



ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA  
AREA SISTEMI DIPARTIMENTALI E DOCUMENTALI (ASDD)  
Centro Risorse per la Ricerca Multimediale (CRR-MM)

# DATA MANAGEMENT PLAN

Open Research Data Pilot - Horizon2020

---

GUIDA ALLA REDAZIONE

## Introduzione

I progetti di ricerca che intendono partecipare all'*Open Research Data Pilot - Horizon 2020* sono chiamati a fornire un *Data Management Plan (DMP)* che illustri le specifiche dei dati creati nel corso del progetto e che verranno disseminati in *Open Access*. Nell'ambito del *Work Programme 2014-2015* le seguenti aree partecipano all'*Open Research Data Pilot*:

- Future and Emerging Technologies
- Research infrastructures - part e-Infrastructures
- Leadership in enabling and industrial technologies - Information and Communication Technologies
- Societal Challenge: 'Secure, Clean and Efficient Energy' - part Smart cities and communities
- Societal Challenge: 'Climate Action, Environment, Resource Efficiency and Raw materials' - except raw materials
- Societal Challenge: 'Europe in a changing world - inclusive, innovative and reflective Societies'
- Science with and for Society.

(cfr. <http://www.sussex.ac.uk/library/research/researchdatamanagement/create/biddingforfunding/horizon2020dataplan>)

Nonostante non siano stati posti vincoli ai progetti non partecipanti al Pilot, i progetti inerenti a *Research and Innovation actions* e *Innovation actions* sono chiamati a redigere un sintetico DMP, in quanto rilevante per la pianificazione del lavoro e valutabile sotto il parametro "*impact*".

## Il Data Management Plan

Innanzitutto va detto che il Data Management Plan non è un documento immutabile: è previsto che una prima versione debba essere redatta entro 6 mesi dall'inizio del progetto, ma è possibile rielaborarne di nuove (o arricchire la precedente) nelle fasi successive di avanzamento del progetto, ogni qualvolta intervenga un cambiamento significativo (e.g., la creazione di un nuovo dataset, un cambiamento nella *policy*), sino al completamento di un piano definitivo.

Scopo del DMP è fornire informazioni sulla *Data Management Policy* che si intende adottare per rendere disponibili i dati della ricerca in *Open Access*, ovvero una analisi e una descrizione dei dati stessi, degli standard, delle tecniche e dei *workflow* che li caratterizzano e delle politiche di accesso, riuso e preservazione.

## Il template

Il seguente template, basato sulle linee guida fornite da *Horizon 2020* e dal *Digital Curation Center (DCC)*, prevede una struttura in cinque sezioni per la redazione di un primo DMP ed è già predisposto con alcune indicazioni di contenuti per la successiva stesura in fase avanzata del progetto di ricerca.

Con ciò si intende porre insieme tutti i valori richiesti dalla versione finale di DMP, lasciando al compilatore la scelta di quali inserire o meno a seconda della fase in cui si va a redigere il piano.

Idealmente, questo template è inteso per ogni singolo dataset prodotto dal progetto di ricerca ed è quindi reiterabile.

Di seguito viene fornita una versione in italiano del DMP, completa ed esplicativa, con una descrizione delle sezioni, del contenuto richiesto per ciascuna nonché alcuni consigli sull'impostazione con cui caratterizzare le risposte.

Per prendere visione di alcuni esempi precompilati (in inglese), consigliamo la consultazione dell'apposita sezione sul sito di DCC (<http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>).

In allegato proponiamo un template vuoto per la compilazione.

## Come possiamo esservi utili?

Il [Centro Risorse per la Ricerca Multimediale \(CRR-MM\)](#) offre assistenza e consulenza ai ricercatori che intendono partecipare alla call di *Horizon 2020*. Nello specifico offriamo:

- consulenza nella redazione del Data Management Plan;
- supporto per la definizione delle fasi di creazione e gestione dei dati della ricerca;
- assistenza nella scelta e nello sviluppo dei software necessari all'accesso e alla manipolazione dei dati;
- un sistema di storage e preservazione dei dati prodotti;
- soluzioni per la valorizzazione sul web dei progetti di ricerca.

## Informazioni generali sul Progetto di Ricerca

### Nome del Progetto di Ricerca

Inserire il nome del progetto di ricerca.

### ID del Progetto di Ricerca

Inserire un identificativo del progetto attribuito dalla propria istituzione di riferimento.

### Grant Number

Inserire il riferimento del progetto in Horizon 2020 (solo per DMP post-award).

### Ricercatore Responsabile

Inserire il nome del ricercatore responsabile del progetto.

### ID del Ricercatore Responsabile

Inserire un identificativo del ricercatore di riferimento (e.g., ORCID).

### Contatti

Inserire nome (se diverso), numero di telefono ed email del ricercatore responsabile.

### Descrizione

Inserire una breve descrizione degli obiettivi del progetto di ricerca e dei dati che si intende produrre.

### Data di Creazione del DMP

Inserire la data di creazione del piano.

# Data Management Plan

## Nome | Identificativo del Dataset

Inserire il nome e l'identificativo univoco del dataset prodotto o che si intende produrre per garantirne la *discoverability*.

L'Area dei Sistemi Dipartimentali e Documentali - Settore AlmaDL ha aderito al progetto coordinato dal gruppo di lavoro Open Access della CRUI (Conferenza Permanente dei Rettori delle Università Italiane) per l'attribuzione del DOI (*Digital Object Identifier*), nella sua versione non commerciale, ai prodotti della ricerca. Il DOI è l'identificatore persistente più utilizzato nella letteratura scientifica, diventato negli ultimi anni standard de facto e condizione necessaria per le pubblicazioni di qualità (cfr. <http://www.cruir.it/HomePage.aspx?ref=2120> per la convenzione tra CRUI e DataCite).

## Descrizione del Dataset

Inserire una descrizione del dataset che si intende collezionare o che si è già collezionato, indicandone la provenienza, la natura dei dati e l'ordine di grandezza.

- Inserire un riferimento ad un potenziale pubblico di interesse e come potrebbe essere loro utile la collezione;
- inserire i riferimenti di *data sources* utilizzati per la creazione del nuovo dataset (se esistenti), quali similarità presentano col progetto ed eventualmente come integrarli e riusarli;

Motivare la creazione di un nuovo dataset: specificare i legami con data sources disponibili e i loro limiti rispetto alle esigenze del progetto di ricerca, definendo il valore aggiunto che apporterà il nuovo progetto (e.g., fornire una breve descrizione dei dati riusati, esplicitando le carenze quantitative o qualitative, l'insufficiente livello di specificità rispetto ai fini dell'analisi...);

- citare, se esistenti, le pubblicazioni inerenti ai dati prodotti.

## Standard e Metadati

Descrivere i dati che verranno creati dal progetto di ricerca, in termini di contenuto e qualità.

- Descrivere i procedimenti utilizzati per la creazione della collezione di dati (e.g., collazione tra data sources, creazione ex novo, analisi di casi esitenti...);
- definire la tipologia, il volume ed il formato dei dati, motivandone la scelta (e.g., conversione dei dati utile ai fini della preservazione, utilizzo di formati standard e accessibili via software open source);
- descrivere i criteri previsti per garantire ai dati *quality assurance* (e.g., procedimenti di *data cleaning*).

Descrivere i metadati e la documentazione che accompagneranno il dataset.

- fornire sufficienti informazioni finalizzate alla autodescrittività del dataset, offrendo quindi una efficace interpretazione del contesto di produzione dello stesso (e.g., documento di overview allegato al dataset con esempi, descrizione dei focus di ricerca, date e riferimenti di *provenance*);
- specificare gli standard utilizzati per la metadattazione (una buona documentazione aumenta la *discoverability* dei propri dati).

Inserire i riferimenti a standard di disciplina (se) utilizzati, sia per la collezione dei dati sia per il management: se attualmente le scelte ricadono su tipologie di dati non standard (o non largamente utilizzate dalla comunità di riferimento del progetto), prevenirne una conversione ai fini dell'interoperabilità.

Prediligere formati non proprietari e accessibili via software open source.

## Accesso e Condivisione dei Dati

Descrivere come si intende condividere i dati in Open Access e fornire istruzioni per l'accessibilità.

- Specificare (se previsto) il periodo di *embargo*;
- spiegare i meccanismi tecnici per la disseminazione dei dati;
- descrivere software e tools richiesti per l'accesso ed il riuso dei dati: se creati unitamente ai dati, garantirne un'identificazione univoca e la loro preservazione (e.g., depositando il codice sorgente del software su

GitHub, <https://github.com>; AMS Acta, <http://amsacta.unibo.it>);

- definire la licenza con cui vengono forniti i dati (garantendo i diritti specificati dal Model Grant Agreement): tra le licenze fortemente consigliate, le Creative Commons *CC0* e *CCBy* (cfr. <http://www.creativecommons.it>).

Indicare il repository in cui verranno depositati i dati: se già esistente e identificabile, indicare la tipologia di repository utilizzato (e.g., archivio aperto istituzionale AMS Acta, <http://amsacta.unibo.it>).

Prediligere i repository istituzionali interoperabili con OpenAIRE (cfr. <https://www.openaire.eu>).

Specificare i metodi per l'utilizzo dei dati, affinché vengano riutilizzati prevenendo incomprensioni, fornendo indicazioni sullo standard qualitativo minimo prestato all'accesso ed il riuso dei dati.

- Specificare le norme per l'accesso ai dati e le specifiche di *data security*: se l'accesso è ristretto ad alcuni gruppi di utenti, motivarne le ragioni, dimostrando di aver preso visione delle norme sulla privacy e la proprietà intellettuale;

- in caso di dati sensibili o soggetti a diritti di proprietà intellettuale, spiegare come si intende tutelarsi (e.g, anonimità dei dati riferiti a persone, rispetto e riferimento alla tipologia di copyright).

Se non si intende disseminare i dati pubblicamente, specificarne le ragioni: e.g, motivazioni etiche, informazioni che inficiano la privacy, dati personali, finalità commerciali, proprietà intellettuale, norme sulla sicurezza (cfr. i casi di *opt out* previsti dal Model Grant Agreement di Horizon2020).

## Archiviazione e Preservazione dei Dati

Descrivere le procedure che si intendono adottare per la preservazione dei dati a breve e lungo termine.

- Definire quali dati verranno preservati/archiviati a lungo termine (e.g, dati unici, dati secondari...);

- specificare il periodo di preservazione dei dati ed i sistemi di back-up utilizzati durante il ciclo di vita del progetto (e.g, storage sui server del CRR-MM ospitati dal CeSIA);

- indipendentemente dalle scelte effettuate per la preservazione, garantire il deposito e l'accesso ai dati utilizzati per validare una pubblicazione scientifica (*underlying data*) in un formato machine-readable;

- definire una stima (o un dato certo) del volume finale dei dati, della qualità ed il formato che verranno garantiti nello storage.

Definire i costi necessari a queste operazioni (come si intende coprirli) e il team responsabile di *data management* e *data curation*. La definizione dei costi varia proporzionalmente al volume dei dati, quindi alla qualità e al formato che si intende mantenere, nonché ai sistemi di storage e backup che si vogliono predisporre e agli eventuali servizi aggiuntivi (hosting di banche dati, data integration, data visualization...).

Nel caso in cui sia previsto l'accesso remoto ai dati in un network distribuito:

- definire i responsabili del monitoraggio;

- dichiarare le convenzioni utilizzate per il naming dei file;

- definire il sistema di versioning fornito per l'accesso ai dati.



ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA  
AREA SISTEMI DIPARTIMENTALI E DOCUMENTALI (ASDD)  
Centro Risorse per la Ricerca Multimediale (CRR-MM)

---

<http://www.unibo.it/it/ricerca/progetti-e-iniziative/centro-di-risorse-per-la-ricerca-multimediale-crr-mm>  
asdd.crrmm@unibo.it - 051 209 4318  
Viale Filopanti 7 - 40126 Bologna - Italia