

<https://www.openaccessrepository.it>

## Open Access Repository e FAIR data

### Open Access Repository

- L'archivio istituzionale dell'INFN (OAR) è stato installato, gestito e sviluppato a Catania a partire dal 2015, basato su Zenodo/Invenio v3.
- Costituisce uno dei principali strumenti INFN per rendere **FAIR** i Prodotti della Ricerca.
- I suoi contenuti sono stati migrati su una nuova istanza (RDM Invenio v12) ospitata presso i Servizi Nazionali al CNAF e attiva dal 13 Maggio 2025.
- Ambisce a costituire una raccolta completa dei prodotti della ricerca collegati a INFN (i.e. prodotti con almeno un autore INFN) organizzati per gruppi (**communities**).
- Ospita inoltre contenuti di altri EPR (per es. **ISPR**, MoU in preparazione).
- Può raccogliere molte tipologie di prodotti: pdf, immagini, multimedia file, dataset
- Per ogni inserimento viene assegnato un Digital Object Identifier, registrato su **DATA CITE** (cfr. <https://datacite.org/>)
- La pubblicazione avviene previa validazione per curatela (Disciplinare INFN, art. 5, doi:10.15161/oar.it/143269)

### Vocabolari

- Possono essere definiti vocabolari per molte tipologie di metadati: nomi autori (con codice ORCID), affiliazione (con codice ROR), subjects, keywords, experiment. Dalla versione 12 possono essere aggiornati, il che aggiunge un grado di flessibilità molto importante.
- Sfruttando questa possibilità, possiamo arricchire il vocabolario per autori e affiliazioni contestualmente all'arrivo di nuove informazioni.

### Arricchimento metadati

- Molte Informazioni inserite personalmente dagli utenti erano spesso inesatte (es. Affiliazione: la stessa sede ricorre espressa in molti modi differenti). Impiegando un classificatore basato su una CNN (credits: M. Gattari) abbiamo potuto uniformare le affiliazioni INFN per molti record migrati. Inoltre, quando possibile sono stati associati i codici ORCID agli autori, per record che ne erano inizialmente privi.

### Custom Fields

- La possibilità di definire maschere di inserimento locali per casi specifici permette di ottemperare a disposizioni normative (es. Accettazione di policy) e inserire valori non ordinariamente considerati in altri OAR (es. Article Processing Charge).

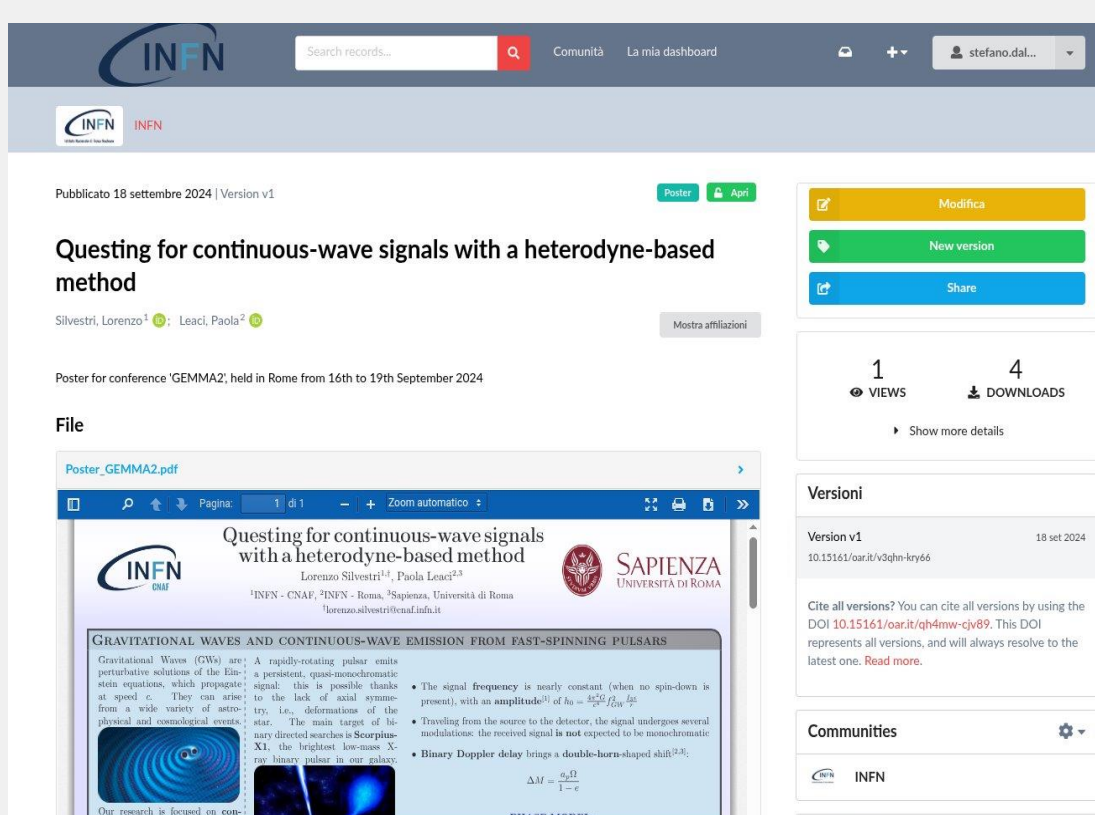
### Attività in corso

- Ingestion dei preprint INSPIRE e arXiv, archivio delle note tecniche INFN (a partire dal 1955) collana Frascati Physics Series e altro.
- Correzione / perfezionamento di metadati post migrazione (es. errata Separazione Nome / Cognome, assegnazione ad altra community, o altro).

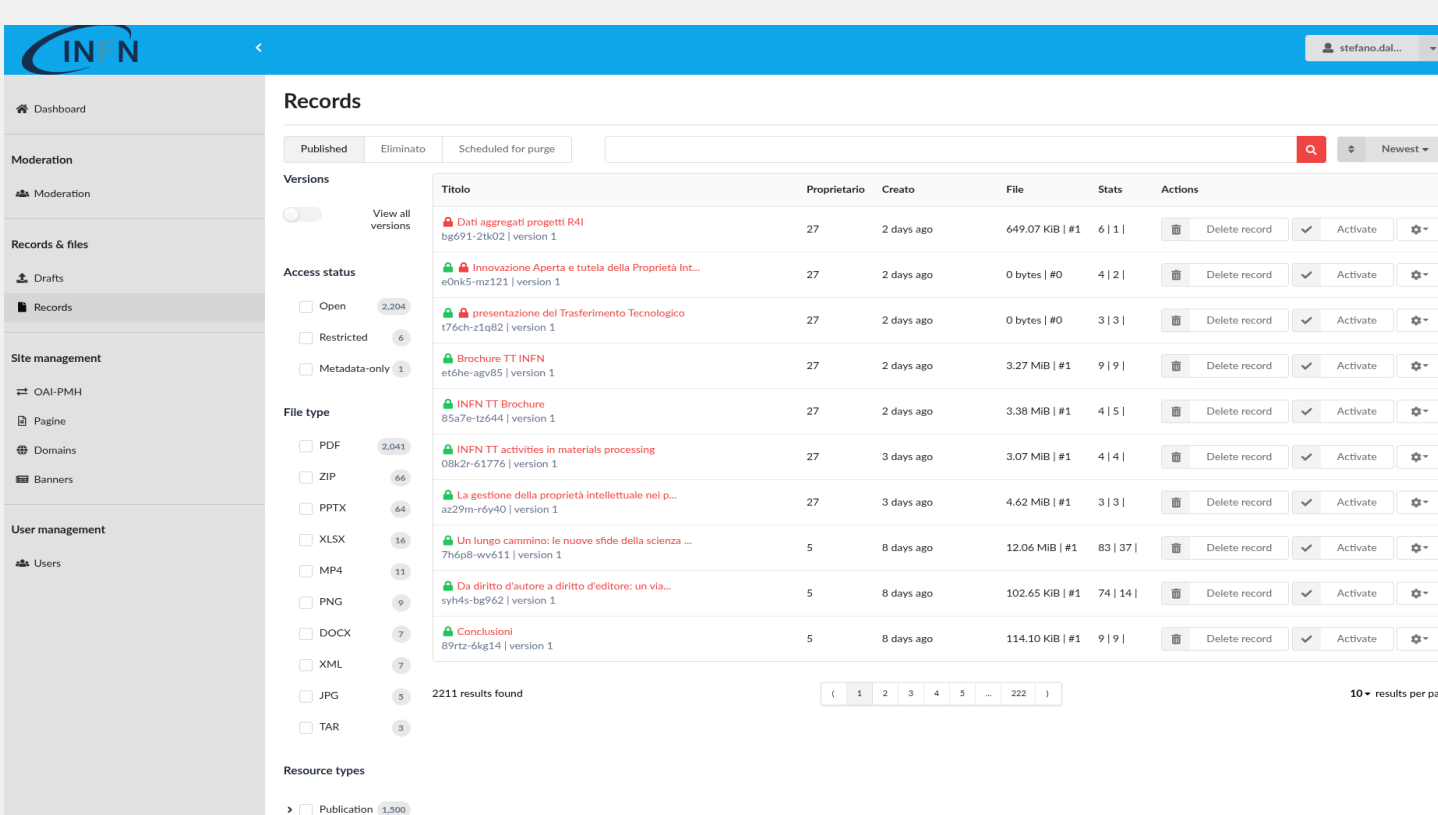
### Management e coordinamento

- Riunioni settimanali, minute e documenti su <https://openscience.infn.workplace.garr.it>
- attività interne: GitLab Ticketing System (<https://baltig.infn.it/invenio/invenio-infn/-/issues>)
- Script, configuration file, documentazione interna etc. sempre su <https://baltig.infn.it>.

### Published record



### Admin dashboard



### Deployment

- L'archivio OAR (**Open Access Repository**) dell'INFN è installato sull'infrastruttura **GHA** (Alta Affidabilità Geografica) dei Servizi Nazionali presso il **CNAF**
- Il servizio si basa sul software **InvenioRDM**, sviluppato al CERN
- Installazione customizzata, predisposta per **scalabilità orizzontale**
- Autenticazione** attualmente: IdP INFN, ORCID. In corso integrazione con un sistema di autenticazione federata.

### Note sulla migrazione

Invenio fornisce tool per l'upgrade da una versione alla successiva. A causa di customizzazioni locali e della grossa differenza tra le versioni, si è proceduto autonomamente, implementando una procedura ETL (Extract, Transform, Load)

**Extraction:** metadati su PostgreSQL database, dati su filesystem, copiati con rsync al CNAF

**Transform:** I valori estratti vanno adattati al formato dello schema destinazione, superando diverse incompatibilità.

**Load:** I metadati trasformati vengono caricati su RDM v12 attraverso le web API.

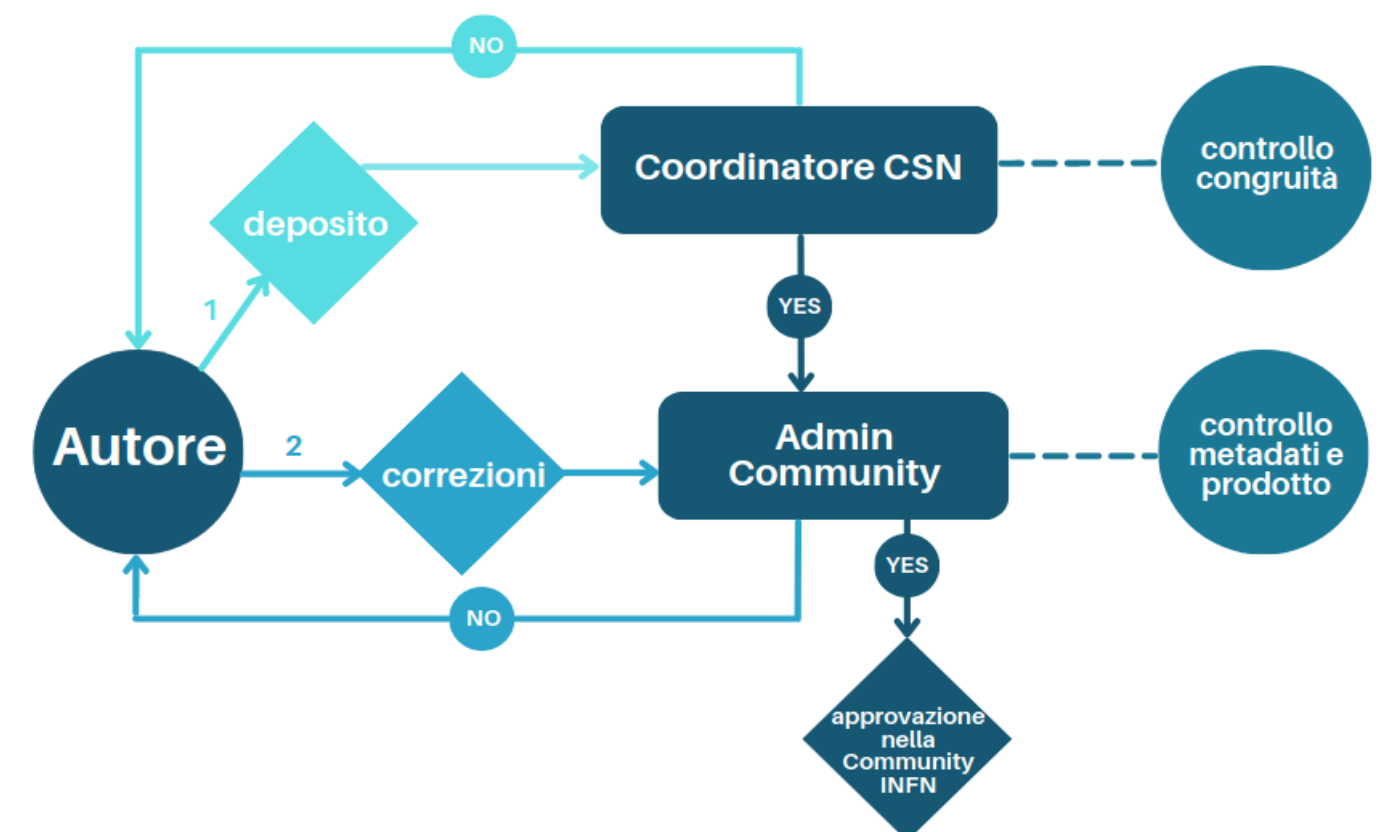
La migrazione è avvenuta partendo da una copia del backend database PostgreSQL di Invenio v3. In questa sono state create diverse **tabelle ausiliarie** con trasformazioni e/o arricchimenti successivi (in particolare normalizzazione affiliazioni, codici ORCID) fino ad ottenere una tabella aux\_records finale, contenente metadati e riferimenti ai datafiles per ogni record. Da questa tabella si è prodotto un dump di ogni record in formato json, con relativi puntatori ai files. Infine questi json files sono stati caricati sulla nuova istanza attraverso le web API di invenio impiegando python scripts prodotti allo scopo.

### Ringraziamenti

R. Rotondo, S. Monforte (INFN Sez. Catania), M. Gattari, Gruppo di Lavoro Open Science, Direzione Sistemi Informativi

### Approvazione dei Prodotti per la pubblicazione

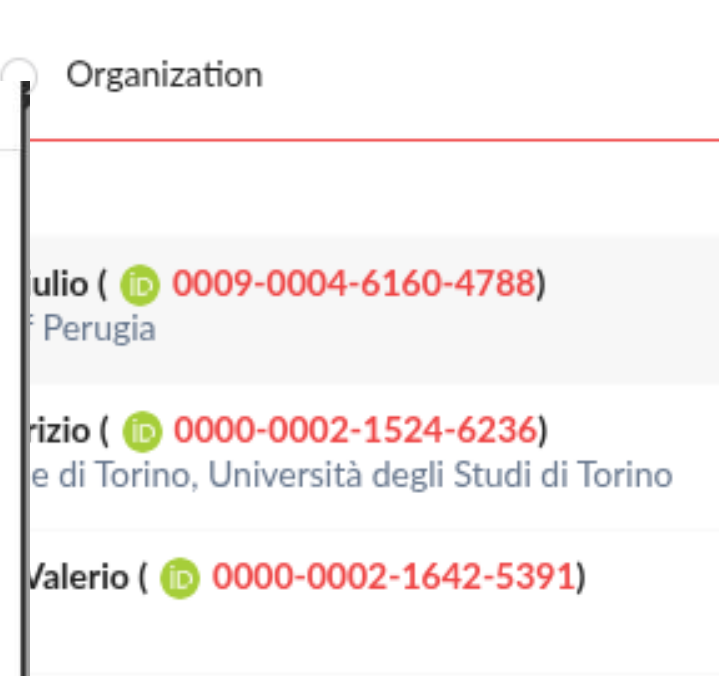
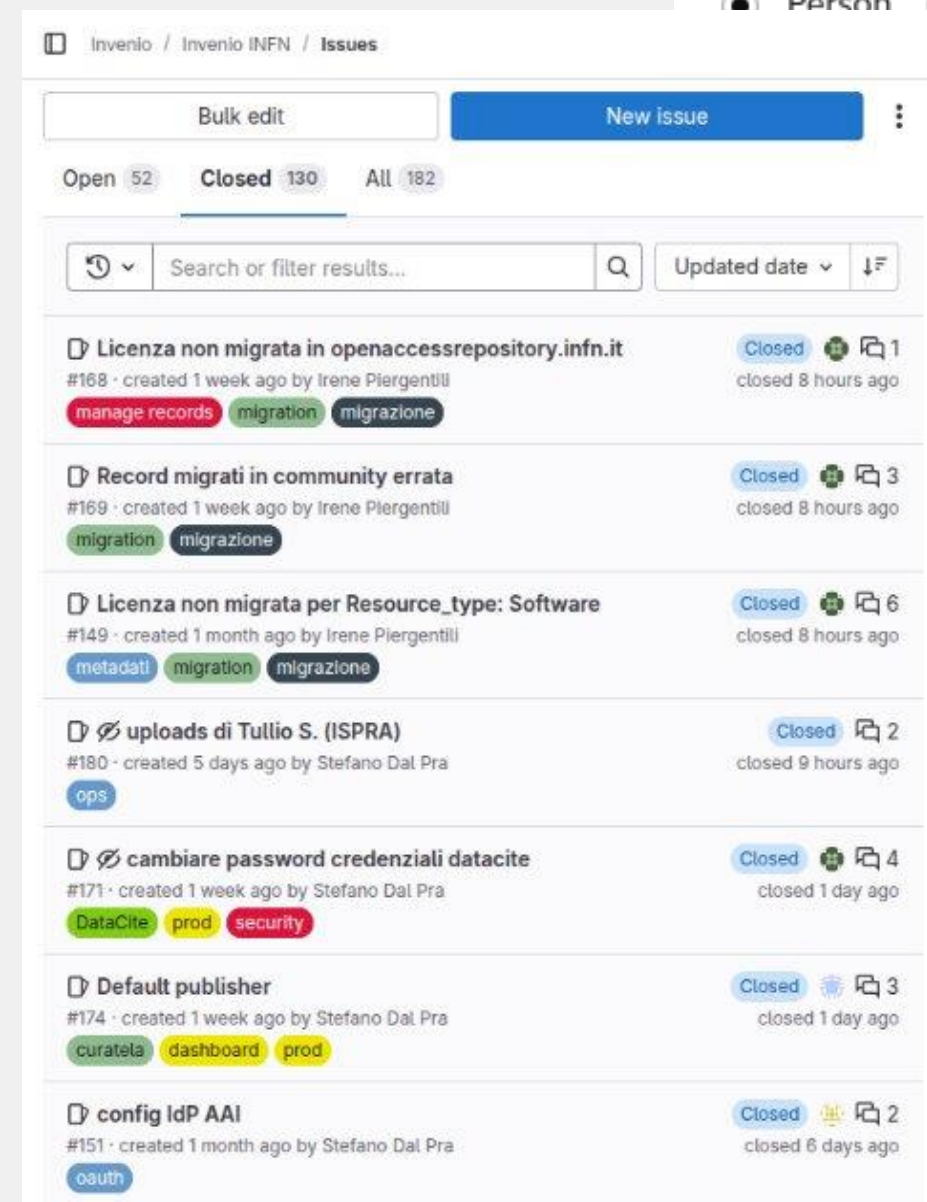
#### Flowchart approvazione dei prodotti in OAR INFN



#### Author completion

##### Add creator

#### Ticketing system (gitlab)



Il vocabolario dei nomi associa il codice ORCID e l'affiliazione (non mostrato in figura) anche in base al codice ROR (<https://ror.org>)

#### Published records by type

