

# Il progetto RDM in IIT: infrastrutture istituzionali per i dati della ricerca

**Stefano Bencetti**

Direttore ICT

Francesca Cagnoni, Valentina Pasquale, Ugo Moschini, Elisa Molinari

**Research Organization Directorate**

Alessandro Bruchi, Felipe Ferreira, Daniele Rossetto

**ICT Directorate**

[rdm@iit.it](mailto:rdm@iit.it)

GenOA Week 2021, 27 ottobre 2021

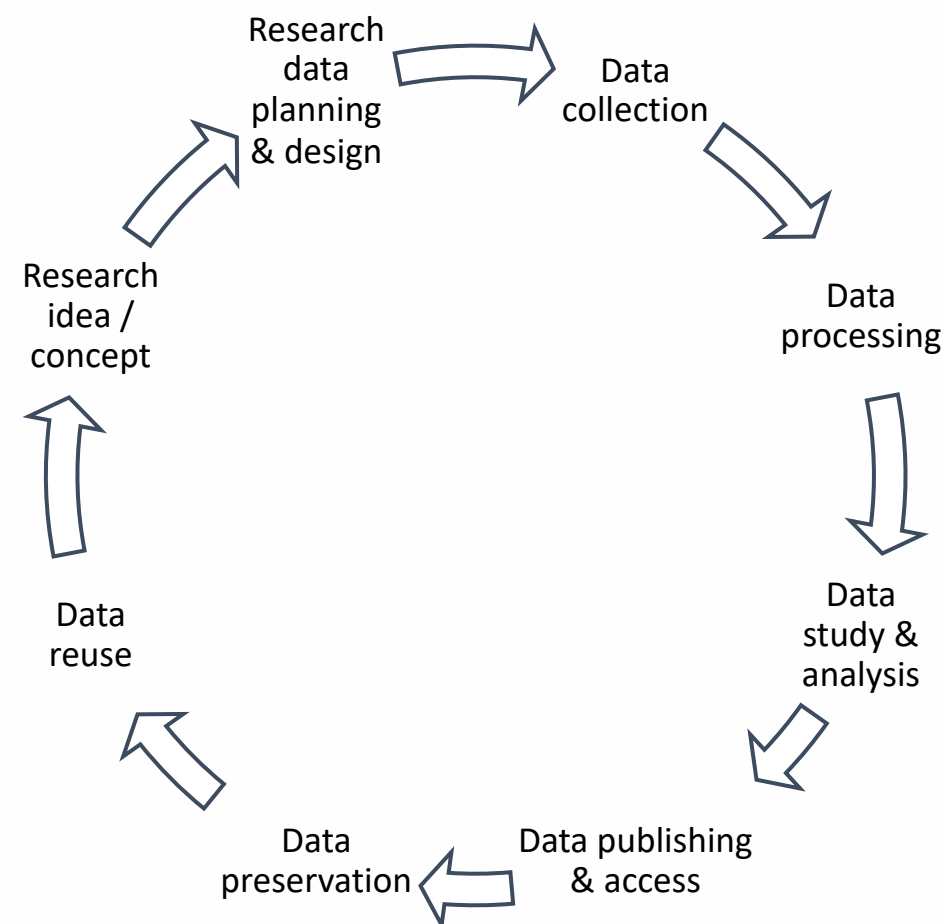
# Obiettivo del progetto RDM

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.

Findable  
Accessible  
Interoperable  
Reusable

Fully  
AI  
Ready

## Ciclo dei dati della ricerca



### Data management planning

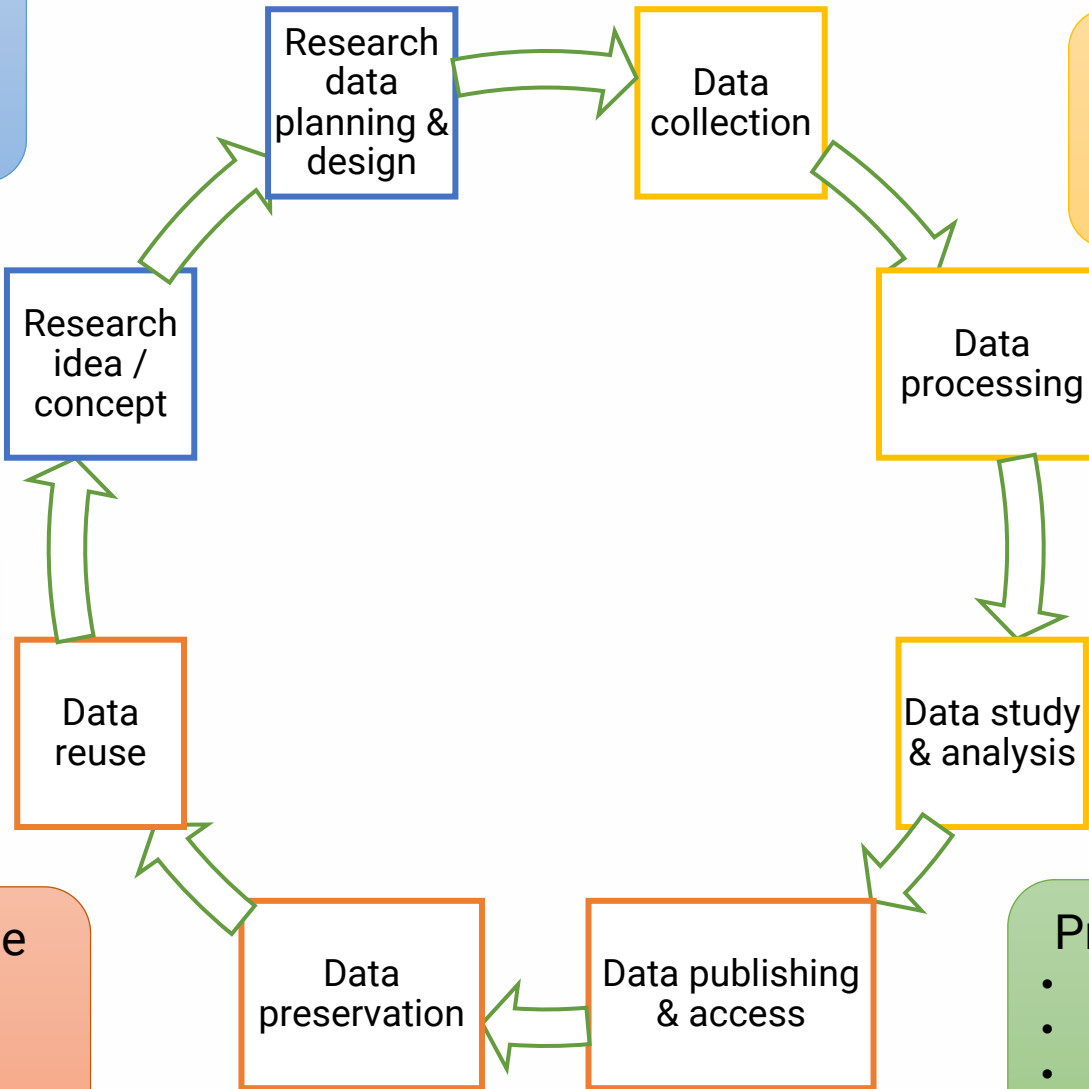
- Strumenti di collaborazione
- Protezione dei dati / privacy
- Supporto alla scrittura dei DMP

### Supporto per i dati attivi

- Infrastruttura istituzionale storage
- Cyber-security policy

# RDM@IIT

Il progetto **RDM@IIT** è una collaborazione delle Direzioni di **Research Organization** e **ICT**, con il contributo di **Affari Legali**.



### Repository istituzionale

- Conservazione dei dati a lungo termine
- Condivisione aperta
- Data curation

### Promuovere Open Science

- Linee guida istituzionali
- Webinars
- Formazione
- Eventi

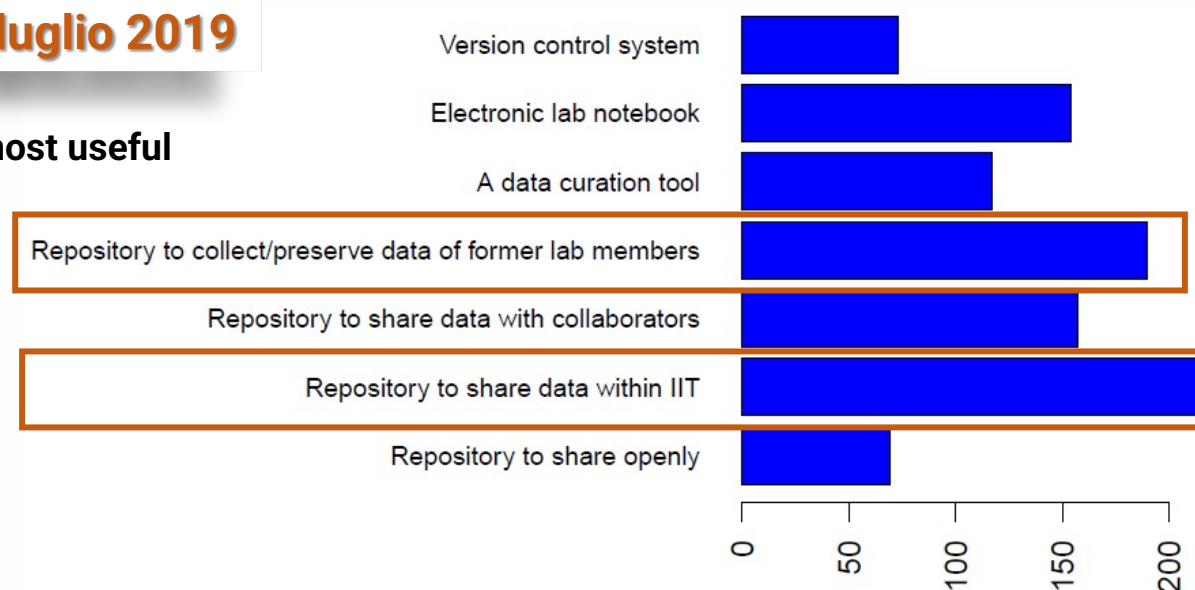
# Il punto di partenza: le esigenze dei ricercatori



## Dal sondaggio proposto ai ricercatori IIT a **luglio 2019**

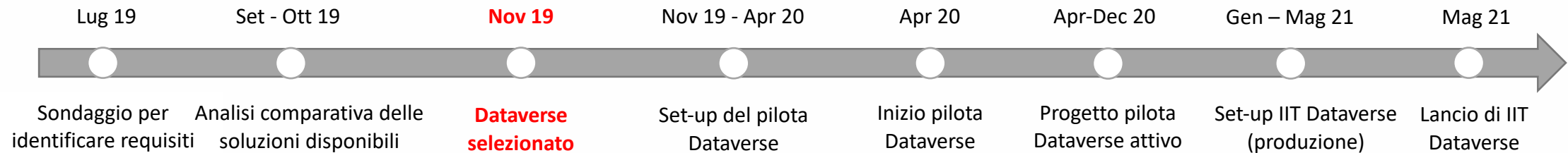
**Which RDM tool do you think it would be most useful for your research activity?**

##	answer	count
## 1	Repository to share openly	69
## 2	Repository to share data within IIT	217
## 3	Repository to share data with collaborators	157
## 4	Repository to collect/preserve data of former lab members	190
## 5	A data curation tool	117
## 6	Electronic lab notebook	154
## 7	Version control system	73



# Dataverse

## repository open source per i dati della ricerca



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

- Facilmente configurabile e integrabile con i nostri sistemi
- 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



# Progetto pilota di Dataverse (2020)



**iit** ISTITUTO ITALIANO DI TECNOLOGIA

Search ▾ User Guide Support Log In

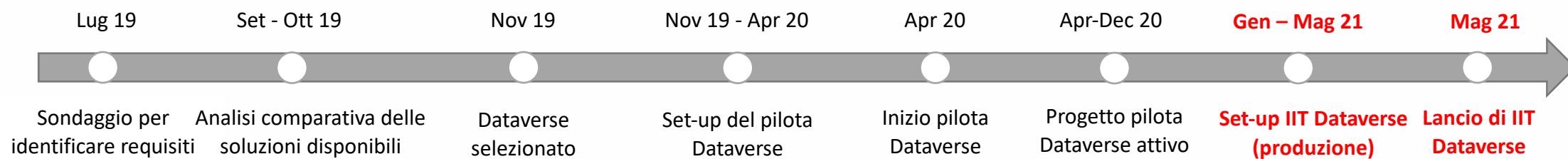
**IIT Dataverse** (Istituto Italiano di Tecnologia)

Metrics 0 Downloads

The root dataverse.

Nel 2020, abbiamo avviato un **progetto pilota** con alcune Linee di Ricerca, selezionate nei diversi Domini di Ricerca di IIT, per testare Dataverse e raccogliere il *feedback* dei ricercatori.

# Maggio 2021: lancio di IIT Dataverse



**Why use IIT Dataverse?**

The infographic depicts a conversation between two scientists. On the left, a scientist says: "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!". A response box says: "Preserve research datasets in the long-term".

In the middle, a scientist says: "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!". A response box says: "Document datasets with metadata".

On the right, a scientist says: "For my ERC I have to preserve and possibly share data making them FAIR.". A response box says: "Make data FAIR and open when possible".

Another scientist says: "Do you know we have an institutional repository for research data? It can be of help!". A response box says: "That's why documentation and metadata are important!!".

At the bottom, a scientist says: "After a while...". A response box says: "That's why documentation and metadata are important!!".

The infographic includes the logos for The Dataverse Project and IIT (Istituto Italiano di Tecnologia) and is licensed under CC BY 4.0.

- ✓ **19 maggio 2021:** evento interno di presentazione di IIT Dataverse.
- ✓ Rilascio di linee guida sulla nostra Intranet.
- ✓ Realizzazione di un servizio di supporto per la creazione di *dataset*.



<https://dataverse.iit.it>

Metrics 1 Download

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 IIT Dataverse.

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 resources

Search this dataverse...  [Advanced Search](#)

### FINDABLE

- ✓ Integrato con DataCite per l'assegnazione di DOIs.
- ✓ Censito in [www.re3data.org](http://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*.

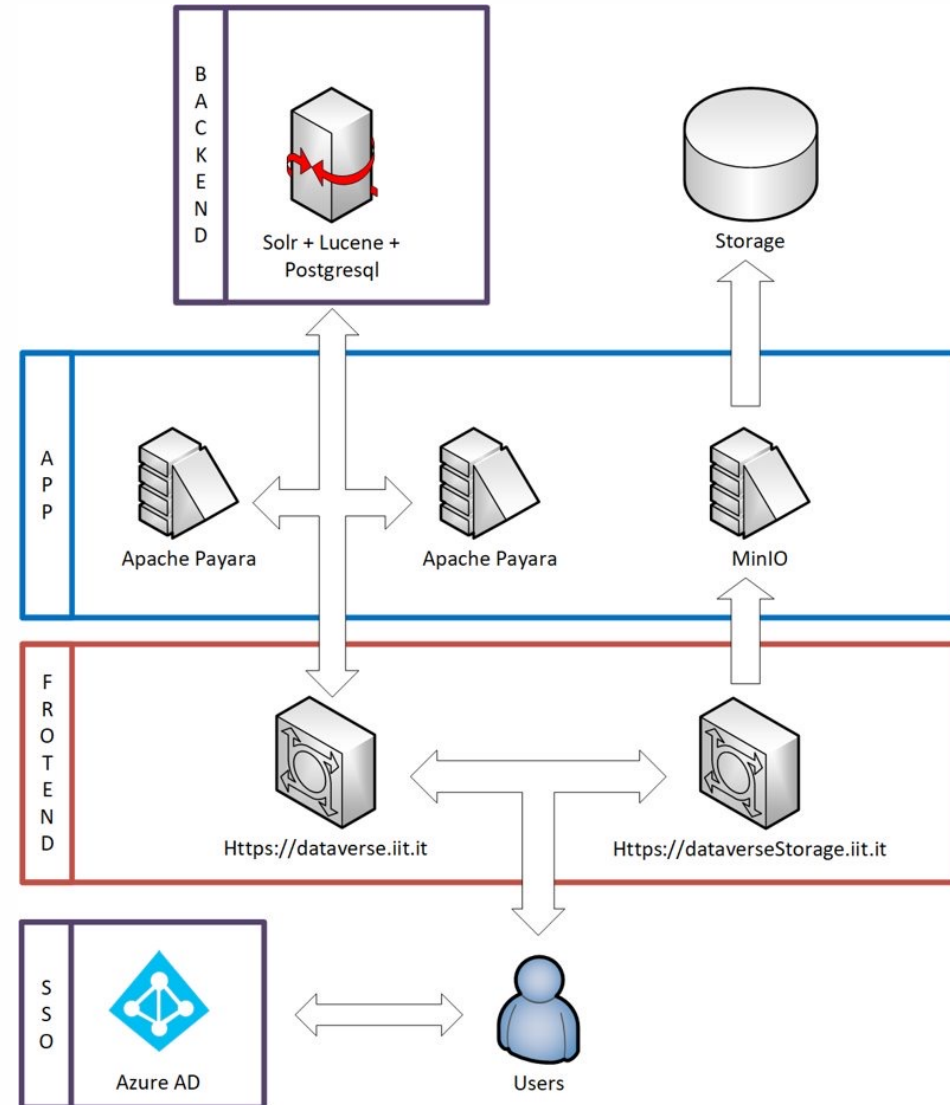




# Dataverse Infrastructure

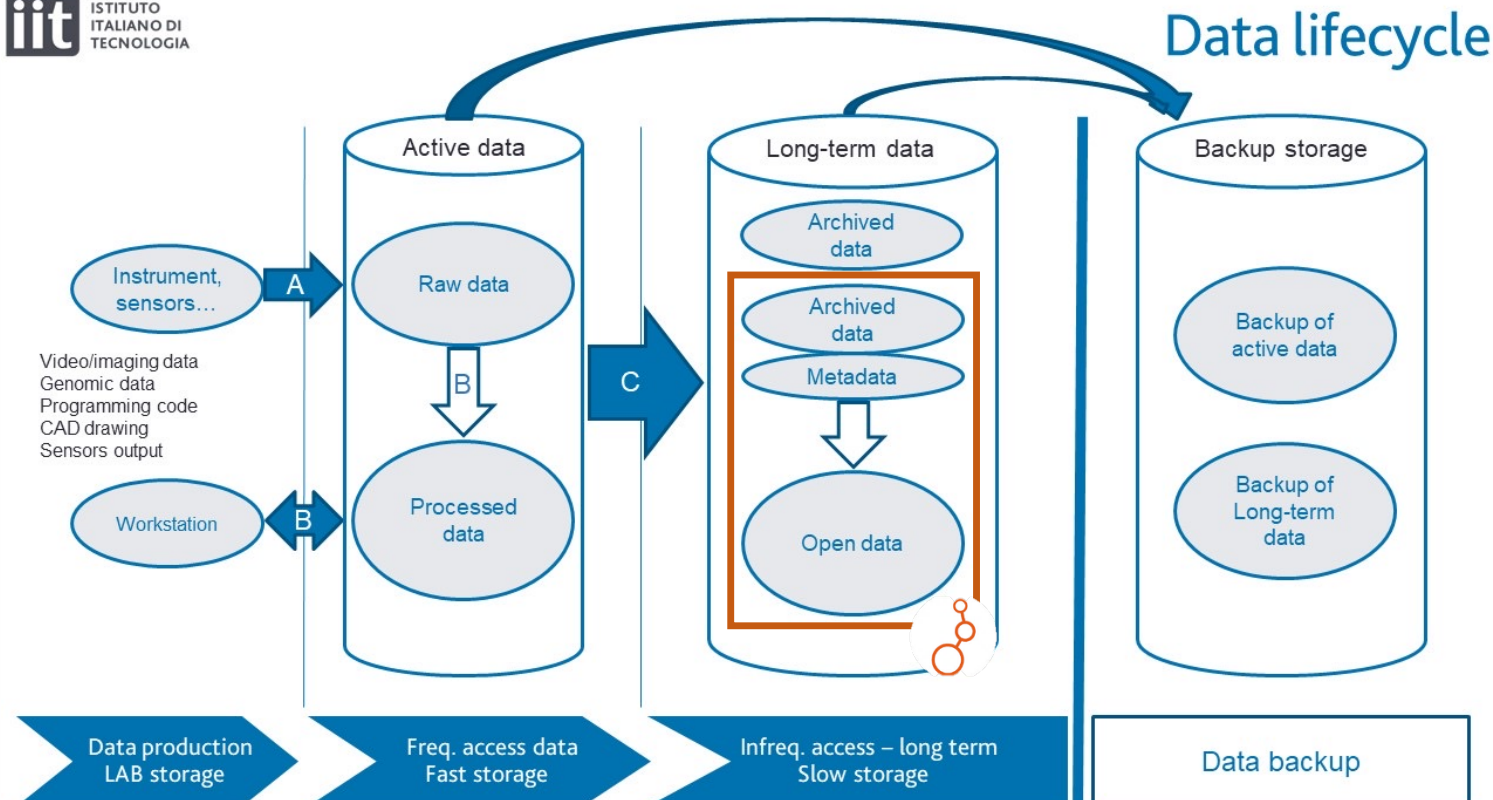
## INFRASTRUCTURE

Architettura scalabile  
4vm on vmware vCenter  
12 CPU  
40 GB RAM  
200GB disk  
1 Storage



# Quali dati in IIT Dataverse?

IIT Dataverse è primariamente un **repository interno per dataset finalizzati**.

ISTITUTO  
ITALIANO DI  
TECNOLOGIA

I ricercatori IIT possono depositare in IIT Dataverse una volta che hanno terminato il loro progetto di ricerca e opzionalmente possono decidere di pubblicarli in **accesso aperto**, con l'assegnazione di un DOI persistente.

# Infrastruttura storage

- Gara in corso
- Sistema on prem: velocità, prossimità agli strumenti
- Dimensioni totali >5PB
- Distribuzione sul territorio: sistemi primario sulla sede di Genova, ogni centro della rete IIT ha un sistema locale da ~40TB
- 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
- Dataverse ospitato su sito primario



# Policy di sicurezza delle informazioni: classificazione rischio sul dato

## Dati ad alto rischio

### RISCHIO:

la perdita di Riservatezza, Integrità o Disponibilità delle informazioni può avere impatti significativi sulla Proprietà Intellettuale o sulla missione o reputazione della Fondazione, o significativi impatti economici o legali

### ESEMPI:

Informazioni militari, dati genomici, dati sensibili, dati di ricerca ecc

## Dati a medio rischio

### RISCHIO:

la perdita di Riservatezza, Integrità o Disponibilità delle informazioni può avere impatti limitati sulla Proprietà Intellettuale o sulla missione o reputazione della Fondazione, o limitati impatti economici o legali.

### ESEMPI:

Dati di ricerca, dati personali ecc

## Dati a basso rischio

### RISCHIO:

la perdita di Riservatezza, Integrità e Disponibilità delle informazioni non ha impatti apprezzabili sulla Proprietà Intellettuale, sulla missione e sulla reputazione della Fondazione, e non ha apprezzabili impatti economici né legali.

### ESEMPI:

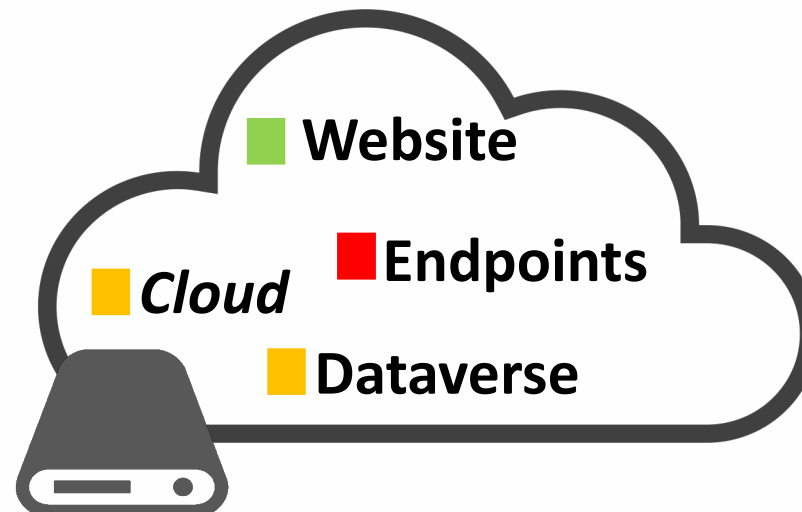
Dati pubblici, open data, dati di ricerca ecc

# Policy di sicurezza delle informazioni: data management

## Data Risk Classification



## Misure di sicurezza



# Le sfide del prossimo futuro

- Offrire un supporto sempre più specifico per le diverse discipline presenti in IIT
  - Introduzione schemi metadati disciplinari (Dataverse 5.7, integrazione con [SKOSMOS](#))
- Knowledge-base MS, derivare conoscenza grazie alle informazioni (dati e metadati)
- Sviluppo e mantenimento infrastruttura di storage, integrazione dei metadati
- Integrare IIT Dataverse a livello europeo, con i servizi e le infrastrutture della European Open Science Cloud
  - Integrazione con OpenAIRE
- Incrementare il servizio di supporto con servizi di *data stewardship*