

Linee guida per la gestione dei dati della ricerca

Guidelines for Research Data Management (RDM)

1) Premessa

Lo scopo della gestione dei dati della ricerca è di favorire la reperibilità, l'accesso e lo scambio dei dati, massimizzando quindi l'efficacia e la riproducibilità dell'attività di ricerca.

Nel seguito si informa su tutti gli aspetti pratici della RDM, in particolare su:

- la scelta del repository *open data*
- la scelta dei dati da archiviare e le modalità per farlo
- il software di ricerca
- le licenze di utilizzo
- il modello suggerito per la preparazione di un Data Management Plan (DMP)

Le linee guida costituiscono un documento in continua evoluzione al fine di mantenere aggiornate le informazioni utili per chi debba applicare la "Policy RDM". A questo scopo, è in corso la preparazione di una versione web delle linee guida che considera le particolari caratteristiche dell'INRiM e dei suoi ambiti di ricerca.

2) Ambito di applicazione

In questa fase iniziale, l'applicazione della "Policy RDM" riguarda i **progetti finanziati** da terzi, nel rispetto delle normative date dal finanziatore. Tali progetti includono, ad esempio, quelli finanziati nell'ambito dei programmi Horizon Europe ed EURAMET. La documentazione allegata a queste linee guida contiene copia dei regolamenti e delle indicazioni per la preparazione di DMP attualmente in vigore.

3) Trattamento e gestione dei dati della ricerca

Premessa: Secondo la *Policy sulla gestione dei dati della ricerca INRiM*, i dati della ricerca (RD) devono essere archiviati in un sistema *appropriato* e gestiti in modo tale da risultare *findable*,

*accessible, interoperable and reusable*¹ (reperibili, accessibili, utilizzabili e disponibili per utilizzi successivi) oltre che valutabili e comprensibili.

Sistema appropriato: Per sistema *appropriato* si intende un sistema che sia in grado di garantire un livello ragionevole di sicurezza e affidabilità.

Ad esempio, sono considerati appropriati:

- Data repository (ad es. Zenodo): *scelta consigliata per la condivisione di RD legati a progetti finanziati da terzi: Zenodo, community INRiM (vedi sotto)*
- Servizi di cloud storage (ad es. Google Drive): *scelta consigliata per lo storage di tutti i RD e condivisione all'interno dell'INRiM*
- NAS/server di storage, purché gestiti secondo certi criteri (ridondanza tramite RAID, backup periodici, aggiornamento continuo del software, monitoraggio dei parametri degli hard disk e pronta sostituzione in caso di guasti, ..).
- Hard disk esterni, purché dotati di backup aggiornato (ossia una copia completa e aggiornata dei dati deve essere tenuta su un secondo supporto).

NON sono invece considerati appropriati:

- Computer personale o di laboratorio.
- Sito internet del singolo ricercatore/gruppo di ricerca.
- Chiavette USB.
- CD/DVD.

Dati di ricerca: per “dati di ricerca” (research data RD) si intendono tutti i dati in formato elettronico (file .dat, .txt, .avi, .mp3, etc.) utilizzati per *convalidare i risultati presentati su pubblicazioni scientifiche* prodotte nell’ambito di un progetto di ricerca (vedi ambito di applicazione).

I dati di ricerca includono inoltre i *software* sviluppati per la valutazione ed elaborazione dei dati grezzi, i metadati (vedi sotto), e ogni altro materiale (immagini, audio, etc.) necessario per comprenderne i contenuti e riutilizzarli.

Repository: I dati della ricerca che vengono resi pubblici, per esempio tramite pubblicazione su rivista, devono essere depositati su un repository specifico della disciplina, se esiste, e sul data repository *Zenodo community INRiM* (<https://zenodo.org/deposit/new?c=inrim>), indicando il progetto di ricerca.

Metadati: per “metadati” si intendono tutte le informazioni aggiuntive che identificano un insieme di dati di ricerca. Si distingue fra metadati di tipo “amministrativo” e “tecnico”.

I metadati “amministrativi” contengono informazioni quali:

- Titolo.
- Descrizione.

¹ FAIR è l’acronimo di: *Findable; Accessible; Interoperable; Reusable* (<https://www.go-fair.org/>).

- Elenco degli autori e relative affiliazioni.
- Parole chiave.
- Lingua.
- Licenza d'uso.
- Anno di pubblicazione.
- Elenco e codice dei progetti che hanno finanziato la ricerca.

I metadati amministrativi sono solitamente caricati direttamente in fase di upload sul repository certificato. È molto importante compilare tutti i metadati amministrativi, con la massima accuratezza possibile, per rendere i propri dati *Findable*.

I metadati “tecnici” contengono invece informazioni che servono a caratterizzare ulteriormente i dati. Purtroppo, tale caratterizzazione dipende dalla particolare tipologia di dato, e non è quindi possibile fornire un elenco esaustivo.

Metadati “tecnici” sono ad esempio:

- Data e ora di ciascuna misura.
- Condizioni ambientali.
- Informazioni tecniche sugli strumenti utilizzati durante la misura.
- ...

I metadati tecnici vengono normalmente scritti sotto forma di file di testo, o come *header* dei file contenenti i dati.

Per rendere la ricerca *Interoperable*, i metadati devono essere il più possibile *standard* e *formali*. Gli standard descrivono modi comuni per strutturare e rendere comprensibili i dati. Per esempio, possono prendere la forma di vocabolari controllati e/o ontologie. Nel riportare metadati metrologici, può essere utile seguire il vocabolario internazionale metrologico *VIM*.

Licenza d'uso: I dati della ricerca, salvo diverse e specifiche indicazioni, devono essere associati ad una licenza di utilizzo *open* secondo la “[open definition](http://opendefinition.org/)”², come la licenza [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) 4.0 (CC-BY).

Le licenze Creative Commons non si applicano ai prodotti software, che utilizzano invece licenze *open source*, quali la [licenza MIT](https://opensource.org/licenses/MIT) o la [licenza Apache 2.0](https://www.apache.org/licenses/LICENSE-2.0).

I metadati normalmente non possiedono una licenza d'uso. Nel caso sia obbligatorio indicarne una, *deve* essere applicata la licenza [Creative Commons Zero \(CC0\)](https://creativecommons.org/licenses/by/4.0/) (che si applica solo ai metadati).

² <http://opendefinition.org/>

4) Supporto

Per favorire la corretta implementazione della RDM, l'INRiM metterà a disposizione dei propri afferenti adeguati strumenti e informazioni di supporto:

- pagine web dedicate alla RDM/Open science (<https://www.inrim.it/it/ricerca/open-science>) contenenti tutti i documenti e le informazioni utili, tra cui:
 - ❖ “Policy RDM” e “Linee guida RDM”: <https://www.inrim.it/it/ricerca/open-science>
 - ❖ template per la preparazione del *Data Management Plan* (DMP)³ EURAMET EMPIR/EPM:
https://msu.euramet.org/downloads/documents/09_Data_Management_Plan_template.docx
 - ❖ ERC OPEN Science; information and templates (<https://erc.europa.eu/manage-your-project/open-science>)
 - ❖ EURAMET EMPIR/EPM Reporting Guidelines, Part 9 – Preparing data management plans:
https://msu.euramet.org/downloads/documents/EMPIR_Reporting_Guidelines_Part9.pdf
 - ❖ Annotated Model Grant Agreement di Horizon Europe, ANNEX 5 (p.154)
https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf
 - ❖ Horizon Europe Data Management Plan (DMP) template
https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
 - ❖ Guidelines on FAIR data management in Horizon 2020
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- documenti informativi ed esempi di applicazione (in preparazione)
- assistenza diretta da parte dei membri del gruppo Open Science
- una sezione FAQ accessibile dal sito web (in preparazione)

Allegato 1 - Definizioni (dalla Policy)

³ *Data management Plan* (DMP) è un documento che progetta l'intero ciclo di vita dei dati della ricerca e può essere continuamente aggiornato. Il DMP assicura che i dati della ricerca siano accessibili, tracciabili, disponibili, autentici, citabili e che rispondano a requisiti legali definiti con chiarezza e a misure appropriate di sicurezza in relazione al loro riutilizzo. Nella loro forma ideale i DMP saranno in un formato *machine actionable*.

Dati della ricerca: sono i dati in formato digitale prodotti dalle attività di ricerca, o che ne illustrano le fasi di lavoro. In particolare, sono da considerarsi dati della ricerca le registrazioni oggettive dei dati numerici, simboli, testi, audiovisivi utilizzati nei processi della ricerca scientifica e necessari per convalidare i risultati, così come i programmi software utilizzati nel corso del processo di elaborazione che produce i risultati finali della ricerca.

Esempi tipici di dati della ricerca sono quelli raccolti in fase di misura, i valori di riferimento di grandezze e campioni utilizzati in laboratorio, risultati di indagini, oggetti provenienti da raccolte, procedure di prova o simulazioni, protocolli. La gamma delle diverse tipologie di dati riflette la diversità e lo sviluppo metodologico delle discipline scientifiche e dei processi di ricerca. I dati della ricerca possono assumere forme diverse durante la durata dei progetti (diverse varianti dei dati primari, dati elaborati inclusi risultati negativi o ambigui, dati condivisi, dati pubblicati) e possono pertanto avere associate autorizzazioni di accesso diverse nel corso di tale evoluzione.

Gestione dei dati della ricerca: include la pianificazione, la registrazione, l'elaborazione, la documentazione e l'archiviazione, secondo le modalità delineate nella "*Policy*" e specificate nelle *linee guida*. Garantisce l'accesso, il riutilizzo, la riproducibilità dei dati della ricerca in quanto presupposti della qualità e trasparenza su cui si basano i risultati scientifici.

Afferenti all'INRiM: tutti i soggetti che contribuiscono alle attività di ricerca dell'INRiM inclusi il personale dipendente dell'INRiM, i titolari di assegno di ricerca e il personale che è in formazione o in visita presso l'INRiM e afferisce ad altri Enti o a titolo privato.

Progetto di ricerca: è l'impresa scientifica che risponde a un'esigenza di ricerca, indipendentemente dalla fonte di finanziamento. I progetti di ricerca possono includere: attività sperimentali, studi teorici, studi di fattibilità, analisi dei dati Anche le attività di ricerca previste per il conseguimento di un titolo di dottorato o di laurea magistrale si configurano come progetto di ricerca. Ogni progetto di ricerca deve avere un responsabile che ha il compito di stesura del DMP e supervisione sulla gestione dei dati di ricerca.