



# Scienza Aperta @ INFN

Concezio Bozzi, INFN Sezione di Ferrara

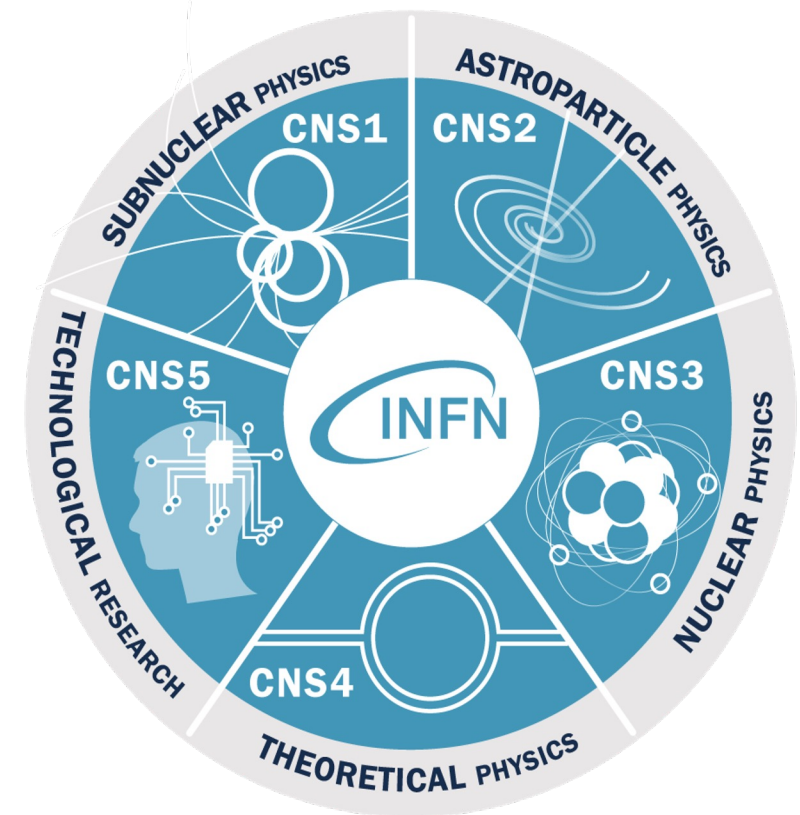
11 giugno 2024

# L'Istituto Nazionale di Fisica Nucleare



L'INFN è stato fondato nel 1951 e ha una lunga e prestigiosa storia che discende da Enrico Fermi e i ragazzi di via Panisperna

Una comunità di oltre 6,000 persone ~  
**25%** con borse di dottorato, post-dottorato e contratti di ricerca.



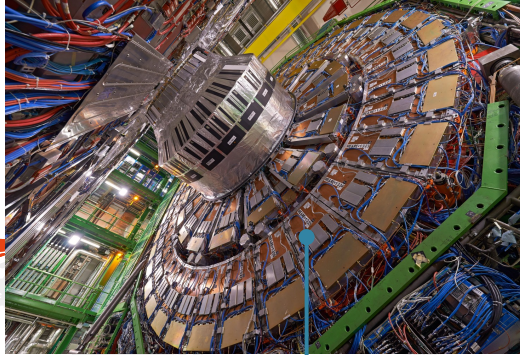
5 linee di ricerca/commissioni scientifiche nazionali + Comitato di Coordinamento Nazionale del Calcolo (CNC)

# Infrastrutture dell'INFN

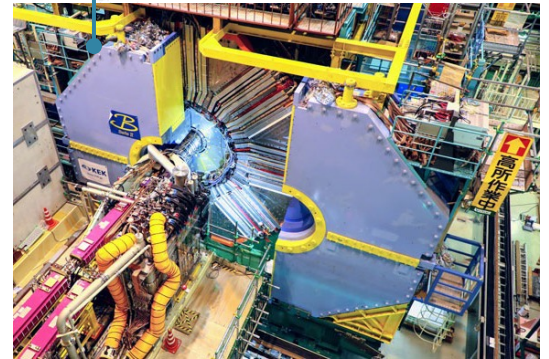
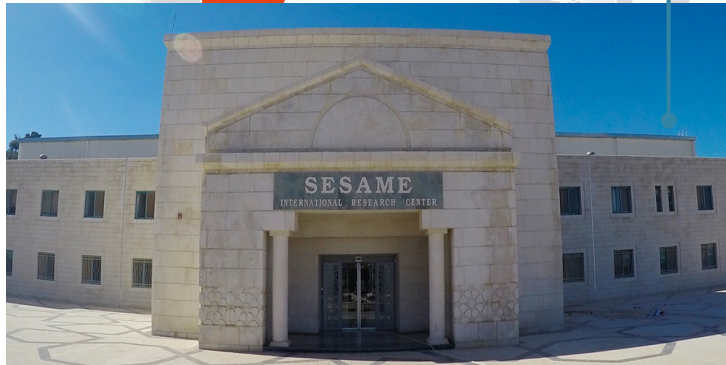
- 4 Laboratori Nazionali
- 20 Sezioni
- 6 Gruppi associati
- 3 Centri Nazionali e Scuole
- 1 Consorzio internazionale







# Un DNA internazionale

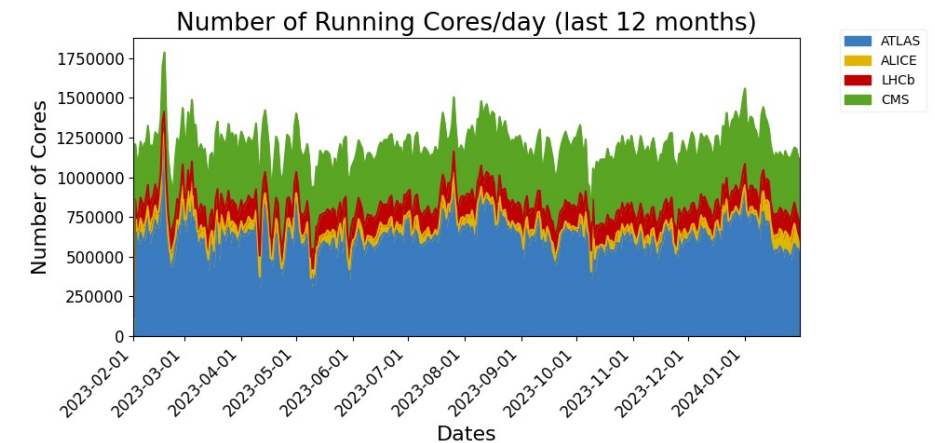
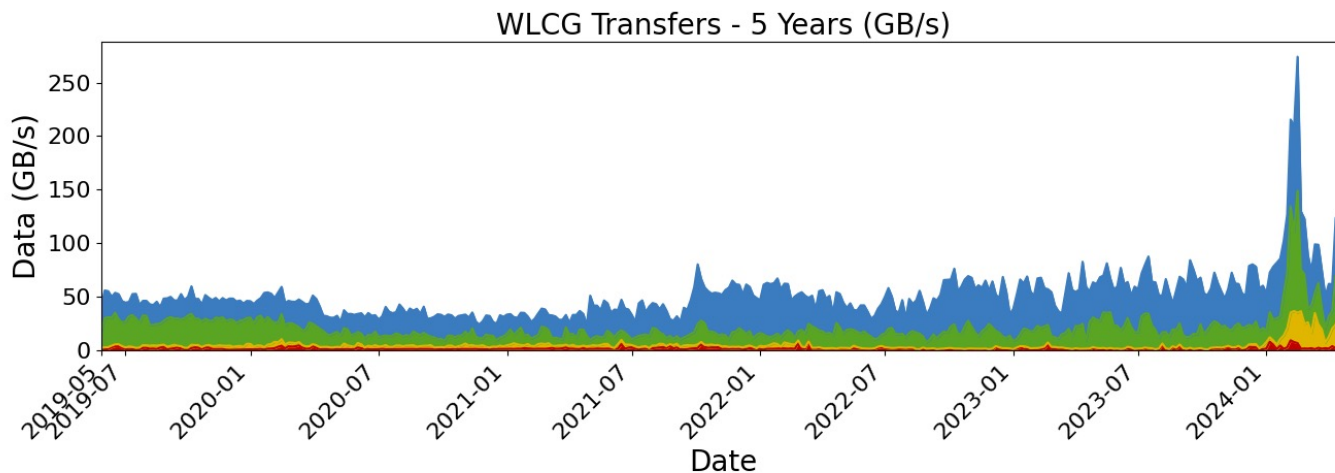
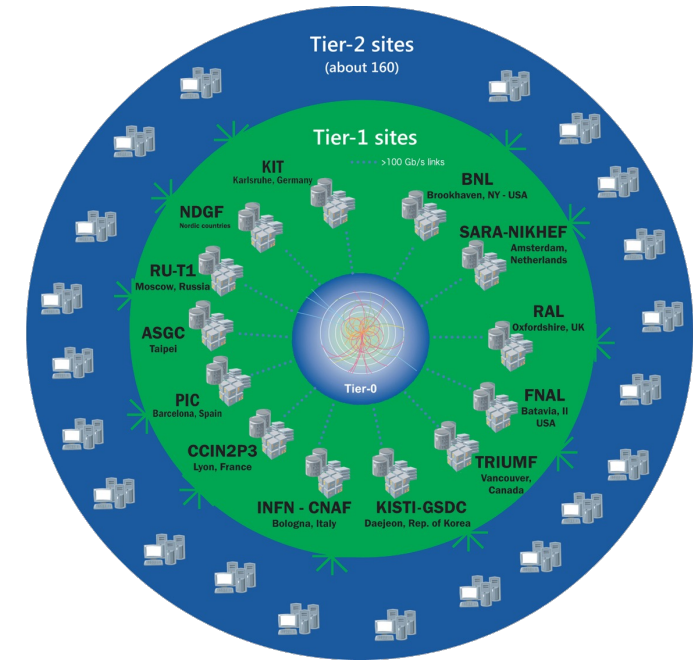




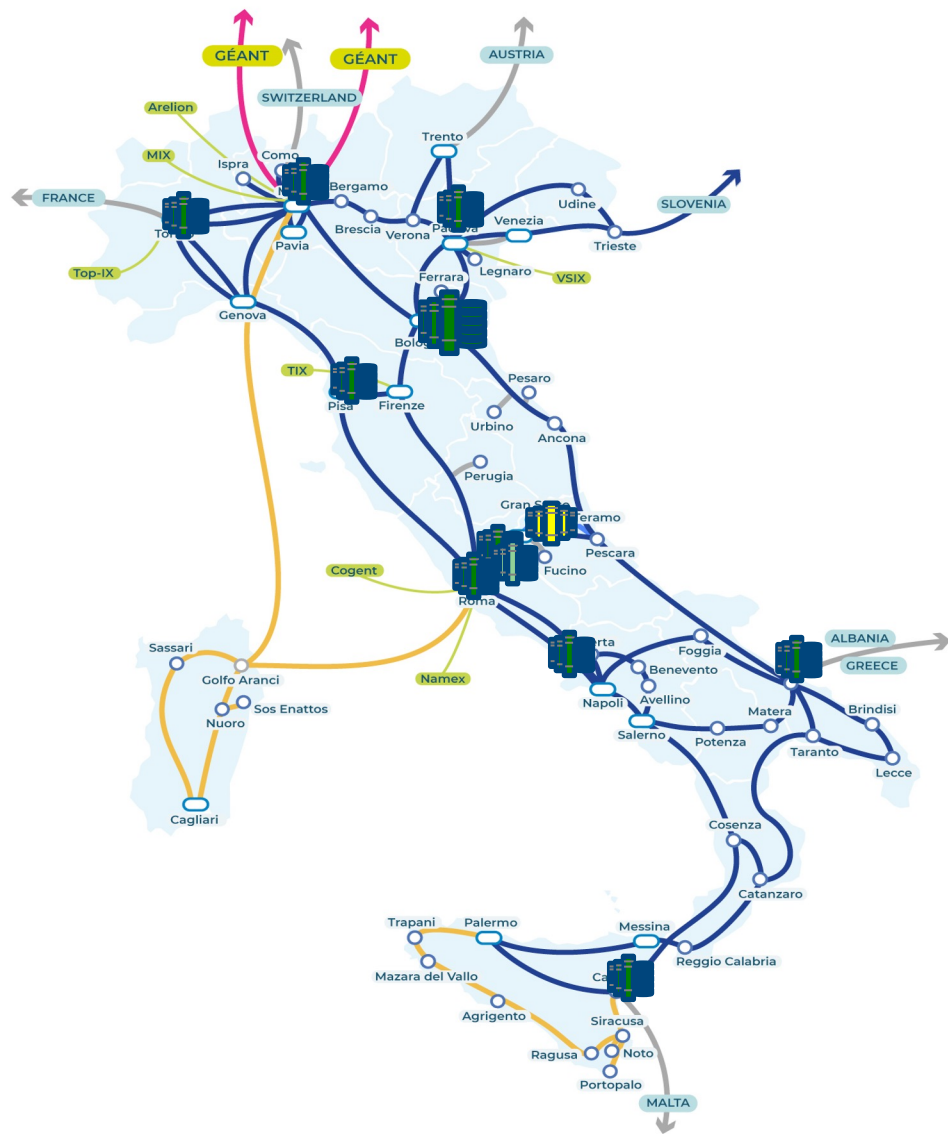
# Big data, collaborazioni internazionali

Grandi risorse di calcolo distribuito, e.g.  
“Worldwide LHC Computing Grid”

- ~175 centri di calcolo
- ~1.25 milioni di core usati di continuo
- ~1EB storage (disco e nastro)
- ~50-100GB/s dati continuamente trasferiti



# DataCloud INFN: infrastruttura di calcolo e big data



Attività consolidate in tecnologie e soluzioni IT distribuite, infrastruttura "stato dell'arte" **basata su grid e cloud.**

L'INFN opera servizi Grid e Cloud su un'infrastruttura federata:

- **1 grosso centro nazionale**, al CNAF (Bologna) – con area certificata ISO/IEC 27001, 27017, 27018
- **9 centri di media grandezza**

I centri di calcolo sono connessi tramite link dedicati a **10-100 Gbit/s** attraverso la **rete GARR.**

Complessivamente:

- **circa 190,000 CPU cores**
- **250 PetaBytes di spazio disco** enterprise-level
- **200 PetaBytes di archivio su nastro**

L'accesso alle risorse viene fornito tramite interfacce cloud.



# Data preservation / scienza aperta in HEP

- L'INFN è nativamente coinvolto nei programmi di conservazione dei dati di laboratori HEP (CERN), esperimenti (LHC et al) e collaborazioni (DPHEP)
- [Rassegna recente](#)
  - capitolo su dati aperti
  - politiche su dati aperti
  - definiti 4 livelli di conservazione
- Conservazione dei dati ispirata a criteri di dati aperti and principi FAIR
- Aspetto fondamentale: cura dei dati su tutto l'orizzonte temporale
  - inclusa la possibilità di riprocessarli
- Sviluppo di strumenti quali [Zenodo](#) e [Reana](#) per facilitare l'adozione dei principi della scienza aperta

Eur. Phys. J. C (2023) 83:795  
https://doi.org/10.1140/epjc/s10052-023-11885-1

THE EUROPEAN  
PHYSICAL JOURNAL C

Review

**Data preservation in high energy physics**

DPHEP Collaboration

T. Basaglia<sup>1</sup>, M. Bellis<sup>2,b</sup>, J. Blomer<sup>1</sup>, J. Boyd<sup>1</sup>, C. Bozzi<sup>3</sup>, D. Britzger<sup>4</sup>, S. Campana<sup>1</sup>, C. Cartaro<sup>5</sup>, G. Chen<sup>6</sup>, B. Couturier<sup>1</sup>, G. David<sup>7,c</sup>, C. Diaconu<sup>8,a</sup>, A. Dobrin<sup>9</sup>, D. Duellmann<sup>1</sup>, M. Ebert<sup>10</sup>, P. Elmer<sup>11</sup>, J. Fernandes<sup>1</sup>, L. Fields<sup>21</sup>, P. Fokianos<sup>1</sup>, G. Ganis<sup>1</sup>, A. Geiser<sup>12</sup>, M. Gheata<sup>9</sup>, J. B. Gonzalez Lopez<sup>1</sup>, T. Hara<sup>13</sup>, L. Heinrich<sup>1</sup>, M. Hildreth<sup>21</sup>, K. Herner<sup>14</sup>, B. Jayatilaka<sup>14</sup>, M. Kado<sup>1</sup>, O. Keeble<sup>1</sup>, A. Kohls<sup>1</sup>, K. Naim<sup>1</sup>, C. Lange<sup>20</sup>, K. Lassila-Perini<sup>15</sup>, S. Levonian<sup>12</sup>, M. Maggi<sup>22</sup>, Z. Marshall<sup>18</sup>, P. Mato Vila<sup>1</sup>, A. Mečionis<sup>1</sup>, A. Morris<sup>17</sup>, S. Piano<sup>16</sup>, M. Potekhin<sup>7</sup>, M. Schröder<sup>1</sup>, U. Schwickerath<sup>1</sup>, E. Sexton-Kennedy<sup>14</sup>, T. Šimko<sup>1</sup>, T. Smith<sup>1</sup>, D. South<sup>12</sup>, A. Verbytskyi<sup>4</sup>, M. Vidal<sup>1</sup>, A. Vivace<sup>1</sup>, L. Wang<sup>6</sup>, G. Watt<sup>19</sup>, T. Wenaus<sup>7</sup>



[Zenodo](#): multi-disciplinary open research repository hosted by CERN and commissioned by the European Commission through OpenAIRE project to support the Open Data and Open Access movements in Europe. Data, software and all research related digital artifacts can be submitted. All uploads are assigned with a digital object identifier (DOI) to make them citeable. Zenodo provides versioning, metrics, integration with GitHub.

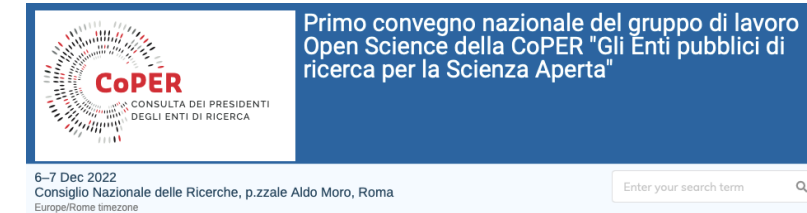


Reproducible research data analysis platform

[Reana](#) re-executes analyses on a workflow-aware, container-based computing backend ("Workflow as a service")

# Attività INFN in progetti di scienza aperta

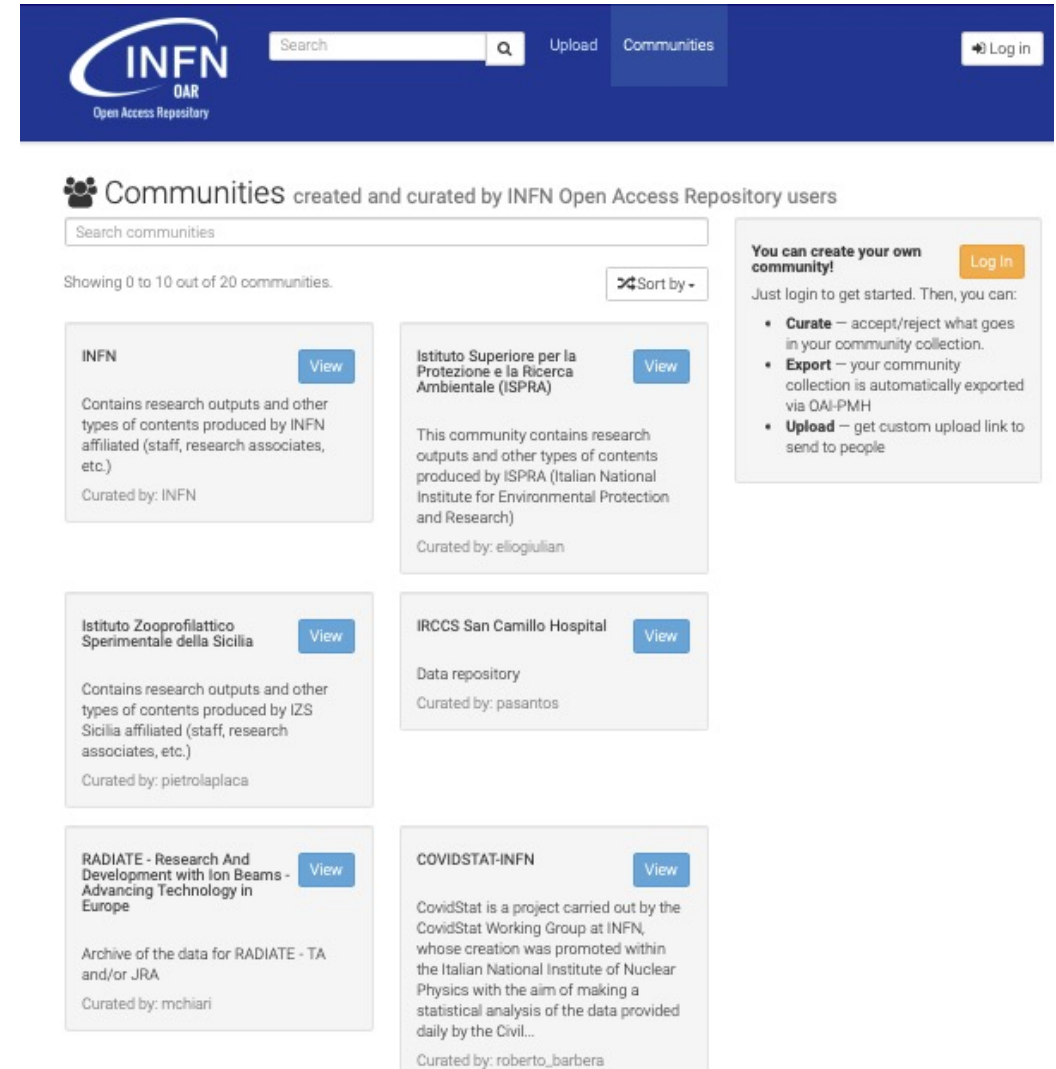
- Partecipazione a iniziative italiane
  - Co-coordinamento del [gruppo Scienza Aperta di CoPER](#)
  - Membro dell'Italian Computing and Data Infrastructure ([ICDI](#))
- Partecipazione in progetti collegati all'European Open Science Cloud ([EOSC](#))
- Co-coordinamento e ruoli chiave in progetti multi-disciplinari PNRR ([ICSC/TeraBit/...](#)) che prevedono
  - Implementazione di politiche e pratiche di scienza aperta
  - Piani di data management FAIR





# Attività INFN in pratica

- Strumenti chiave già presenti e diffusi naturalmente in HEP/INFN ben prima che si codificassero ufficialmente i principi della scienza aperta
- Software: <https://baltig.infn.it>
  - basato su GitLab, con Control Version, permette di individuare univocamente prodotti sw
- Software + Dati: </cvmfs/>
  - filesystems accessibili tramite installazione di un client.
- Archivi aperti
  - implementato il primo [archivio Zenodo](#) fuori dal CERN
  - Facilmente adattabile e clonabile per altre comunità
  - Prototipo di un'infrastruttura nazionale di archivi interoperabili
- Una pletera di prodotti e piattaforme open-source
  - [Indico](#), [sympa](#), [pandora](#), [newdle](#),...



The screenshot shows the 'Communities' page of the INFN Open Access Repository. The header includes the INFN logo, a search bar, and navigation links for 'Upload' and 'Communities'. Below the header, there's a search bar for communities and a 'Log In' button. The main content area displays a grid of community cards, each with a 'View' button. The communities listed are:

- INFN**: Contains research outputs and other types of contents produced by INFN affiliated (staff, research associates, etc.). Curated by: INFN.
- Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA)**: This community contains research outputs and other types of contents produced by ISPRA (Italian National Institute for Environmental Protection and Research). Curated by: ellogiulian.
- Istituto Zooprofilattico Sperimentale della Sicilia**: Contains research outputs and other types of contents produced by IZS Sicilia affiliated (staff, research associates, etc.). Curated by: pietrolaplaca.
- IRCCS San Camillo Hospital**: Data repository. Curated by: pasantos.
- RADIATE - Research And Development with Ion Beams - Advancing Technology in Europe**: Archive of the data for RADIATE - TA and/or JRA. Curated by: mchiari.
- COVIDSTAT-INFN**: CovidStat is a project carried out by the CovidStat Working Group at INFN, whose creation was promoted within the Italian National Institute of Nuclear Physics with the aim of making a statistical analysis of the data provided daily by the Civil... Curated by: roberto\_barbera.

On the right side, there is a call to action: 'You can create your own community!' with a 'Log In' button. Below this, it says 'Just login to get started. Then, you can:' followed by a list of actions: 'Curate' (accept/reject what goes in your community collection), 'Export' (your community collection is automatically exported via OAI-PMH), and 'Upload' (get custom upload link to send to people).

# Attività INFN in pratica

- [ALEPH open data su EUDAT](#)
- [CovidStat19](#): un esempio di scienza aperta basato su principi FAIR
- ESCAPE Virtual Research Environment (VRE) sulla materia oscura
- Elaborazione di Data Management Plans per gli esperimenti ai 4 laboratory nazionali



The screenshot shows the COVIDSTAT INFN website. At the top, there is a navigation bar with links for 'Sommary', 'Grafici ISS', 'Grafici DPC', 'Tabelle', 'Mappa', 'Animazioni', 'Blog', and 'Aiuto'. Below the navigation bar, the main header reads 'COVIDSTAT INFN'. A sub-header states: 'Il sito CovidStat INFN è aggiornato su base settimanale solo con i dati dell'Istituto Superiore di Sanità.' Below this, a paragraph explains that other graphics are no longer updated due to data source changes. A row of six icons represents different data visualization tools: 'Grafici ISS', 'Grafici DPC', 'R<sub>t</sub>', 'Tabelle', 'Mappa', and 'Blog'. The 'Ultimi post dal blog' section features three articles: 'Stima del tasso di mortalità da COVID-19', 'Attenzione, la "discesa" sembra finita', and 'Una stima della mortalità della variante Omicron dai dati del Sud Africa'. Each article includes a short summary and a 'Continua a leggere' link. At the bottom, there are four preview images of data visualizations: a line graph of 'Tempo Raddoppio (giorni), Casi totali, fit a 5 gg', a map of Italy with regional data, a map of Africa with regional data, and a table of data.



# Attività INFN in pratica

- [ALEPH open data su EUDAT](#)
- [CovidStat19](#): un esempio di scienza aperta basato su principi FAIR
- ESCAPE Virtual Research Environment (VRE) sulla materia oscura
- Elaborazione di Data Management Plans per gli esperimenti ai 4 laboratory nazionali

## EOSC Future – Virtual Research Environment

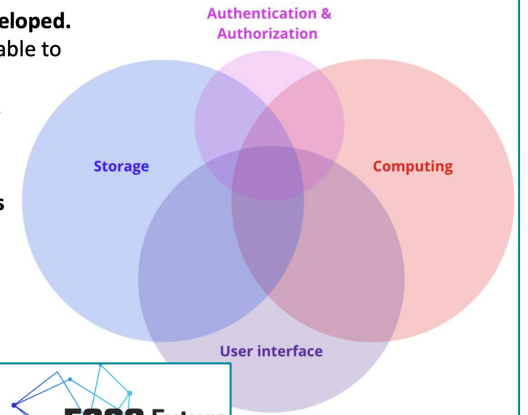


The ESCAPE Virtual Research Environment (VRE) is being developed.

- Collaborative online platform where science projects are able to develop and share end-to-end workflows.
- Access all digital content needed, in compliance with FAIR principles.

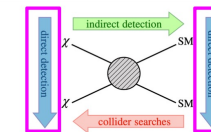
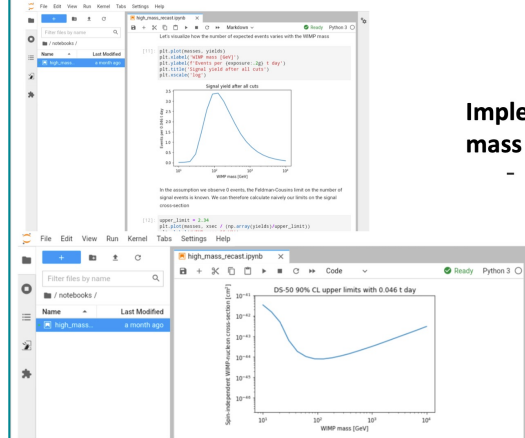
The VRE is actively used by members of both science projects

- Analyze, preserve, run, share analyses.
- Education and outreach.



See talk by E. Gazzarrini

## DarkSide Plans



Implemented a reanalysis tool for a high-mass search on the VRE platform.

- Output: **DarkSide50** exclusion curve for WIMP-nucleon cross section.

Further work is ongoing.

- Low mass analysis to be implemented.
- Different **theoretical models** (WIMP halo, argon response...) can be inserted by the user to produce different limit results.
- Working towards first open implementation.

# Piano Nazionale Scienza Aperta (PNSA)

Approvazione nel 2022

*Dati della ricerca: "As open as possible, as closed as necessary"*



## PIANO NAZIONALE PER LA SCIENZA APERTA (2021 – 2027)

Assi di intervento	Piano di intervento
1. Pubblicazioni scientifiche	<ul style="list-style-type: none"><li>• Accesso aperto alle pubblicazioni</li><li>• Forme non commerciali di pubblicazione</li><li>• Quadro normativo in materia di diritto d'autore</li><li>• Sistema di monitoraggio</li><li>• Risorse formative aperte</li></ul>
2. Dati della ricerca	<ul style="list-style-type: none"><li>• <i>FAIRification</i> nel sistema ricerca</li><li>• Integrazione in EOSC</li><li>• Produzione collaborativa di dati</li><li>• Formazione delle figure tecniche</li></ul>
3. Valutazione della ricerca	<ul style="list-style-type: none"><li>• Processi e criteri di valutazione</li><li>• Collaborazione tra istituzioni e tra ricercatori</li><li>• Pubblicare in accesso aperto</li><li>• Revisione paritaria aperta</li><li>• Infrastruttura nazionale</li></ul>
4. Scienza aperta, comunità scientifica e partecipazione europea	<ul style="list-style-type: none"><li>• Percorso organico verso la scienza aperta</li><li>• Attività di coordinamento a livello europeo</li></ul>
5. Apertura dei dati della ricerca su SARS-COV-2 e Covid-19	<ul style="list-style-type: none"><li>• Portale nazionale per dati FAIR e testi su COVID19</li><li>• Modelli di dati aperti sulla salute pubblica</li></ul>

**Dal 2007:** coordinatore italiano di [scoap3.org](http://scoap3.org),  
[oa2020.org](http://oa2020.org), [scienceeurope.org](http://scienceeurope.org), [coalition-s.org](http://coalition-s.org)

Tab. 1 - Struttura e obiettivi del piano

# Piano Nazionale Scienza Aperta (PNSA)

Approvazione nel 2022

Dati della ricerca: "As open as possible, as closed as necessary"



## PIANO NAZIONALE PER LA SCIENZA APERTA (2021 – 2027)

Assi di intervento	Piano di intervento
1. Pubblicazioni scientifiche	<ul style="list-style-type: none"> <li>• Accesso aperto alle pubblicazioni</li> <li>• Forme non commerciali di pubblicazione</li> <li>• Quadro normativo in materia di diritto d'autore</li> <li>• Sistema di monitoraggio</li> <li>• Risorse formative aperte</li> </ul>
2. Dati della ricerca	<ul style="list-style-type: none"> <li>• FAIRification nel sistema ricerca</li> <li>• Integrazione in EOSC</li> <li>• Produzione collaborativa di dati</li> <li>• Formazione delle figure tecniche</li> </ul>
3. Valutazione della ricerca	<ul style="list-style-type: none"> <li>• Processi e criteri di valutazione</li> <li>• Collaborazione tra istituzioni e tra ricercatori</li> <li>• Pubblicare in accesso aperto</li> <li>• Revisione paritaria aperta</li> <li>• Infrastruttura nazionale</li> </ul>
4. Scienza aperta, comunità scientifica e partecipazione europea	<ul style="list-style-type: none"> <li>• Percorso organico verso la scienza aperta</li> <li>• Attività di coordinamento a livello europeo</li> </ul>
5. Apertura dei dati della ricerca su SARS-COV-2 e Covid-19	<ul style="list-style-type: none"> <li>• Portale nazionale per dati FAIR e testi su COVID19</li> <li>• Modelli di dati aperti sulla salute pubblica</li> </ul>

- **Politiche di dati aperti** per esp. LHC@CERN, OPERA@LNGS,...
- **Primo archivio aperto italiano** basato su ZENODO-INVENIO
- Gruppo di lavoro per la stesura dei **DMP dei laboratori nazionali**
- Percorso di **formazione data steward** attivato con UNIBO
- Approvato disciplinare per **accesso aperto ai prodotti della ricerca**

Tab. 1 - Struttura e obiettivi del piano



# Piano Nazionale Scienza Aperta (PNSA)

Approvazione nel 2022

*Dati della ricerca: "As open as possible, as closed as necessary"*



## PIANO NAZIONALE PER LA SCIENZA APERTA (2021 – 2027)

Assi di intervento	Piano di intervento
1. Pubblicazioni scientifiche	<ul style="list-style-type: none"> <li>• Accesso aperto alle pubblicazioni</li> <li>• Forme non commerciali di pubblicazione</li> <li>• Quadro normativo in materia di diritto d'autore</li> <li>• Sistema di monitoraggio</li> <li>• Risorse formative aperte</li> </ul>
2. Dati della ricerca	<ul style="list-style-type: none"> <li>• FAIRification nel sistema ricerca</li> <li>• Integrazione in EOSC</li> <li>• Produzione collaborativa di dati</li> <li>• Formazione delle figure tecniche</li> </ul>
3. Valutazione della ricerca	<ul style="list-style-type: none"> <li>• Processi e criteri di valutazione</li> <li>• Collaborazione tra istituzioni e tra ricercatori</li> <li>• Pubblicare in accesso aperto</li> <li>• Revisione paritaria aperta</li> <li>• Infrastruttura nazionale</li> </ul>
4. Scienza aperta, comunità scientifica e partecipazione europea	<ul style="list-style-type: none"> <li>• Percorso organico verso la scienza aperta</li> <li>• Attività di coordinamento a livello europeo</li> </ul>
5. Apertura dei dati della ricerca su SARS-COV-2 e Covid-19	<ul style="list-style-type: none"> <li>• Portale nazionale per dati FAIR e testi su COVID19</li> <li>• Modelli di dati aperti sulla salute pubblica</li> </ul>

•Sottoscritto **accordo con EC** sulla riforma **RA** e **COARA**

Tab. 1 - Struttura e obiettivi del piano

# Piano Nazionale Scienza Aperta (PNSA)

Approvazione nel 2022

Dati della ricerca: "As open as possible, as closed as necessary"



## PIANO NAZIONALE PER LA SCIENZA APERTA (2021 – 2027)

Assi di intervento	Piano di intervento
1. Pubblicazioni scientifiche	<ul style="list-style-type: none"> <li>• Accesso aperto alle pubblicazioni</li> <li>• Forme non commerciali di pubblicazione</li> <li>• Quadro normativo in materia di diritto d'autore</li> <li>• Sistema di monitoraggio</li> <li>• Risorse formative aperte</li> </ul>
2. Dati della ricerca	<ul style="list-style-type: none"> <li>• FAIRification nel sistema ricerca</li> <li>• Integrazione in EOSC</li> <li>• Produzione collaborativa di dati</li> <li>• Formazione delle figure tecniche</li> </ul>
3. Valutazione della ricerca	<ul style="list-style-type: none"> <li>• Processi e criteri di valutazione</li> <li>• Collaborazione tra istituzioni e tra ricercatori</li> <li>• Pubblicare in accesso aperto</li> <li>• Revisione paritaria aperta</li> <li>• Infrastruttura nazionale</li> </ul>
4. Scienza aperta, comunità scientifica e partecipazione europea	<ul style="list-style-type: none"> <li>• Percorso organico verso la scienza aperta</li> <li>• Attività di coordinamento a livello europeo</li> </ul>
5. Apertura dei dati della ricerca su SARS-COV-2 e Covid-19	<ul style="list-style-type: none"> <li>• Portale nazionale per dati FAIR e testi su COVID19</li> <li>• Modelli di dati aperti sulla salute pubblica</li> </ul>

Tab. 1 - Struttura e obiettivi del piano

- 2021— costituito GLOS (**gruppo di lavoro SA**)
- Proponente e co-coordinatore del [gruppo di lavoro SA di CoPER](#)  
Contributo alla [roadmap](#) ("prossimi passi")
- Membro di **EOSC**, **ICDI**, e dell'**OS WG di nupecc.org**

# Piano Nazionale Scienza Aperta (PNSA)

Approvazione nel 2022

*Dati della ricerca: "As open as possible, as closed as necessary"*



## PIANO NAZIONALE PER LA SCIENZA APERTA (2021 – 2027)

Assi di intervento	Piano di intervento
1. Pubblicazioni scientifiche	<ul style="list-style-type: none"><li>• Accesso aperto alle pubblicazioni</li><li>• Forme non commerciali di pubblicazione</li><li>• Quadro normativo in materia di diritto d'autore</li><li>• Sistema di monitoraggio</li><li>• Risorse formative aperte</li></ul>
2. Dati della ricerca	<ul style="list-style-type: none"><li>• <i>FAIRification</i> nel sistema ricerca</li><li>• Integrazione in EOSC</li><li>• Produzione collaborativa di dati</li><li>• Formazione delle figure tecniche</li></ul>
3. Valutazione della ricerca	<ul style="list-style-type: none"><li>• Processi e criteri di valutazione</li><li>• Collaborazione tra istituzioni e tra ricercatori</li><li>• Pubblicare in accesso aperto</li><li>• Revisione paritaria aperta</li><li>• Infrastruttura nazionale</li></ul>
4. Scienza aperta, comunità scientifica e partecipazione europea	<ul style="list-style-type: none"><li>• Percorso organico verso la scienza aperta</li><li>• Attività di coordinamento a livello europeo</li></ul>
5. Apertura dei dati della ricerca su SARS-COV-2 e Covid-19	<ul style="list-style-type: none"><li>• Portale nazionale per dati FAIR e testi su COVID19</li><li>• Modelli di dati aperti sulla salute pubblica</li></ul>

Tab. 1 - Struttura e obiettivi del piano

• **Progetto COVIDSTAT** <https://covid19.infn.it>



# Sommario



2021- Costituito GLOS (**gruppo di lavoro per l'Open Science**)



L'INFN aderisce agli assi di intervento codificati dal **Piano Nazionale per la Scienza Aperta**



Con la Delibera n. 16717 del 21/07/2023 è stato approvato il "**Disciplinare per l'accesso aperto ai prodotti della ricerca dell'Istituto Nazionale di Fisica Nucleare**".



**archivio istituzionale OAR** disponibile ad altre comunità e clonabile



Creato un **gruppo di lavoro** per la **stesura dei DMP** dei laboratori nazionali



**Data steward: percorso di formazione** attivato con UNIBO

# Sommario

- Necessaria la **cura di dati** (e metadati) su **tutto il loro ciclo di vita**
  - Inclusa la possibilità di riprocessare e il supporto a un network nazionale di repository interoperabili e open-source
- Necessari **data steward**
  - con conoscenza delle specificità e dei requisiti della comunità scientifica
- Progetti non banali, provenienti "dal basso" e su base volontaria
  - I **fondi europei aiutano**
    - a patto che le attività previste nei progetti siano allineate con le necessità dei ricercatori e delle infrastrutture di ricerca
  - Necessario un **finanziamento complementare** al livello nazionale e istituzionale

# Grazie per l'attenzione

Concezio Bozzi (INFN-FE) – [concezio.bozzi@fe.infn.it](mailto:concezio.bozzi@fe.infn.it)

A nome del gruppo:

Francesca Marchegiani (INFN-LNGS) – [francesca.marchegiani@lngs.infn.it](mailto:francesca.marchegiani@lngs.infn.it)

Luciano Gaido (INFN-TO) – [luciano.gaido@to.infn.it](mailto:luciano.gaido@to.infn.it)

Stefano Bianco (INFN-LNF) – [stefano.bianco@lnf.infn.it](mailto:stefano.bianco@lnf.infn.it)

Daniele Bonacorsi (UNIBO & INFN-BO) – [daniele.bonacorsi@bo.infn.it](mailto:daniele.bonacorsi@bo.infn.it)

Stefano Dal Pra (INFN-CNAF) – [stefano.dalpra@cnaf.infn.it](mailto:stefano.dalpra@cnaf.infn.it)

Sandra Parlati (INFN-LNGS) – [sandra.parlati@lngs.infn.it](mailto:sandra.parlati@lngs.infn.it)

Irene Piergentili (INFN-LNF) – [irene.piergentili@lnf.infn.it](mailto:irene.piergentili@lnf.infn.it)

Lorenzo Rinaldi (UNIBO) – [lorenzo.rinaldi@unibo.it](mailto:lorenzo.rinaldi@unibo.it)

Licenza Creative Commons  
Attribuzione 4.0 Internazionale

