



OpenAIRE



Che cos'è un Data Management Plan: presentazione e casi d'uso

Marisol Occioni – Marialaura Vignocchi
Università Ca'Foscari Venezia – Università di Bologna



Benvenuti da:



RDA - Research Data Alliance

è una **organizzazione internazionale** i cui membri si incontrano e discutono con lo scopo di facilitare la condivisione e il riuso dei **dati**. I membri RDA lavorano in Gruppi spontanei e volontari e si incontrano due volte all'anno nelle Plenary. Il capitolo Europeo di RDA sta formando una rete di nodi nazionali attraverso il progetto RDA Europe 4.0 che possano portare i bisogni della comunità Nazionale in RDA e viceversa. Il progetto finanzia la partecipazione ad RDA attraverso diversi tipi di **grant**. Per saperne di più, <https://rd-alliance.org/groups/rda-italy>



OpenAIRE

è la grande **infrastruttura** per il monitoraggio e la verifica delle politiche sull'Open Access finanziata dalla Commissione Europea dal 2009. OpenAIRE collega i progetti ai risultati della ricerca (pubblicazioni, dati, software ecc). Ad oggi serve 16 enti finanziatori in Europa e nel mondo, offre supporto all'Open Science a 360 gradi attraverso la rete di **National Open Access Desks** (NOADs). Per saperne di più, www.openaire.eu



IOSSG - Italian Open Science Support Group

è un **gruppo di lavoro** nato dalla collaborazione di professionisti che lavorano in diversi atenei con competenze trasversali, operanti nell'ambito delle aree: supporto alla ricerca, biblioteche digitali, Open Science, legale, ICT. Per saperne di più, IOSSG

Definizione di dato della ricerca

Cos'è un dato della ricerca

I dati della ricerca sono registrazioni fattuali, che possono assumere la forma di numeri, simboli, testi, immagini o suoni, utilizzati come fonti primarie per la ricerca, comunemente accettate nelle comunità di ricerca come necessarie per convalidare i risultati della ricerca (The Australian Griffith University)

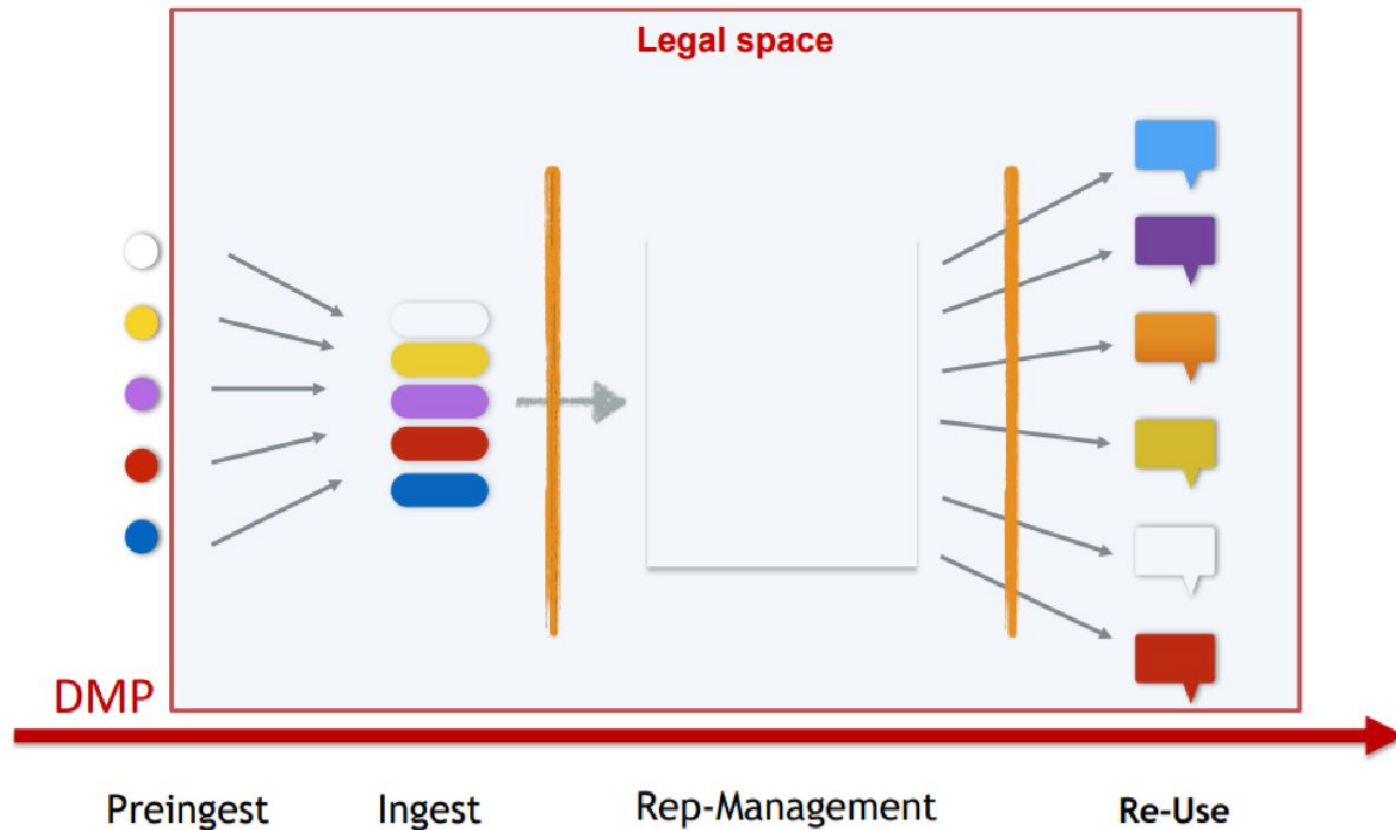
Definizione di DMP

Cos'è il Data Management Plan

1. un piano formale di gestione dei dati che accompagna il lavoro di ricerca e descrive quali **dati** sono raccolti, processati e generati, dove vengono archiviati, a quali licenze e limitazioni sono legati, chi ha diritto al loro utilizzo
→ **DATA**
2. un riferimento per **gestire** responsabilmente i dati e renderli Findable, Accessible, Interoperable, Reusable (FAIR) → **MANAGEMENT**
3. un **piano** per programmare e comunicare la raccolta, conservazione, riuso e divulgazione dei dati + metadati associati, sin dall'inizio dell'attività → **PLAN**

Cos'è il Data Management Plan (2)

4. documento redatto dal ricercatore come template o tool online ([DCC](#), [IOSSG](#), [ERC](#), [H2020](#), [Data Stewardship Wizard](#)) e rappresenta **l'intero ciclo di vita del dato**, assicurandone tracciabilità, disponibilità, autenticità, citabilità, conservazione appropriata, adesione a parametri legali chiari e l'adozione di misure di sicurezza adeguate, che ne assicurano e disciplinano gli usi successivi. Il dato dovrà essere: affidabile, robusto, completo, esaustivo, unico, certificato.



Graphic by Raman Ganguly, University of Vienna, Project e-Infrastructures Austria

Cos'è il Data Management Plan (3)

5. un *living document*: possono essere redatte **varie versioni** a seconda delle fasi di avanzamento della ricerca (ad es.: creazione di un nuovo dataset)
6. esteso o sintetico a seconda delle necessità del progetto
7. un *deliverable*: da redigere **entro 6M** dall'approvazione del progetto
8. un impegno verso i finanziatori: **Open Access by default with some opt outs/as open as possible, as closed as necessary**

Redigere un DMP ma **non** rendere accessibili i dati

H2020, pur estendendo l'Open Data a tutte le aree del programma, prevede la possibilità di **opt-out** , mantenendo chiusi i dati nei casi di:

- a) tutela della privacy delle persone coinvolte nella ricerca
- b) ragioni di sicurezza
- c) produzione di dati sensibili
- d) sfruttamento economico dei risultati della ricerca (brevetti)
- e) nei progetti in cui non è prevista la produzione di dati
- f) altre situazioni legittime e **adeguatamente motivate**

Un esempio: Sensible Data o Ethic Issues

La Commissione Europea prevede la possibilità di mantenere riservati e protetti i dati sensibili o con problematiche legate alla privacy delle persone coinvolte nella ricerca.

In questo caso è necessario che il DMP riceva un avallo preventivo su questi aspetti da parte della Commissione Etica di Ateneo.

Il caso di UniVE: il Data Monitoring Board
Il caso di UniBO: Comitati etici, referente GDPR, es. “Ethical aspects” in DMP;
importanza del DMP per sensibilizzare su problemi etici e di privacy

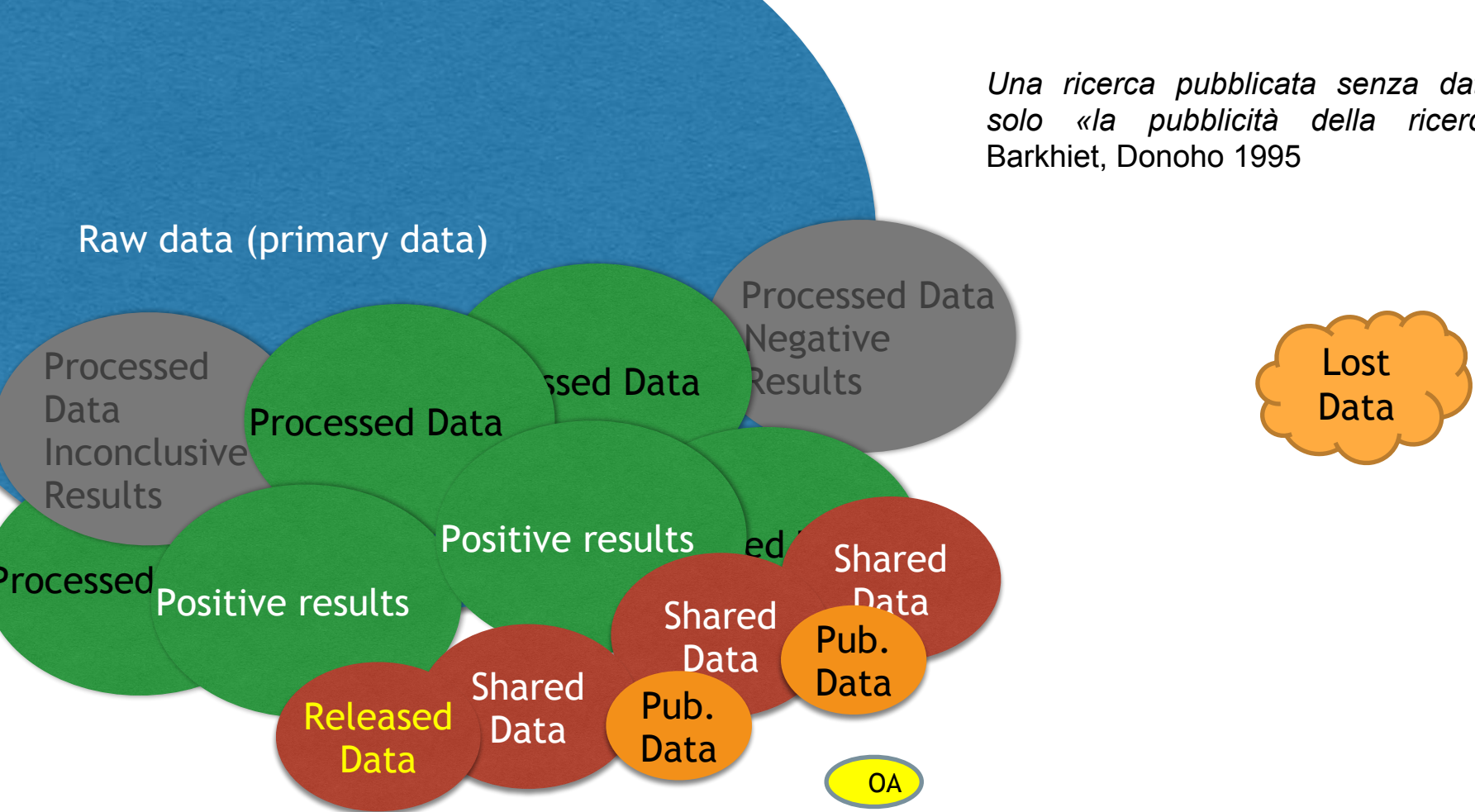
Gli attori del DMP

- l'**ente finanziatore** (fornitore di finanziamento, ad es: [CE](#))
- il **ricercatore** (fruitore del finanziamento e redattore del DMP)
- le **strutture di supporto** (consulenza e supporto competenze trasversali): per saperne di più, appuntamento con «[Data Literacy](#)» nel 2019
- l'**istituzione** (governance che rilascia la policy, dà valore alla ricerca, organizza il supporto, etc...): per saperne di più, appuntamento con «[Open Science Policies](#)» nel 2019

Un mondo di dati

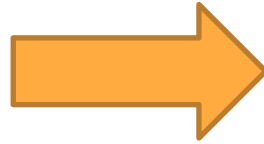


Una ricerca pubblicata senza dati è solo «la pubblicità della ricerca», Barkhiet, Donoho 1995



Workflow dei dati

data collection
data documentation
data storage & back up
data access & security
data preservation & reuse



Sezioni del DMP IOSSG

Administrative plan details
Dataset description
Standard & metadata
Data management, documentation & curation
Data security & confidentiality of potentially
disclosive information
Data sharing & access
Responsibilities & resources
Relevant institutional policies on data sharing
& data security

Struttura del template di IOSSG

1. **Administrative plan details:** Informazioni generali sul progetto di ricerca
2. **Dataset description:** descrivere i dati esistenti o che si intendono creare, indicandone provenienza, natura e ordine di grandezza. Motivare la creazione di nuovi dataset e relativo valore aggiunto
3. **Standards & metadata:** descrivere tipo e standard di metadati ([DC](#), [DCC](#), [RDA](#), ...), la metodologia di raccolta e generazione dei dati con riferimento alla prassi della comunità di pratica. Garantire [qualità](#), ricchezza, accuratezza, riusabilità con riferimento a standard e vincoli (se esistenti)

Struttura del template di IOSSG

4. **Data management, documentation & curation:** descrivere i processi adottati per gestione, documentazione, cura e strategie di conservazione a lungo termine in repos certificati (Istituzione, [re3data](#), [Zenodo](#)). Stima dei costi
5. **Data security & confidentiality of potentially disclosive information:** specificare le procedure adottate per sicurezza, conservazione, trasmissione e scambio dei dati. E' la sezione più importante per i progetti che trattano dati sensibili (soggetta a validazione della Commissione Etica).
6. **Data sharing & access:** descrivere quali dati e come saranno condivisi e resi

Struttura del template di IOSSG

disponibili, le politiche di accesso ai repos, le licenze ([CC](#), [DCC](#),), sw e tool per trasmissione, circolazione e riuso. Sezione rilevante per progetti con open data. Da non completare, ma motivare per gli altri

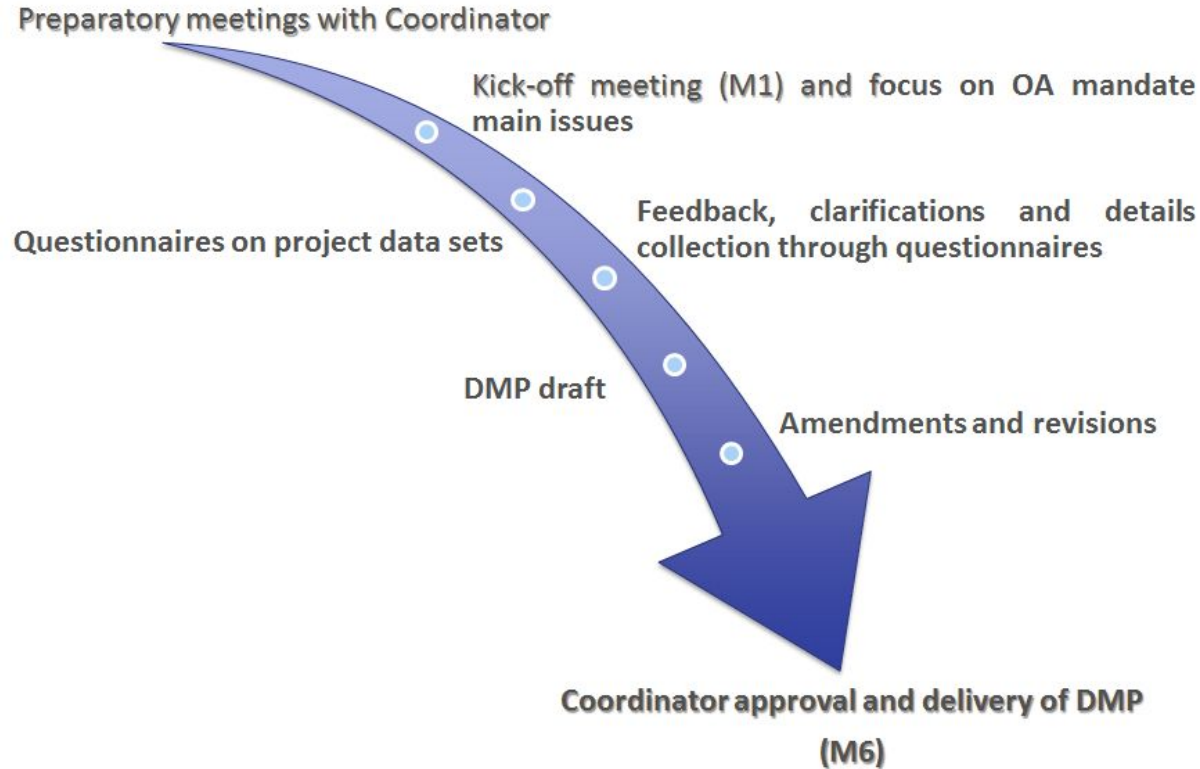
7. **Responsibilities & resources:** dichiarare le responsabilità a livello di progetto/Ateneo (raccolta, creazione, qualità, sicurezza, archiviazione). Evidenziare la necessità di risorse aggiuntive (consulenze, HW, SW). Dichiarare la compatibilità del progetto con la policy sui dati della ricerca, i requisiti legali ed etici
8. **Relevant institutional policies on data sharing & data security:** riferirsi alle policy di Ateneo (dati e pubblicazioni) e al DMP come deliverable



Casi d'uso



Le tappe per la redazione del DMP in H2020



Preventivare i costi in H2020

I costi per la gestione FAIR dei dati sono **spese coperte dal finanziamento** nei progetti H2020. E' importante preventivarli correttamente fin dall'inizio e comprenderli nel DMP. Esistono delle **checklist con stime approssimative dell'impegno orario delle varie attività che aiutano a considerare tutti i costi**. I costi si abbassano se le attività necessarie per l'elaborazione dei dati ai fini della loro condivisione FAIR sono ben identificate dall'inizio e sono portate avanti nel corso della ricerca (es. controllo di qualità, *data cleaning* e armonizzazione dei dati, *codebooks*, anonimizzazione...)

[OpenAIRE costing tool](#)

[UK Data Service Costing Tool](#)



Punti di attenzione per chi fa ricerca (1)

- **accesso e protezione dei dati**: quali dati devono essere protetti per ragioni di confidenzialità o per un possibile impiego industriale/commerciale? L'esigenza di tutela è temporanea o permanente? Copre tutti i dati del progetto o solo una parte? Posso applicare un embargo? Quali dati posso condividere?
- **dati di terzi**: utilizzerò dati di fonti terze? Quali sono le loro condizioni di utilizzo? Devo chiedere un permesso per analizzare i dati e per pubblicare i dati derivati?
- **dati personali e sensibili**: la ricerca coinvolge persone, le loro opinioni, il loro stato di salute o giuridico, l'orientamento sessuale? Considera l'impatto della legge privacy, le misure da adottare e i costi per gestire e disseminare i dati (consenso informato, storage sicuro e protetto, trascrizioni anonimizzate, pubblicazione in forma aggregata, distruzione dei dati raw)

Punti di attenzione per chi fa ricerca (2)

- **riuso dei dati**: l'elaborazione dei dati richiede dei sw particolari o proprietari? Considera la necessità di mettere a disposizione il sw o di salvare i dati in formati aperti e riutilizzabili da tutti, fornendo tutta la documentazione utile alla loro comprensione (*codebooks*, istruzioni per la conversione...). Usa gli standard della tua comunità di ricerca e scegli licenze che consentano il riuso senza ambiguità o limitazioni ([CC-BY 4.0](#))
- **salvataggio e archiviazione dei dati**: come conservo i dati durante la ricerca? Considera uno spazio dove salvare e condividere i dati con i partner che sia sicuro e rispetti le normative europee. Quale repository uso per la conservazione e l'accesso pubblico? Scegli i *data repository* in uso presso la tua comunità scientifica e adatti al tipo di dati che stai generando, informati presso la tua istituzione, o in alternativa utilizza [Zenodo](#)

Il questionario per i ricercatori: griglie per DMP

Tools utili: [DMOnline](#); [Data Stewardship Wizard](#); [easyDMP](#); [RDMO](#)

The screenshot shows the Data Stewardship Wizard interface. On the left is a sidebar with an orange header and a grey body. The header contains 'KM Editor', 'KM Packages', and 'DS Planner'. The body contains a 'Core DS Knowledge Model' section with a list of items: 'Design of experiment', 'Data design and planning', 'Data Capture/Measureme', 'Data processing and curat', 'Data integration', 'Data interpretation', and 'Information and insight'. Below this is a user profile for 'Marialaura Vignocchi'. The main content area has a title bar 'Design or experiment' and a 'Text' section with a paragraph: 'Before you decide to embark on any new study, it is nowadays good practice to consider all options to keep the data generation part of your study as limited as possible. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others.' Below the text is a 'Questions' section with four questions: 'Is there any pre-existing data?', 'Will reference data be created?', 'Will you be storing samples?', and 'Will you be collecting experimental data?'. A red circle highlights a '+ Add question' button below the questions.



Fornire supporto ai ricercatori (1)

- Analizzare i dati di progetto e identificare i dataset (cos'è un dataset?)
- Elaborare procedure per il trattamento dei dati personali e sensibili e fornire [modelli per il consenso informato](#), [sw per l'anonimizzazione](#), [template e tool](#) ...
- Definire lo spazio di gestione dei dati nel corso della ricerca (intranet, *cloud*...)
- Identificare i ruoli di responsabilità (data creator, data contributor, contact person)
- Stabilire delle convenzioni per il *file naming* e il *versioning*
- Dare indicazioni sui [formati accettabili per la conservazione e la condivisione](#)
- Definire [i metadati](#) minimi e obbligatori (es. IDs, *funder* e progetto...)
- Spiegare le licenze da adottare
- Fornire template per il [read.me file](#) di documentazione
- Dare indicazioni per la scelta del data repository e istruzioni per il deposito (tempi, metadati, sistemi di *versioning*, embargo, *cross-linking*...)

Fornire supporto ai ricercatori (2)

Caratteristiche di un buon *data repository* (no soluzioni fai-da-te, come il sito di progetto!!!):

- deve essere affidabile, possibilmente certificato, con governance pubblica e politiche di conservazione a lungo termine dei dati e di riuso dei metadati
- deve supportare protocolli standard per l'interoperabilità ([OAI-PMH](#))
- deve supportare metadati standard come quelli richiesti da [OpenAIRE](#) che consentano la relazione con le pubblicazioni e i dati di progetto
- deve essere in grado di attribuire un identificatore persistente e univoco ai dati (es. [DOI](#), [URI](#), [URN](#), [HANDLE](#), [ARK](#))
- deve supportare licenze aperte come le Creative Commons ([CC](#))
- deve garantire l'integrità dei dati
- deve essere possibilmente registrato in [re3data.org](#)

IL DMP va aggiornato!

Esempio di DMP iniziale (M6) secondo le [Guidelines - H2020](http://doi.org/10.6092/unibo/amsacta/5551)

<http://doi.org/10.6092/unibo/amsacta/5551>

Esempio di DMP aggiornato (revisione di mid-term M13)

<http://doi.org/10.6092/unibo/amsacta/5732>

Il DMP finale indicherà come e dove sono stati archiviati gli *underlying data* e gli altri dati del progetto, se viene applicato un periodo di embargo e con quali motivazioni



Table of Contents

The Data Management Plan (DMP)	1
1. Data summary	1
2. FAIR Data	3
2.1 Making data findable, including provisions for metadata	3
2.2 Making data openly accessible	4
2.3 Making data interoperable	7
2.4 Increase data re-use (licensing)	7
2.5 Allocation of resources	8
2.6 Data security	10
2.7 Ethical aspects	11
3. Data sets overview	13
Annex I: Data sets tables	14
WP1 - Framework for comparative analysis: differences, implementation, perceptions and data gathering	14
WP2 - Evaluation of EU citizens' perception of the EU project in relation to regional performance of the Cohesion Policy and institutional quality	17
WP3 - The effectiveness of projects' communication strategies	20
WP4 - Spatial determinants of policy performance and synergies	22
WP5 - The multiplicity of shared meanings of EU and Cohesion Regional and Urban Policy at different discursive levels	25

Lavori in corso: verso un *open* DMP

Oggi il DMP è un documento di testo redatto a mano dal ricercatore che risponde a una serie di domande organizzate secondo un template indicato dall'ente finanziatore. Poco standardizzato, oneroso da compilare, difficile da fruire e da verificare da parte dei vari stakeholder della filiera.

[RDA DMP Common Standards WG](#) sta lavorando per trasformare il DMP in un testo strutturato per automatizzarne gran parte della compilazione e renderlo fruibile in modo automatico (*machine readable and actionable*) dai fornitori di servizi (aggregatori, repository, motori di ricerca...)





Example

- Current DMPs – model questionnaires

<administrative_data>

<question>Who will be the Principle Investigator?</question>

<answer>The PI will be John Smith from our university.</answer>

</administrative_data>

- Machine-actionable DMPs – model information

"dc:creator":[{

"foaf:name":"John Smith",

"@id":"orcid.org/0000-1111-2222-3333",

"foaf:mbox":"mailto:jsmith@tuwien.ac.at",

"madmp:institution":" AT-Vienna-University-of-Technology"

}],

uso di schemi
standard: DC;
PID: Orcid...;
Vocabolari
controllati

06/11/2018

www.rd-alliance.org - @resdataall



OpenAIRE

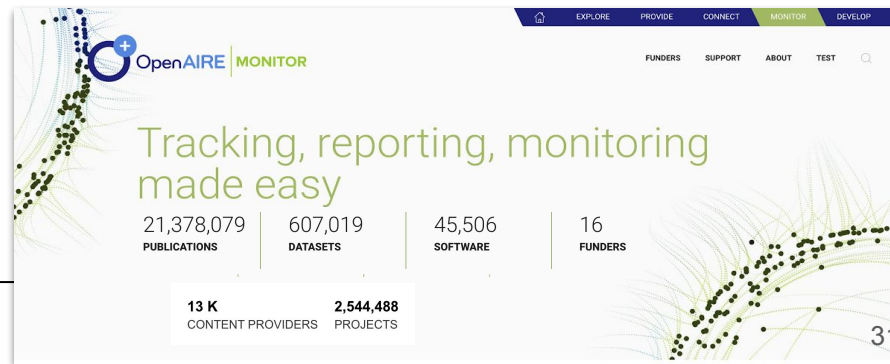


... e la visibilità?

Il progetto OpenAIRE (<https://www.openaire.eu/>) raccoglie e rende visibili tutte le pubblicazioni, i dati e gli altri risultati delle ricerche finanziate dalla Commissione Europea. OpenAIRE comunica direttamente al Participant Portal i risultati della ricerca associati ad un determinato progetto [NB: associare **SOLO i progetti pertinenti**, per rendicontazione scientifica e finanziaria]

Sul nuovo portale appena rilasciato è possibile fare una ricerca

- per grant number per affiliazione
- per pubblicazione
- per DOI
- ecc...



OpenAIRE2020

Open Access Infrastructure for Research in Europe 2020 (643410)

Project

EC

Open Access mandate

Funding: H2020 | RIA

Start Date: 2015-01-01

End Date: 2018-06-30

Organization: LANDSPITALI | JISC | EMBL | UZH | CVTISR | UBO | LIBER | SDU | UB
 UvA | COUPERIN | UL | University of Bielefeld | Athena RC | UH | UNIWARSAW
 UNIVERSITEIT GENT | IMI BAS | LU | IYTE | VSB | FECYT | RBI | SURF | UKON
 Universidade do Minho | UTARTU | UGOE | UOM | CERN | TU Delft | EIFL | CNR | UoD
 DANS | CINECA | TCD | UNIVIE | NKUA | COAR e.V. | University College London | UiO
 KBSE | UniHB | CLARA | UCY | Sarminfo | KTU | NHRF (EIE) | UL

[Detailed project information \(CORDIS\) →](#)

Publications (89)	+
Research Data (20)	+
Software (6)	+
Other Research Products (18)	+
Statistics	+
Metrics	+

Share - Bookmark



Application Box

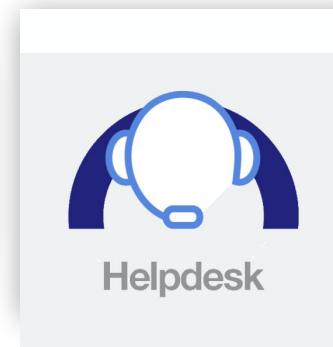
Publications

Research Data

[< > Include in your site \(HTML\)](#)
[≡ Get EC report \(HTML\)](#)
[↓ Get EC report \(CSV\)](#)
[↑ Deposit](#)
[Link this project to...](#)

Servizi OpenAIRE per i Ricercatori

- Catch all Repository: Zenodo
- Servizio di anonimizzazione dei dati: Amnesia
- Associare i dati e le pubblicazioni: Scholexplorer
- Help Desk Nazionale: NOAD (noad-it@openaire.eu)
- OpenDMP presto disponibile!



Research Data Alliance



In cosa può aiutare la Research Data Alliance?



RDA è una organizzazione internazionale i cui membri collaborano allo sviluppo di infrastrutture ed attività che hanno come scopo la riduzione degli ostacoli esistenti alla condivisione e allo scambio di dati, per accelerare l'innovazione derivante dai dati a livello mondiale.

Con più di 7000 membri da 137 Paesi, la comunità di RDA include ricercatori, scienziati e professionisti dei dati che lavorano in diverse discipline, campi e su diverse tematiche e da diverse tipologie di organizzazioni nel Mondo.

Come Opera RDA



I membri di RDA si confrontano attraverso **Working Group**, focalizzati su uno specifico tema, e **Interest Group** più esplorativi, che lavorano su argomenti più generici. Entrambe le tipologie di gruppi nascono spontaneamente nella comunità, su base volontaria e hanno lo scopo di scambiare opinioni, conoscenze, scoperte e discutere ostacoli e possibili soluzioni, esplorando e definendo policy e collaudare ed armonizzare gli standard per migliorare e facilitare la condivisione e il riutilizzo dei dati a livello globale.

Qualche aiuto specifico da RDA

- [Metadata Standards Directory](#): una directory che raccoglie standard e strumenti utili catalogandoli per le diverse discipline. È mantenuta da un [WG di RDA](#)
- [RDA/CODATA Summer Schools](#) in Data Science and Cloud Computing in the Developing World
- [Exposing DMPs WG](#)
- [DMP Common Standards WG](#)
- [FAIRSharing Registry: connecting data policies, standards & databases WG](#) il WG ha già prodotto una [Raccomandazione RDA](#)
- [FAIR Data Maturity Model WG](#)



Cascading grant di RDA Europe

Il progetto H2020 RDA Europe 4.0 ha implementato un programma di finanziamento per incoraggiare la partecipazione in RDA

Cascading Grants RDA Europe 4.0:

- Early Career
- Experts
- Adoption Grants



IOSSG



Il gruppo di lavoro IOSSG

Fornisce supporto e strumenti per rispondere alle sfide che la scienza aperta pone al mondo della ricerca: **processi della ricerca**, **ciclo di vita dei dati**, loro **gestione**, questioni relative alla **governance**, con particolare attenzione ai **servizi** e alle **infrastrutture**. L'attività del gruppo ha come focus l'impatto a livello locale delle politiche **EOSC**.

Materiali prodotti:

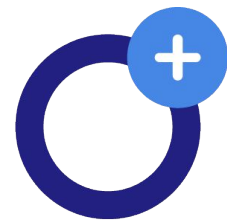
- [Modello di Policy sulla gestione dei dati della ricerca](#)
- [Data Management Plan Checklist](#)

Grazie!

loni@unive.it marialaura.vignocchi@unibo.it



RESEARCH DATA ALLIANCE
EUROPE



OpenAIRE

IOSSSG

