

[Dataverse@unimi.it](https://dataverse.unimi.it): un percorso di crescita graduale e costante

Il progetto Research Data Management @unimi ([https://dataverse.unimi.it/](https://dataverse.unimi.it)) è nato nel 2018 per rispondere alle richieste di trasparenza dei finanziatori e a quelle di riproducibilità o replicabilità delle ricerche.

In quel momento non esisteva un servizio centralizzato per il supporto alla gestione dei dati FAIR, né strumenti che ne permettessero la gestione. Ciò significa che ciascun ricercatore o gruppo di ricerca cercava di provvedere in autonomia e secondo il buon senso alla gestione, archiviazione e conservazione dei dati.

Dalle interviste fatte nei dipartimenti è però risultato subito chiaro che c'era una esigenza forte di strumenti di archiviazione e di formazione alla gestione dei dati in maniera FAIR e alla gestione dei diritti (soprattutto nelle collaborazioni con terzi).

Il primo passo è stato quindi quello di individuare uno strumento che potesse essere adeguato per tutte le aree disciplinari. Lo strumento doveva coprire quelle aree che non avevano repository di riferimento e assicurare una gestione dei dati in modalità FAIR.

Dopo varie sperimentazioni, la scelta è caduta su Dataverse, un software open source, sviluppato dall'università di Harvard, molto utilizzato da sistemi della ricerca nazionali (ad esempio in Danimarca e in Olanda).

Il software è stato implementato con il supporto di 4Science che gestisce l'hosting e i processi di manutenzione e aggiornamento.

Contemporaneamente alla analisi del software veniva approvata la policy su RDM che chiedeva ai ricercatori destinatari di finanziamenti una gestione FAIR dei dati.

Anche il contesto editoriale poneva richieste sempre più precise ai ricercatori rispetto ai dati alla base dei risultati esposti negli articoli.

Qualsiasi strumento che si intenda implementare, ancorché ottimo per le funzionalità che offre, serve però a poco se non viene portato all'attenzione dei ricercatori e se non viene dimostrato in che modo lo strumento può essere loro utile.

Su questo punto l'Ateneo e la Direzione che gestisce i processi di Open Science si sono impegnati moltissimo attraverso una serie di azioni:

- Predisposizione di istruzioni e linee guida per il deposito dei dati.
- Creazione di un [sito](#) dedicato a RDM che, oltre a fornire le istruzioni per l'uso del repository, offre anche una serie di informazioni sul tema della gestione dei dati, degli strumenti, dei servizi a disposizione.
- Incontri informativi mensili su RDM, Dati FAIR e uso di Dataverse.
- Corsi specifici di competenze trasversali (obbligatorie) per gli studenti di dottorato, ripetuti ad ogni ciclo.
- Presentazioni su richiesta ai Dipartimenti e ad aree specifiche.
- Un progetto pilota per i dottorandi dedicato alla stesura del DMP (Data Management Plan) per il loro progetto di ricerca.

- Progetti di formazione ad hoc dedicati a gruppi di ricerca internazionali a guida UniMI.
- Supporto ai ricercatori che hanno necessità di archiviare i propri dati in maniera FAIR.
- Report di monitoraggio sull'uso di Dataverse (chi - cosa - come) comprensivo dei moltissimi download.
- Reclutamento di due (presto tre) curatori dei dati (data stewards) a supporto del progetto RDM.

Particolare attenzione è stata dedicata alla implementazione del software: in primo luogo si è posta attenzione alla connessione con gli identificativi persistenti (ORCID, DOI), poi si è cercato di facilitare l'accesso dei nostri utenti permettendo la registrazione al sistema sia attraverso il collegamento con il sistema di autenticazione di ateneo sia attraverso ORCID.

Il workflow di caricamento dei dati in Dataverse, descritto in una guida lunga e una breve, viene periodicamente rivisto, sulla base dei suggerimenti degli utenti e anche per cercare di rendere il processo il più robusto possibile.

Le due data steward hanno lavorato alla qualità di dati e metadati del repository. L'arrivo di due figure che sono state formate sul tema dei dati della ricerca e della loro gestione ha permesso di implementare un workflow di validazione dei dataset sia già pubblicati che in corso di pubblicazione e ciò ha facilitato il processo di certificazione (Core Trust Seal) che è in corso.

Il servizio non è stato per nulla veloce né facile da implementare, anche perché la tendenza di chi produce dataset è quella di tralasciare tutte le attività che non sono obbligatorie a causa dell'intensità dei processi della ricerca. Tuttavia, con la crescita delle richieste da parte degli enti finanziatori e delle riviste anche l'uso di Dataverse ha cominciato a diffondersi.

Le attività di disseminazione sopra descritte tuttavia, continuano sempre con la stessa intensità. Il servizio per poter funzionare deve avere personale dedicato alla attività di supporto alla gestione dei dati e alla promozione del servizio stesso fra i ricercatori dell'ateneo.

Dataverse è uno strumento che, se gestito con cura, permette una capillare disseminazione dei dati. Lo dimostrano le statistiche: a metà dello scorso anno i download dell'archivio erano qualche migliaio, mentre a un anno di distanza abbiamo quasi raggiunto quota 300 mila download.