



IIT Dataverse:
**il *repository* dell'Istituto Italiano di
Tecnologia per la conservazione e la
condivisione dei dati FAIR della ricerca**

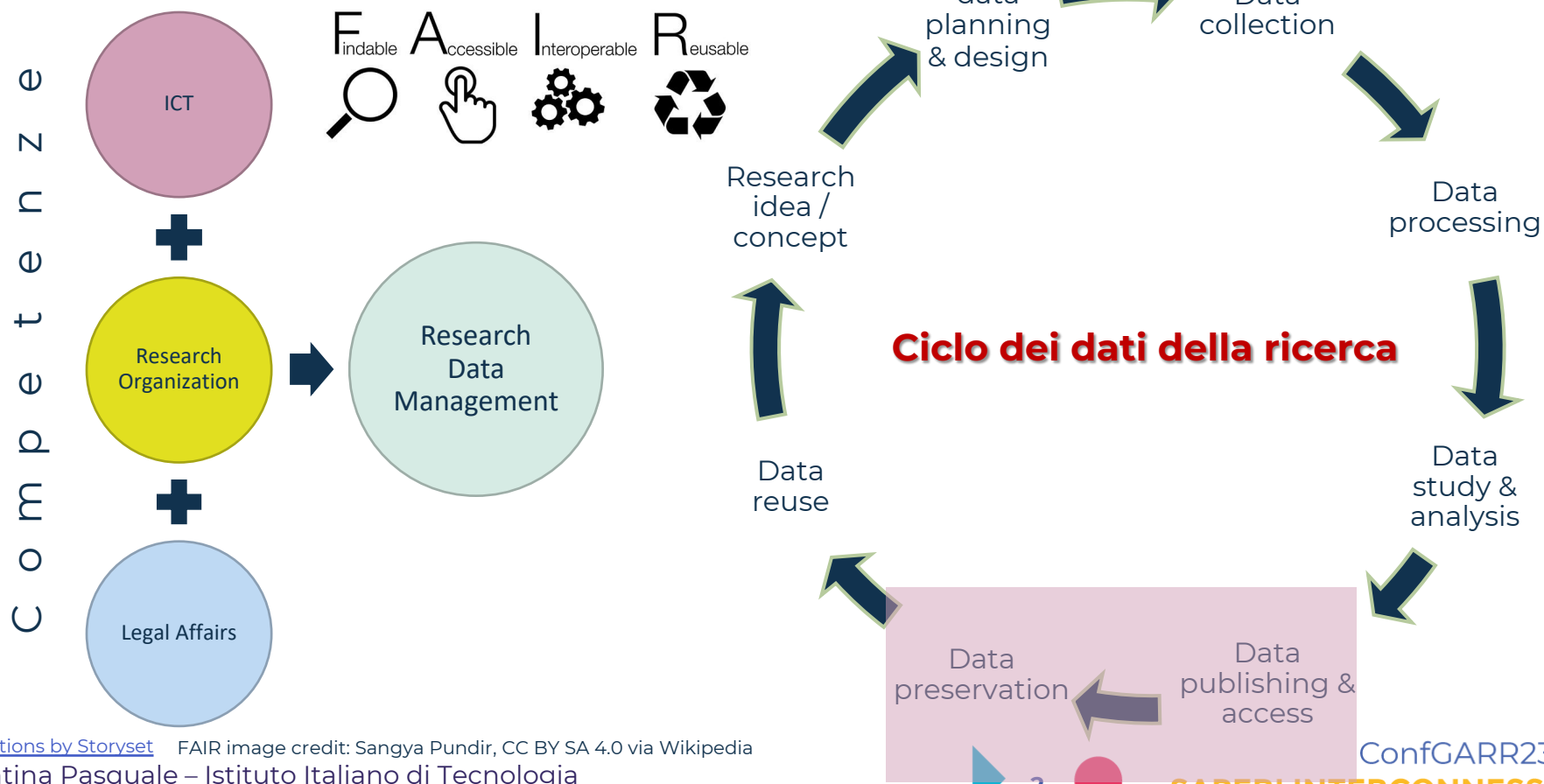
Valentina Pasquale

Istituto Italiano di Tecnologia

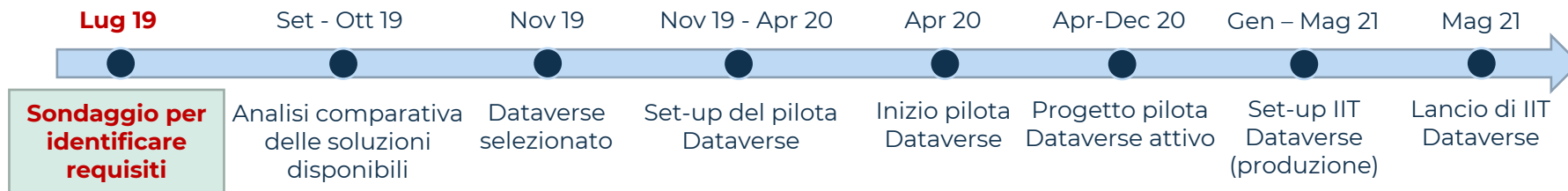


Il contesto

Costruire un servizio di supporto *Research Data Management* che tenga conto di tutte le fasi del ciclo di vita dei dati della ricerca e abiliti la produzione di dati FAIR.



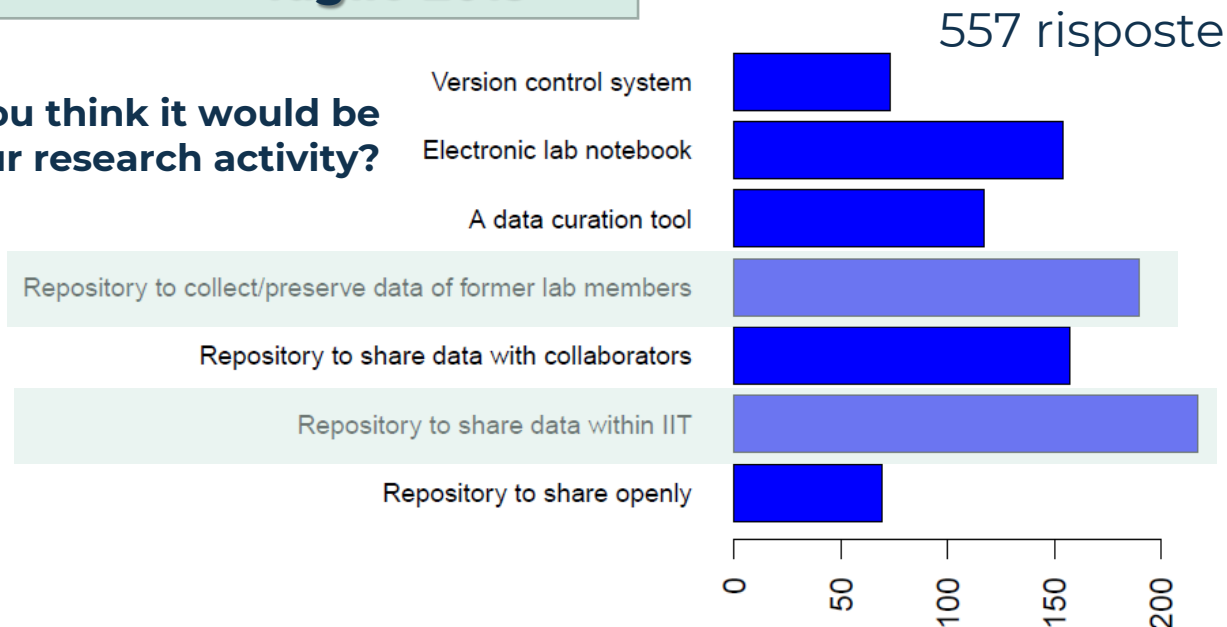
Il punto di partenza: i bisogni dei ricercatori



Survey proposto ai ricercatori IIT a **luglio 2019**

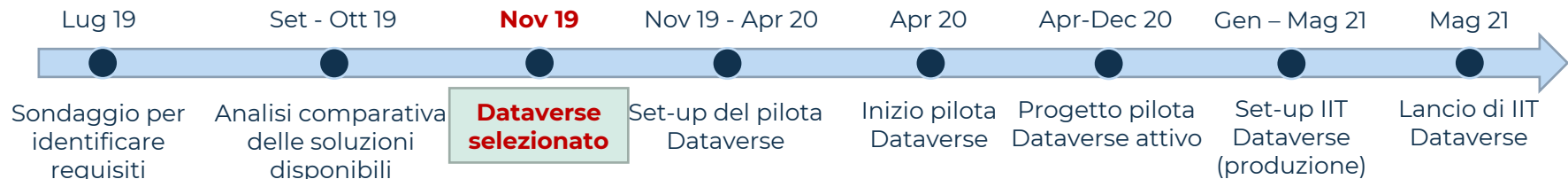
Which RDM tool do you think it would be most useful for your research activity?

Dal sondaggio e' emersa la necessità di dotarsi di un sistema per la conservazione dei dati, soprattutto al fine di abilitare una maggiore collaborazione interna ed esterna.





Dataverse: repository open source per i dati della ricerca

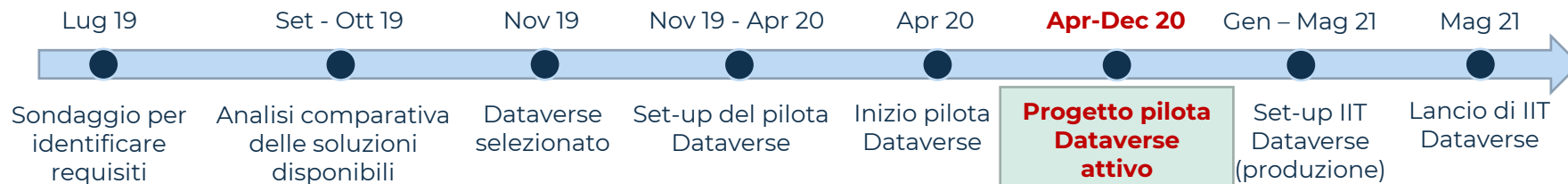


Dataverse e' un'applicazione web *open source* per condividere, conservare, citare, esplorare i dati della ricerca, sviluppata a Harvard University.

- Facilmente configurabile e integrabile con i nostri sistemi interni
- Permette di catalogare e conservare i dati, renderli **FAIR** e aperti quando possibile
- Offre supporto istituzionale per la conservazione a lungo termine
- Adottato a livello mondiale (100 installazioni, 06/2023)



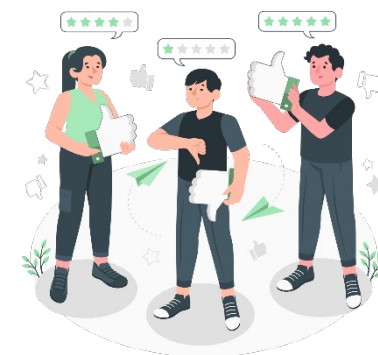
Progetto pilota



Nel 2020, abbiamo avviato un **progetto pilota** per testare Dataverse e raccogliere il feedback dei ricercatori.

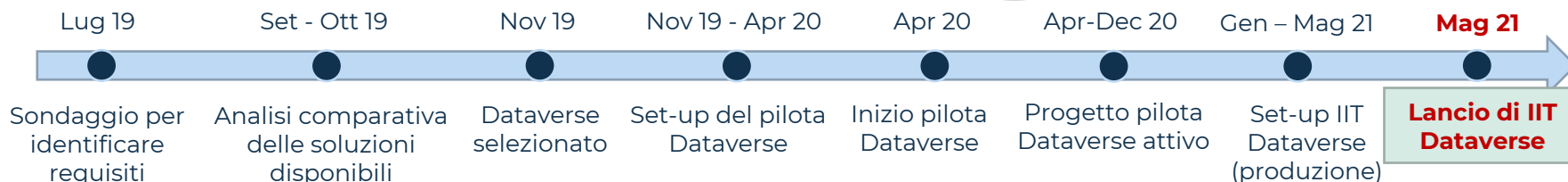
Qualche numero...

- **8** Linee di Ricerca coinvolte
- **4** Domini di Ricerca rappresentati
- **~10** dataset creati



Alla fine del progetto pilota è stato condotto un breve **sondaggio**, al quale sono seguite alcune **interviste** con gli utenti più coinvolti per raccogliere osservazioni e suggerimenti per la configurazione.

Lancio di IIT Dataverse



Why use IIT Dataverse?

Steve left the lab 2y ago. I would like to use his PhD data, but I can't read the HD where he saved them!

He could have used a repository to preserve data for the long-term!

Document datasets with metadata

After a while....

I managed to read Steve's data and the boss asked me to analyze them, but I don't know what "var1" and "var2" mean!

That's why documentation and metadata are important!!

For my ERC I have to preserve and possibly share data making them FAIR.

Do you know we have an institutional repository for research data? It can be of help!

Preserve research datasets in the long-term

Make data FAIR and open when possible

The Dataverse Project iit ISTITUTO ITALIANO DI TECNOLOGIA

CC BY 4.0

19 maggio 2021:

evento di presentazione di IIT Dataverse.

Lancio di IIT Dataverse in produzione per tutte le Linee di Ricerca: inizialmente solo ad uso interno, successivamente abilitato alla pubblicazione dei dataset (fine 2021).

[Illustrations by Storyset](#)

Valentina Pasquale – Istituto Italiano di Tecnologia



ConfGARR23

SAPERI INTERCONNESSI

Attività di formazione e supporto



Formazione: materiale informativo (*demo videos*, linee guida, FAQs, presentazioni) condiviso attraverso la IIT Intranet.

Regolamentazione tramite la pubblicazione di *policy* (sicurezza delle informazioni, privacy, etc.) e termini di utilizzo.



Supporto on demand alla creazione dei dataset, istituzione di un servizio di *data curation* per la pubblicazione dei dataset.



Caratteristiche



Terms of Use



IIT Dataverse 

Search ▾

User Guide

Support

Log In

Metrics

39,184 Downloads

IIT Dataverse is a place for IIT scientists to preserve and share datasets supporting original research and used by Harvard University.

If you have IIT institutional credentials, click on "Log in" in the top right bar to start using the service.

To **get started**, have a look at the links below:

- [FAQ's](#) – more about depositing data in IIT Dataverse
- [Deposit checklist](#) – a short guide to dataset creation
- [IIT Dataverse Intranet page](#) – a more comprehensive reference for IIT Dataverse users

Search this dataverse...



Advanced Search

 **Dataverses (106)**

 **Datasets (9)**

 **Files (3,666)**

Dataverse Category

Research Group (2)

Publication Year

2023 (8)

2022 (12)

2021 (95)

1 to 10 of 115 Results

Replication Data for: Regularized Benne

Mar 28, 2023 - Computational and Chemical Biology public



Decherchi, Sergio; Ciccotti, Giovanni; et al. "Regularized Benne estimators", <https://doi.org/10.48550/dataset.10.48550>

The file contains the simulation data of the p...
input folder there is a file called "lambda.txt"

Replication Data for: On the allosteric p

Mar 27, 2023 - Computational and Chemical Biology public



FINDABLE

- ✓ Integrato con DataCite per l'assegnazione di DOIs.
- ✓ Censito in www.re3data.org.
- ✓ Integrato con OpenAIRE.
- ✓ Indicizzato sui principali motori di ricerca.

ACCESSIBLE

- ✓ Accessibile attraverso protocollo standard HTTPS.
- ✓ APIs per l'accesso ai dati.

INTEROPERABLE

- ✓ Supporto per metadati disciplinari.
- ✓ Integrabile con vocabolari esterni.

REUSABLE

- ✓ Supporto per diversi tipi di licenze.
- ✓ Possibilità di implementare *guestbook* e *restricted datasets*.

<https://dataverse.iit.it>

Valentina Pasquale – Istituto Italiano di Tecnologia



ConfGARR23
SAPERI INTERCONNESSI

Configurazione hw e sw

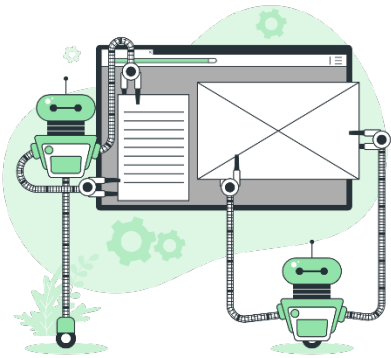
3 virtual machines VMware vCenter

12 CPU

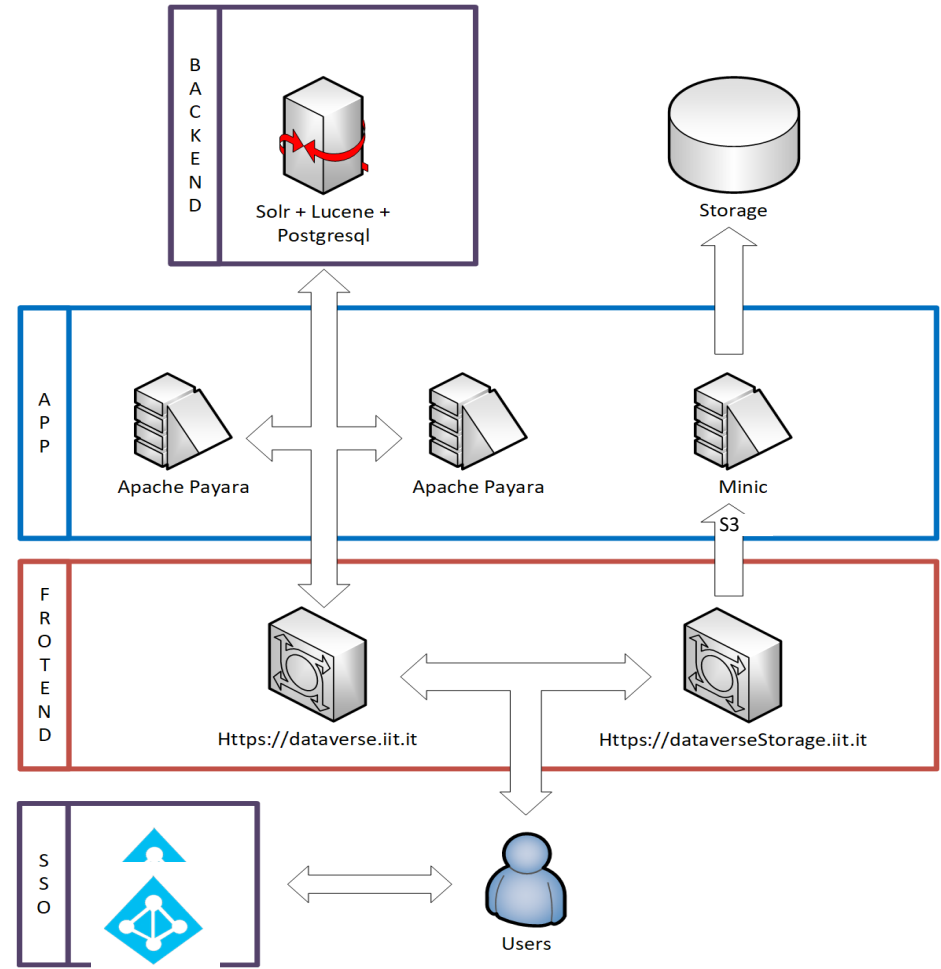
40 GB RAM

200 GB disk

10 TB storage dati

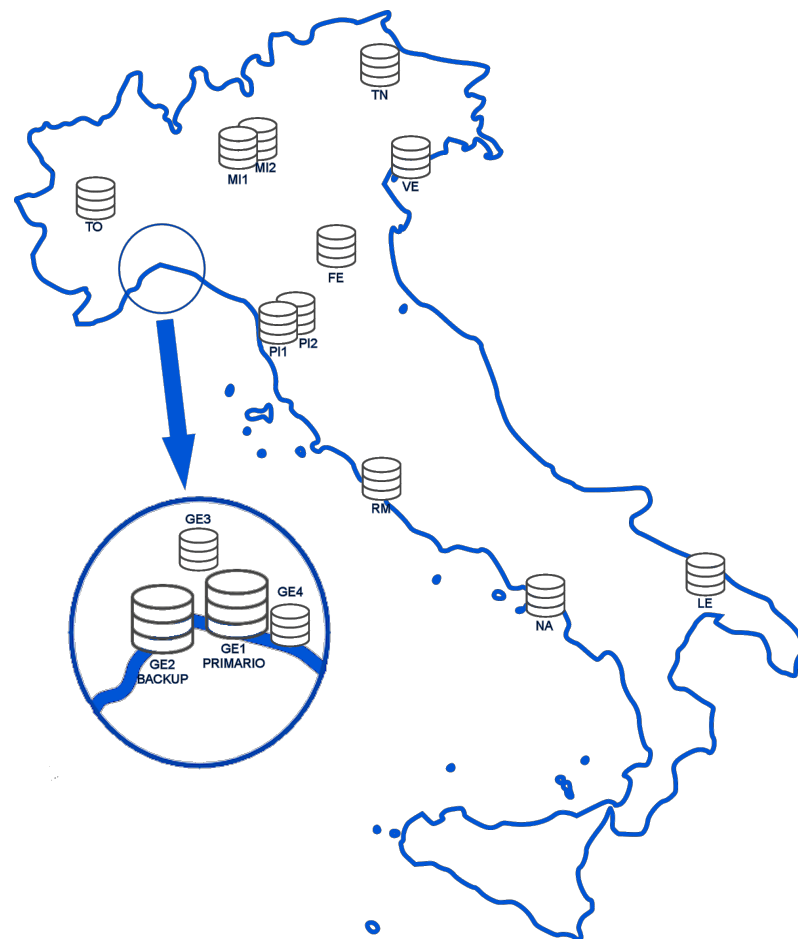


Al primo login, ogni ricercatore trova il suo profilo utente configurato con i permessi necessari per accedere ai *dataverses* creati per la sua Linea di Ricerca.

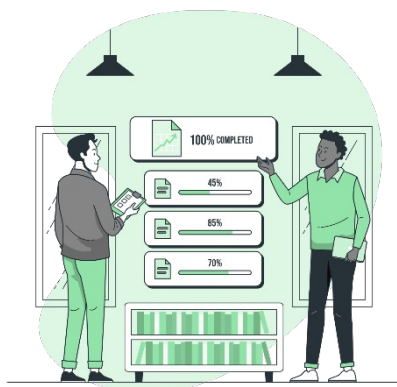


Nuova infrastruttura storage per i dati

- Installazione in corso
- Sistema on prem: velocità, prossimità agli strumenti
- Dimensioni totali 5 PB + backup
- Distribuzione sul territorio: sistemi primari presso la sede di Genova, ogni centro della rete IIT ha un sistema locale in media da ~40 TB
- Il sito primario riceve repliche selezionate dei dati provenienti dalle sedi periferiche
- Il sito primario replica sul sito di backup
- Sistema di gestione dei metadati integrato nei sistemi di storage
- Politiche di sincronizzazione configurabili in base al singolo sito, al singolo set di dati, ai metadati
- IIT Dataverse ospitato su sito primario



Sviluppi futuri



- Migrazione di IIT Dataverse sulla nuova infrastruttura *storage* al fine di incrementare lo spazio disponibile.
- Miglioramento della gestione dei dataset di dimensioni maggiori di 5 GB in *upload* e *download*.
- Abilitazione dei *restricted* dataset e aggiornamento dei termini di utilizzo.

Inoltre, è attualmente allo studio una soluzione infrastrutturale che abiliti la gestione *seamless* di dati *FAIR-by-design* dalla produzione all'analisi, alla condivisione e al riutilizzo.



Grazie per l'attenzione!

Research Data Management @ IIT
rdm@iit.it; valentina.pasquale@iit.it



ConfGARR23
SAPERI INTERCONNESSI