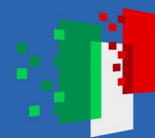




Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



# Il datalake nell'infrastruttura INFN

## Applicazioni e prospettive future

A. Troja, M. Biasotto, D. Ciangottini, M. Delli Veneri, F. Fanzago, A. Italiano, N. Marcelli, L. Morganti, A. Rendina, M. Sgaravatto, D. Spiga, B. Spisso, S. Stalio, M. Verlato

21/05/2024

Workshop sul calcolo nell'INFN - Palau

## Dove eravamo rimasti

Negli ultimi anni, si è molto parlato della federazione degli storage distribuiti nel territorio nazionale.

Nell'ambito del WP6 di DataCloud, sono state fatte diverse sperimentazioni per federare storage distribuiti ed eterogenei.

Il risultato è stato l'implementazione di testbed di datalake, nei quali abbiamo testato e validato diverse configurazioni

La strada fin qui:

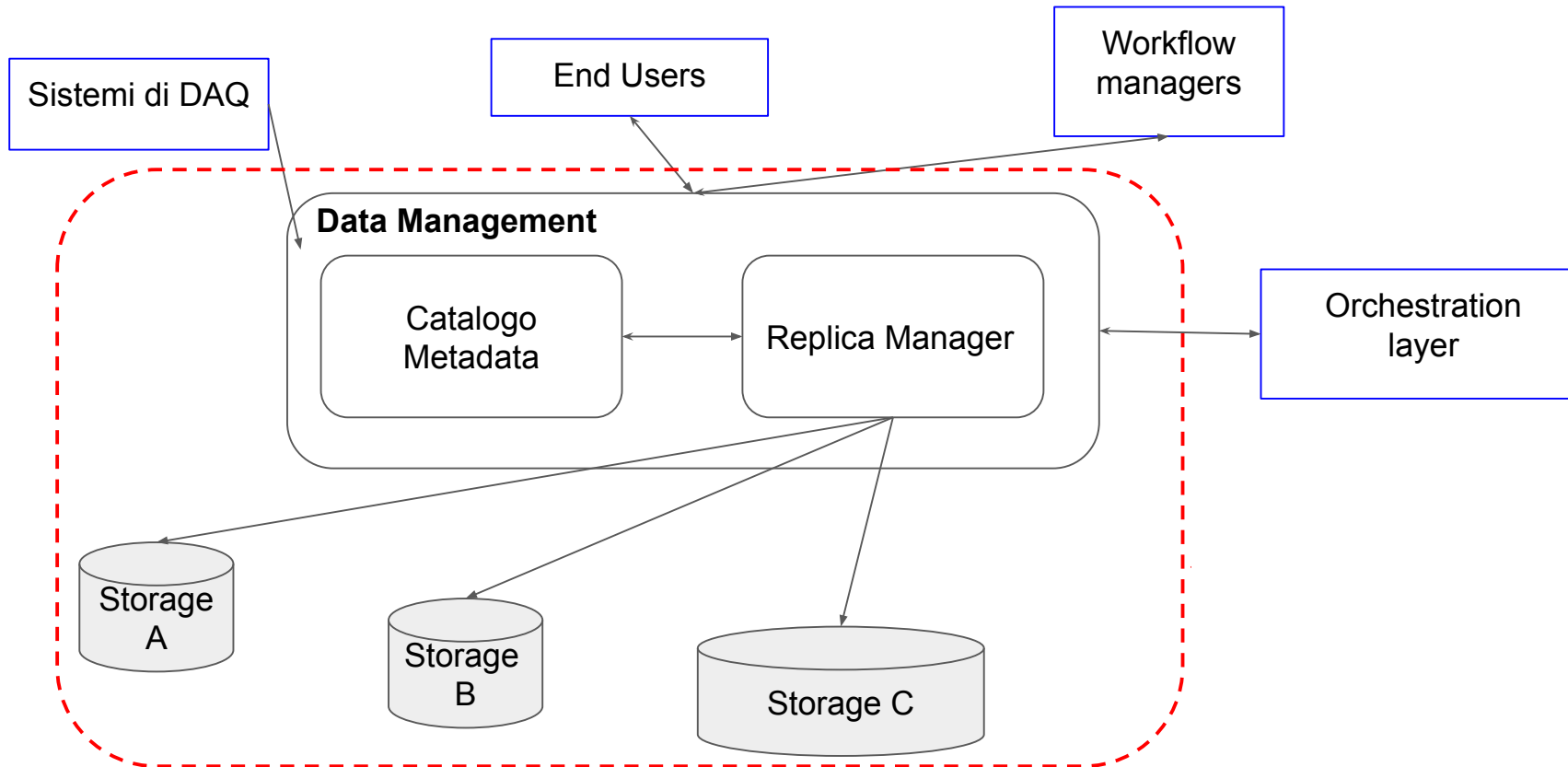
[Storage & Data Magement](#), M. Sgaravatto e D. Spiga, 2022

[Federare lo storage distribuito nazionale](#), D. Ciangottini, 2023

[Attività di R&D in corso sinergiche con i vari progetti e nell'ambito di Datacloud](#), M. Sgaravatto 2024



# Recap: Cos'è il datalake?



[Storage & Data Magement](#), M. Sgaravatto e D. Spiga, 2022

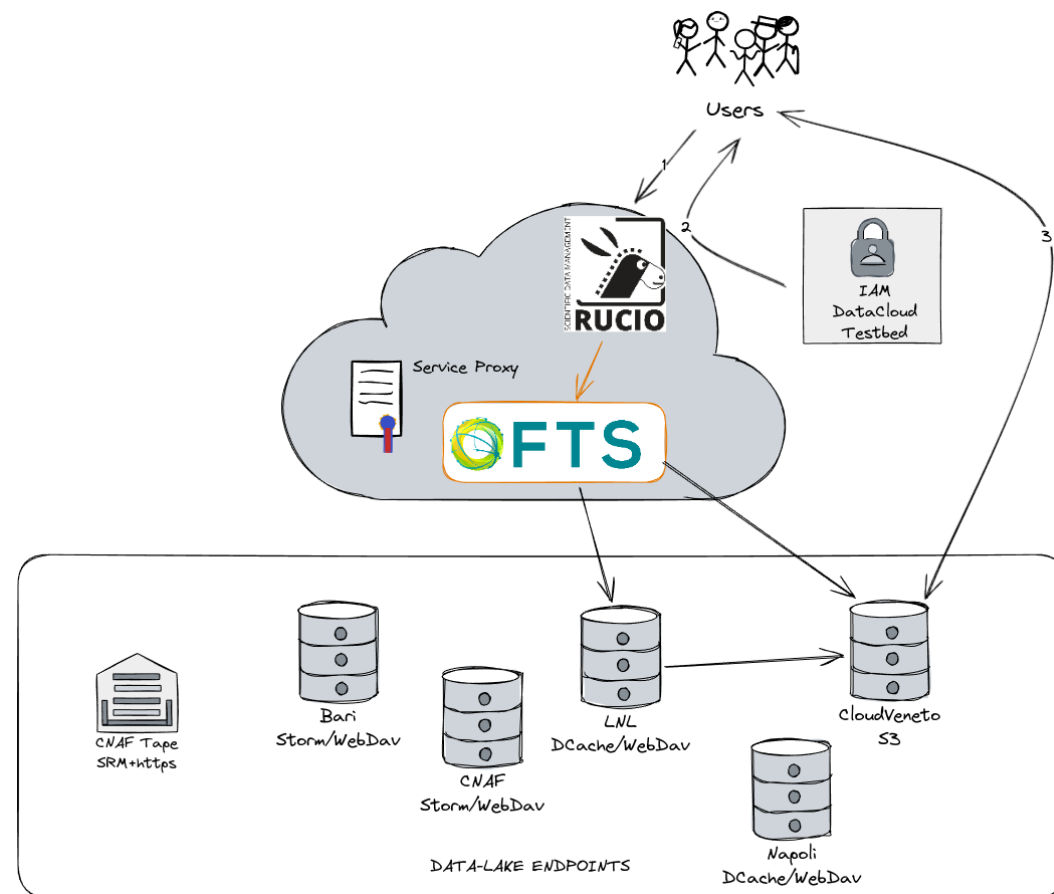
## Il testbed in DataCloud WP6

Abbiamo scelto di avviare la sperimentazione integrando strumenti già in uso in realtà scientifiche a noi vicine (es. LHC):

- Rucio+FTS (replica manager)
- IAM (AuthN/Z)
- Metadata Catalog: embedded in RUCIO

Federati 6 storage system dell'INFN eterogenei:

- Qos (disk, tape);
- Implementazione (dCache, Storm, etc...).



[Federare lo storage distribuito nazionale](#), D. Ciangottini, 2023

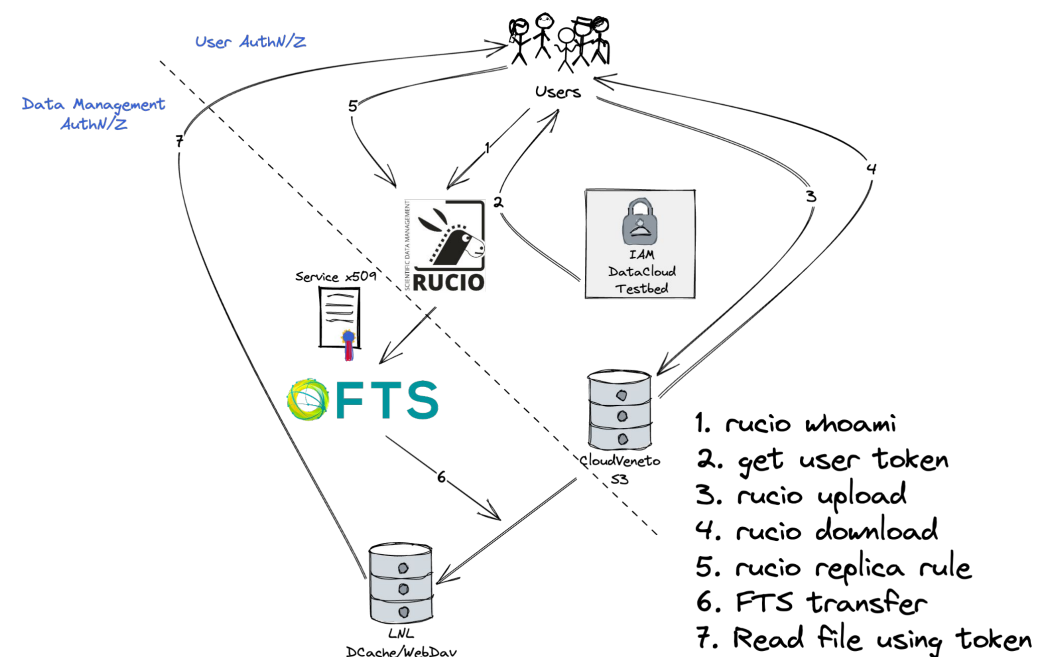
## Modello Autorizzativo

- L'utente interagisce con RUCIO via IAM token;
- Third Part Copy (TPC) con FTS;
  - AuthZ con VOMS proxy di servizio.

Testati diversi modelli autorizzativi sui dati:

- Group based polices;
- Scope based policies definite nell'istanza IAM;
  - Es. isolamento degli utenti  
(es. ogni utente può leggere tutto  
ma scrivere solo nella sua area).

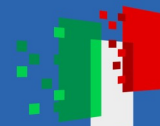
Rucio permette l'implementazione di policy altamente customizzabili. Nel corso degli ultimi mesi abbiamo acquisito molta esperienza a proposito.



[Federare lo storage distribuito nazionale](#), D. Ciangottini, 2023

## Cosa offrono questi strumenti

- **Storage** differenti vengono federati indipendentemente dalla collocazione geografica, dalla loro implementazione, dal loro QoS (disk o tape);
- **L'utente** interagisce coi dati in maniera dichiarativa: ad esempio può dichiarare quante repliche servono per un certo file, su quanti e quali storage, per quanto tempo devono esistere;
- **I dati** possono essere organizzati gerarchicamente (dataset, container);
- Grazie alla gestione dei **metadati**, possiamo implementare funzioni di query;
- **Trasparenza**: Livello di astrazione che nasconde all'utente le eterogeneità e i dettagli implementativi.

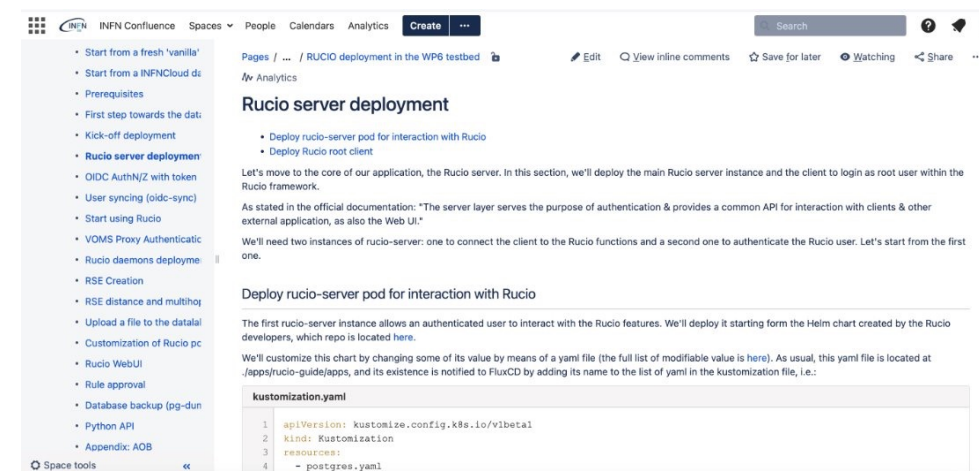
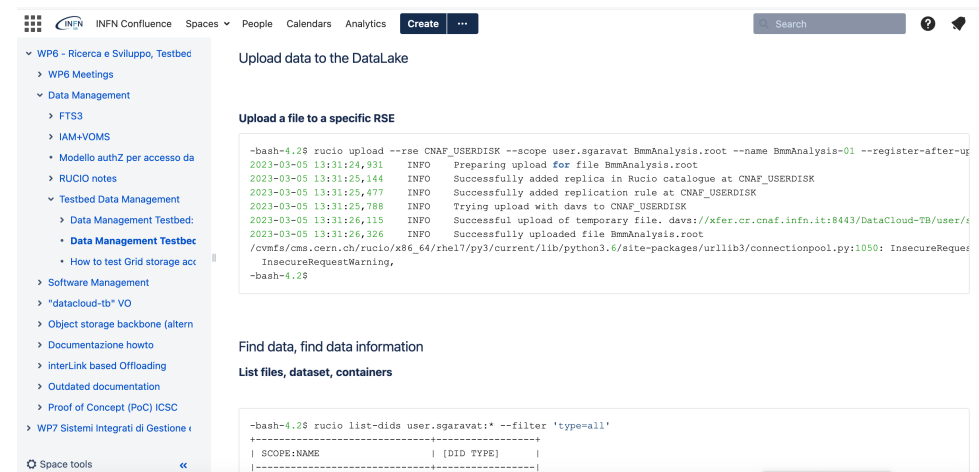


# Documentazione e competenze

Prodotta parecchia documentazione sia per utenti che per operatori (la documentazione ufficiale è piuttosto limitata).

Il numero di persone che hanno iniziato a prendere confidenza con gli strumenti del datalake è aumentato negli ultimi mesi.

Questo evita di concentrare le competenze in poche persone ed il «single point of failure».



## Case study: l'onboarding di Cygno

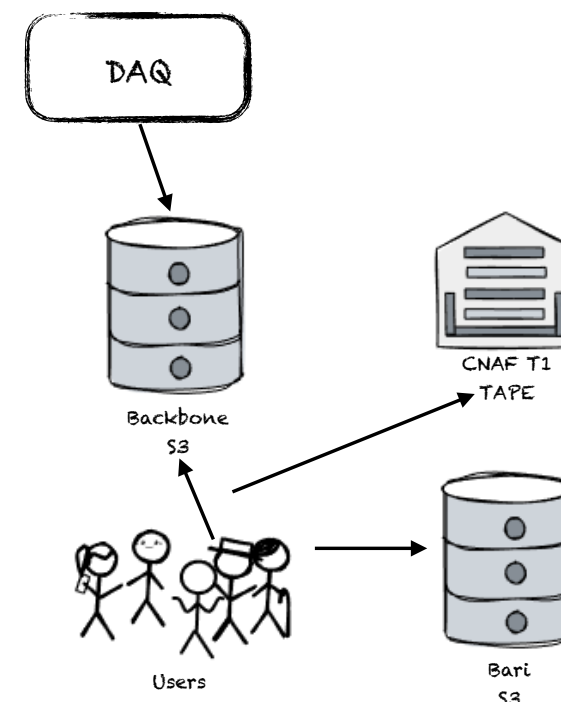
Cygno è un esperimento dell'INFN sulla materia oscura in fase preliminare, ma produce già dati.

I dati vengono caricati prima su uno storage disco S3 sulla Backbone INFNCloud. Da qui, vengono archiviati manualmente su uno storage tape al T1.

Un terzo storage disk (S3 a Bari) è appena stato messo a disposizione dell'esperimento.

Il ciclo vita dei dati prevede che vengano rimossi dai dischi e mantenuti su tape dopo il processamento. Ri-copia su disco in caso di riprocessamento.

Il percorso seguito dai dati di Cygno rappresenta il campo di utilizzo perfetto per il DM.



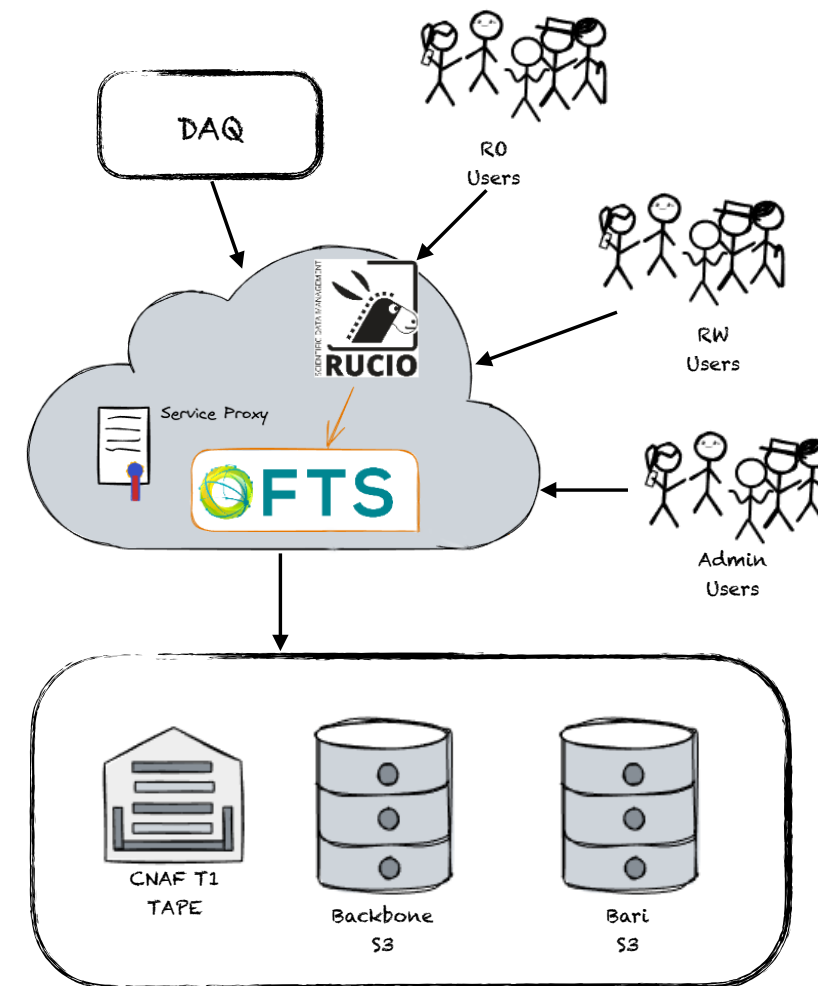


## Case study: l'onboarding di Cygno

Abbiamo quindi sviluppato un datalake che:

- Federa i due storage disk e lo storage tape;
- Implementa delle policy ad hoc (più stringenti di quelle del testbed WP6),
  - In particolare introducendo un layer di autorizzazione su regole specifiche (transfer da/a tape)
    - Tramite una WebUI la persona adeguatamente autorizzata può approvare o rigettare richieste di replica

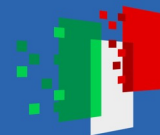
*As we speak*, gli utenti Cygno stanno già registrando in Rucio i dati già archiviati negli storage, mentre i nuovi dati vengono direttamente archiviati tramite gli strumenti del DM.



## Onboarding di altre comunità'

Siamo in contatto con altre comunità che hanno espresso interesse a provare questi tool

- [DarkSide](#): stanno provvedendo (a meno di alcune issue procedurali) a mettere in piedi una istanza RUCIO attraverso la documentazione da noi fornita;
- [DAMPE](#): L'interesse della comunità è sincronizzare i loro dati tra il T1 dell'INFN e un data center in Cina.



## Modelli di deployment

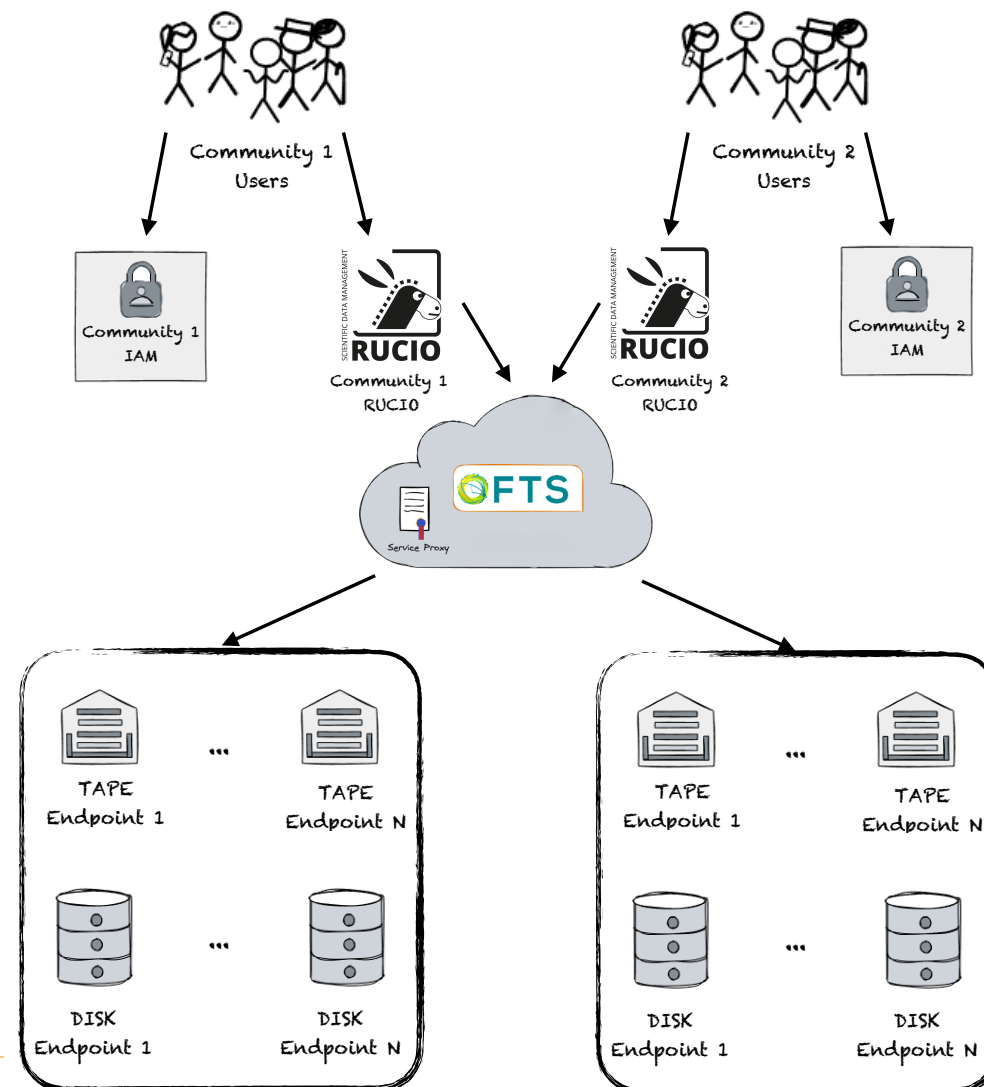
Pensiamo debbano essere previsti due modelli di deployment di un sistema di data management:

- DM di comunità;
- DM «catch-all» nazionale.

## DM di comunità

Ad uso di una comunità di utenti

- Istanza RUCIO di comunità, gestita da essa;
- Integrato con uno IAM che gestisce gli utenti e le policies della comunità;
- FTS multi-VO (gestito centralmente);
- Integrazione degli storage *pledged* della comunità.

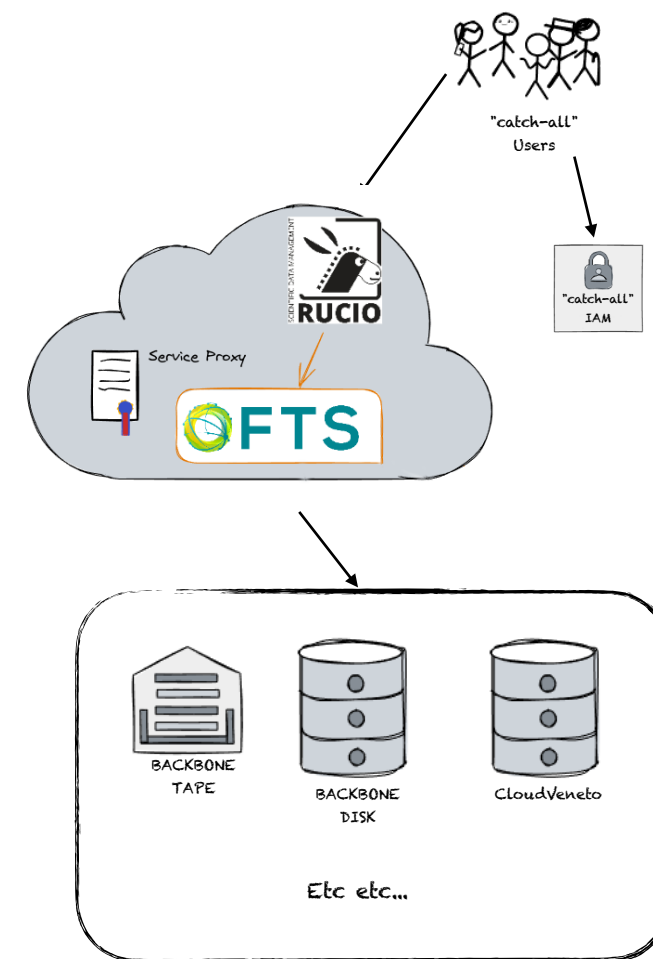


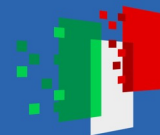
## DM «catch-all»

Istanza di DM messa a disposizione di utenti «singoli».

Generalizzando il caso precedente, permette ad un utente singolo di utilizzare il DM anche per dati «personali», con una singola interfaccia a fronte di diversi backend storage --> semplificazione.

- Istanza di RUCIO gestita centralmente;
- IAM «catch-all» gestito centralmente;
- Uso dell'FTS multi-VO gestito centralmente;
- Federazione degli storage dei vari centri di calcolo dell'INFN.



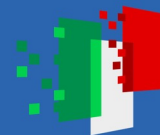


## Verso la produzione

Altri step sono stati compiuti:

- Istanza FTS in high-availability, multi-VO, messa in produzione dal team di operations;
- Infrastruttura PSQL in *high availability* dove ospitare i database RUCIO.

Pensiamo che possa ora partire una sperimentazione per il deployment di una istanza DM «catch-all» nazionale



## Osservazioni finali

- Il sistema presentato è uno dei quattro elementi portanti dell'infrastruttura datalake di ICSC/Terabit;
- È anche elemento del PoC con CINECA, per il quale abbiamo i servizi DM pronti e 5 endpoint INFN federati, ai quali verrà aggiunto un endpoint CINECA;
- Inoltre, è un elemento portante delle attività di Spoke2, in particolare il progetto IDL con Leonardo (innovation grant).

## Alcuni prossimi passi

- Onboarding di altre comunità interessate e ricerca di utenti interessati «personalmente»;
- DM nazionale «catch-all»
- Integrazione delle evoluzioni architetturali:
  - Storm REST API in sostituzione di SRM anche per il tape;
  - Uso di token al posto di VOMS proxy anche per i TPC fatti da FTS;
  - Etc ...
- Integrazione con *metadata catalog* esterni
  - Sinergie con ICSC spoke3 (V. il poster di D. Ciangottini et. al: «[Casi studio ICSC Spoke3 e sinergie con il calcolo dell'INFN](#)»).

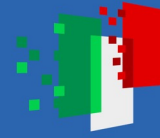




Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Grazie per l'attenzione