

Adattamento da: "Core Requirements for Data Management Plan", in Science Europe. (2021). Practical Guide to the International Alignment of Research Data Management - Extended Edition. <https://doi.org/10.5281/zenodo.4915862>
Creative Commons Attribution 4.0 International

Uno dei pilastri della ricerca scientifica riproducibile è l'esposizione pubblica dei dati. Gli Enti finanziatori, le istituzioni, gli editori chiedono ai ricercatori di depositare i dati della ricerca in repositories dedicati. I ricercatori, quindi, hanno l'esigenza di identificare dei repository affidabili, che consentano loro di depositare i dati e renderli disponibili alla comunità scientifica e alla società.

La maggior parte dei repositories non è certificata da un ente di certificazione accreditato, per cui identificarne uno appropriato è una grande sfida. Alcuni repositories potranno offrire policy e standard specifiche per specifiche comunità scientifiche, altri potranno essere più generalisti.

E' raccomandato ai ricercatori di cercare in prima battuta tra i repositories che sono ampiamente utilizzati e riconosciuti nello specifico ambito disciplinare o tra i repositories certificati.

Di seguito i criteri di selezione raccomandati nel caso in cui i due suggerimenti precedenti non avessero consentito di trovare un repository per depositare i dati.

Identificativi unici persistenti (PIDs)



Consentono la scoperta e l'identificazione dei dati, la loro ricercabilità, citabilità e interrogazione. Supportano la gestione delle versioni.

Metadati



Permettono di trovare i dati. Consentono la referenziazione di informazioni rilevanti collegate, come altri dati o pubblicazioni. Forniscono informazioni pubblicamente disponibili anche per dati che non lo sono. Utilizzano standard di metadatezione largamente accettati dalla comunità scientifica. Assicurano che i metadati siano machine-readable.

Accesso ai dati e licenze d'uso



Consentono l'accesso ai dati sotto condizioni ben specificate. Assicurano l'autenticità e integrità dei dati. Forniscono informazioni, anche machine-readable, sui permessi di utilizzo. Assicurano la riservatezza e il rispetto dei diritti degli interessati e dei produttori dei dati.

Conservazione



Assicura la persistenza di dati e metadati. Nel repository devono essere dichiarati mission, ambito di utilizzo, politiche di conservazione, modalità di gestione, sostenibilità finanziaria, durata della conservazione e piano di mantenimento del repository.



Identificativi unici persistenti (PIDs)

Un repository affidabile dovrebbe:

Consentire la «scoperta» e l'identificazione dei dati, quindi **assicurati che tra i metadati siano previsti identificativi unici persistenti (PID)**. Rendere possibile **ricercare, citare e interrogare** i dati anche se fisicamente dovessero cambiare collocazione. Fornire supporto per il **data versioning**.



Metadati

I dati dovrebbero essere descritti con numerosi metadati che documentino come sono stati generati, sotto quale licenza e come potranno essere riutilizzati; i metadati devono contenere le informazioni di contesto che consentano ad altri ricercatori di interpretarli appropriatamente.

Un repository affidabile dovrebbe:

Consentire di trovare i dati, assicurando l'interoperabilità e il riutilizzo dei dati, fornendo i **dati e i metadati in un linguaggio accessibile**, basato su un ben determinato formalismo. I dati e i metadati dovrebbero essere descritti utilizzando **vocabolari controllati** e formati che consentano di cercarli, combinarli in modo automatico. Consentire la **referenziazione ad informazioni rilevanti correlate**, con un identificativo unico persistente e la relativa descrizione. **Fornire informazioni pubblicamente disponibili anche per dati non pubblici**, assicurando la conservazione a lungo termine dei metadati e la loro ricercabilità, anche se i dati non sono disponibili per svariati motivi. Utilizzare, per i metadati, **standard che siano largamente accettati dalla comunità scientifica**, assicurando che i dati mantenuti nel repository siano **machine-retrievable** (reperibili con strumenti informatici), e assicurando buone pratiche per la gestione dei dati.



Accesso ai dati e licenze d'uso

Un repository affidabile dovrebbe:

Consentire l'accesso ai dati sotto ben specificate condizioni, i cui **termini di riutilizzo dei dati devono essere descritti in modo chiaro**. Le informazioni sulla licenza devono essere incluse nei metadati. **Assicurare l'autenticità ed integrità dei dati** (informazioni dettagliate sulla provenienza dei dati, le modalità di creazione e di elaborazione, in quale contesto possono essere riutilizzati e quanto sono affidabili). **Fornire informazioni sulle licenze e i permessi in un formato machine-readable**. Ove possibile, dovrebbero essere utilizzati sistemi di licensing largamente accettati (ad esempio Creative Commons) e che possano essere referenziati tramite URL. **Garantire la riservatezza e i diritti degli interessati e dei creatori dei dati**, fornendo un sistema di autenticazione e profilazione



Conservazione

Un repository affidabile dovrebbe:

Assicurare la persistenza di metadati e dati, assicurando continuamente disponibilità ed accesso ai dati e ai metadati affidati ad esso dai suoi utenti. **Essere trasparente su missione, ambito, politiche di conservazione, piani** (inclusi governance, sostenibilità finanziaria, periodo di conservazione e piano di continuità), fornendo evidenza in maniera documentata anche di partner esterni per la gestione tecnica e fornendo informazioni su come verrebbero trasferiti i dati in caso di cessazione del repository.

