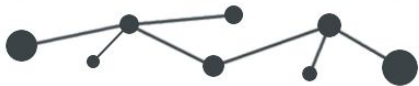




OpenAIRE

IOSSG



Research Data Management e politiche europee sui dati

Elena Giglia

Università di Torino - Unità di progetto Open Access

OpenAIRE NOAD, IOSSG

Emma Lazzeri

CNR-ISTI Pisa

OpenAIRE NOAD, Responsabile Nodo RDA Italia

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

...perché occuparci dei dati?

- perché una buona gestione dei dati è alla base di una buona ricerca
- perché una ricerca pubblicata senza dati è solo “la pubblicità della ricerca”
[[Berkhiet, Donoho 1995](#)]
- perché la scienza risulta più solida e robusta, e si discute di dati e non della loro interpretazione... e le frodi sono più difficili
- perché i dati creano ponti fra le discipline e aprono la strada a ricerche interdisciplinari , es. per le sfide globali come il cambiamento climatico
- perché permettono di trovare errori [il caso Rehinart-Rogoff, [video di K. Birney](#)]
- perché l'accesso ai dati
 - permette il riuso, con economie di scala
 - evita duplicazioni
 - favorisce la Open Science

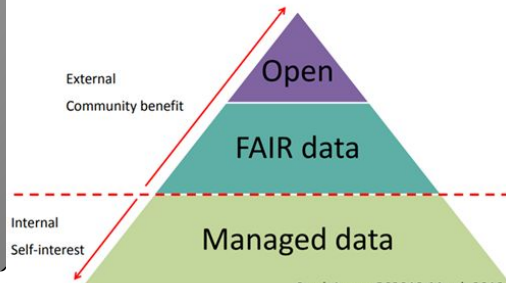
Open science

- fare Open Science significa **rendere aperto ogni passo della ricerca**
- il miglior esempio? Http. Nato come protocollo interno, poi aperto a tutti
- i principi della Open Science sono **trasparenza, riproducibilità, collaborazione, inclusività, accessibilità, rigore, riuso**
- ... in una parola, **la scienza, fatta bene!**
- la scienza aperta si basa anche sull'idea che la **ricerca finanziata con fondi pubblici debba essere pubblicamente disponibile**: “ogni cittadino ha diritto di trarre vantaggio dalla ricerca finanziata con soldi pubblici, poiché provengono anche dalle sue tasse” [Neelie Kroes, Commissione Europea]
- la **Commissione Europea** ha compiuto una **scelta decisa a favore** della Open Science

...per una ricerca
efficiente, i dati

1. vanno gestiti
2. vanno resi FAIR
3. se possibile,
vanno aperti

How do Open, FAIR & RDM intersect?



Sarah Jones, OS2018, March 2018

I dati vanno gestiti

Cosa sono i dati? “ le prove utilizzare per dare informazioni su o per supportare le conclusioni scientifiche” [[Sheffield](#), 2018]

Una buona gestione dei dati comprende:

1. corretta descrizione, organizzazione e scelta del formato (dall'inizio della ricerca)
2. corretta metadatazione [esempio di [set di metadati](#) per singole discipline]
3. conservazione, durante il progetto e sul lungo periodo (esigenze diverse)... perché i dati sono fragili, e si perdono [80% dati delle pubblicazioni degli ultimi 20 anni non è più disponibile, [Gibney 2013](#)]

Se avete 4 minuti, guardate questo [video](#): tutti gli errori di una scorretta gestione dei dati

La gestione dei dati passa da un buon Data Management Plan!

- DMP è un documento che obbliga a pensare ai dati in modo strutturato
- è uno strumento “vivo” che va aggiornato periodicamente
- esistono checklist per farvi le giuste domande sui vostri dati ([DCC](#), [IOOSG](#))
- ... e uno strumento preziosissimo per gestirli bene: [Data Stewards Wizard](#)
- esistono modelli di DMP online
- ...per saperne di più, appuntamento al 28 novembre, 14.30 con il 4. webinar

PER IMPARARE A GESTIRE I DATI:

- corso [Essentials4data](#)
- [Open Science training handbook](#), un manuale per ogni aspetto della Open Science
- Data Management ([DDC](#), [ANDS](#), [DANS](#))

I dati vanno resi FAIR

- FAIR significa Findable, Accessible, Interoperable, Reusable
- per conoscere [i principi](#) e [come applicarli](#), vi invitiamo al prossimo webinar, il 20 novembre alle 12.00
- significa in sostanza applicare i corretti metadati, formati, descrizioni e licenze per rendere i vostri dati riusabili
- la [EOSC, European Open Science Cloud](#) che nascerà il prossimo 23 novembre, si baserà sui dati FAIR. La Commissione Europea investe 4 miliardi e mezzo in EOSC. Tutta la ricerca europea - e l'innovazione - passerà di lì. Forse conviene esserci...
- ...i vostri dati sono già FAIR? scopritelo con questa [checklist](#)

Se possibile, i dati vanno aperti

- ... perché la “A” di FAIR significa solo “Accessibile”, non “aperto”: è necessario sapere dove trovare il dataset
- il principio adottato dalla Commissione Europea è **“as open as possible, as closed as necessary”**
- lo stesso principio andrebbe adottato per ogni dataset, indipendentemente dal fatto di essere finanziato in H2020. Una volta che il dataset sia stato utilizzato dal suo creatore, può essere reso pubblico
 - depositandolo in un archivio aperto come Zenodo, che assegna un DOI
 - pubblicandolo in un [Data Journal](#), una rivista che pubblicano solo dataset+scheda riassuntiva
 - pubblicandolo come Supplementary material dell’articolo relativo
- Il primo requisito per rendere aperti i dati è ovviamente aprirli anche dal punto di vista giuridico, rilasciandoli con una [Licenza Open](#).

Dati e diritti / 1

- **“Ripeti con me: i dati della ricerca non sono miei, i dati della ricerca non sono miei”.** Così Wainer Lusoli, funzionario della Commissione Europea, introduce il tema dei diritti sui dati.
- Il diritto d'autore (L.633/1941) non si applica infatti ai dati grezzi, poiché il diritto tutela la specifica forma espressiva con cui l'informazione è presentata, non l'informazione in sé (art. 1 e 2).
- Esiste poi un “diritto sui generis” sulle banche dati, che tutela lo sforzo economico del costituente della banca dati, ed è regolamentato dalla Direttiva 96/9/EC, ripresa dagli articoli 102 bis e 102 ter della Legge sul Diritto d'autore (L.633/1941). Esiste quindi una tutela non sui dati in sé, bensì sui dati quando essi siano aggregati e organizzati in una banca dati.

Dati e diritti / 2

Una licenza è uno strumento giuridico attraverso il quale il detentore dei diritti permette alcuni utilizzi della propria opera; senza licenza, e data la complessità dei principi sulla tutela dei dati, eventuali riutilizzatori si troverebbero in una zona grigia del diritto, senza sapere cosa poter fare.

Poiché sui dati grezzi non esiste un vero diritto d'autore, lo strumento più appropriato è la [Creative Commons Zero](#) (CC0): il costituente della banca dati rinuncia all'esercizio dei suoi diritti di tutela, rilasciando quindi i dati in un regime di pubblico dominio, riutilizzabili da chiunque senza alcun vincolo.

Dati e diritti / 3

Con la Licenza CC0, ogni tipo di riutilizzo dei dati è permesso. Ciò non toglie che chi riutilizza un dataset debba citare la fonte, secondo ogni buona prassi scientifica. Per aiutare gli utenti, è utile inserire una citazione già predisposta che possa essere copiata per attribuire il credito dovuto.

L'adozione di una licenza [Creative Commons Attribution](#) (CC BY), che prevede l'attribuzione, può non essere corretta né legalmente sostenibile, proprio perché i dati non sono tutelati dal diritto d'autore. A maggior ragione non sono opportune le licenze [NC – Non Commercial e ND – No Derivatives](#), che di fatto bloccano il riutilizzo dei dati

Perché aprire i dati?

I dati correttamente gestiti con un Data Management Plan e poi aperti

- rendono la vostra ricerca più trasparente
- favoriscono la riproducibilità
- rendono i vostri dati accessibili (anche se non aperti)
- fanno risparmiare tempo al momento della redazione di un articolo
- riducono il rischio di perdere i dati
- facilitano la condivisione e il riuso (non dimentichiamo che si tratta di dati prodotti con fondi pubblici)
- aumentano le citazioni dell'articolo corrispondente

[Drachen, T.M. et al., (2016). Sharing data increases citations. LIBER Quarterly. 26(2), pp.67–82]

**Gestire i dati e
renderli aperti
favorisce
trasparenza,
riproducibilità
e riuso**



Cosa prevede Horizon2020 per i dati

- riferimento normativo: [Art. 29.3 Annotated Model Grant Agreement](#), p. 243 e 248
- inizialmente era previsto solo un progetto pilota per 9 discipline; da gennaio 2017 il progetto pilota (Open Research Data Pilot) è stato esteso a tutti i progetti finanziati, e l'opt-in è diventato default
- si applica
 - ai dati che validano quanto pubblicato nell'articolo scientifico (cosiddetti 'underlying data')
 - ogni altro tipo di dataset specificato nel Data Management Plan
- c'è la possibilità di opt-out per ragioni di
 - sicurezza
 - incompatibilità con la normativa di protezione dei dati personali
 - rischio che gli obiettivi del progetto non vengano ottenuti
 - mancanza di dati prodotti dal progetto stesso

La premessa: il Data Management Plan

Horizon 2020 richiede che ogni progetto si assicuri che vengano adottate le misure atte a permettere a terze parti

“accesso, estrazione, sfruttamento, riproduzione e disseminazione”

Come prima cosa è necessario stilare il DATA MANAGEMENT PLAN

- per ogni progetto va stilato un Data Management Plan relativo alla gestione di tutti i dataset che si produrranno (se ci sono esigenze diverse, produrre diversi DMP)
- DMP è un deliverable e va consegnato entro M6
- va aggiornato periodicamente se le condizioni del progetto lo richiedono

I dati in H2020: 3 passi

1. DEPOSITO: i dati vanno depositati in un repository “trusted”
 - Zenodo è perfetto perché è già connesso con OpenAIRE
 - se ci sono repositories disciplinari già utilizzati è meglio depositare lì
 - su [Re3data](#) si trovano repositories per ogni disciplina
2. OPEN ACCESS: i dati vanno resi aperti
 - si può imporre embargo a discrezione del progetto
 - occorre associare una licenza CC più aperta possibile: CC0 o CC-BY
3. DOCUMENTAZIONE: occorre fornire ogni informazione utile su come è stato costituito il dataset e possibilmente DEPOSITARLI
 - software ([GitHub](#)), protocolli ([protocols.io](#)), algoritmi ([MyExperiment](#))

...avete mai provato?

- a utilizzare un Open Lab Notebook? Si tratta di notebook digitali in cui potete inserire
 - testo
 - dati
 - codice eseguibile
 - data visualization

... quindi, al momento di scrivere l'articolo, avete tutto pronto. E in più, potete condividere direttamente l'intero notebook, aumentando la trasparenza sul vostro lavoro. [Per saperne di più...](#)

- a pre-registrare l'esperimento? Anche questo contribuisce alla trasparenza e vi assegna una priorità intellettuale. Provate [AsPredicted](#) o [OSFRegistry](#)

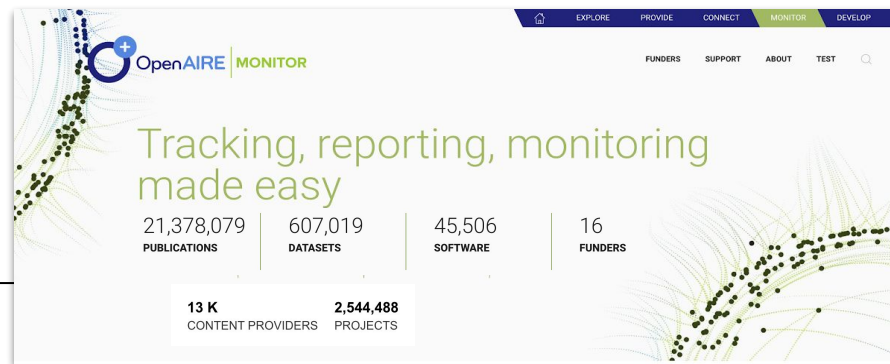
... e la visibilità?

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

e sono presenti servizi utili alla ricerca



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)

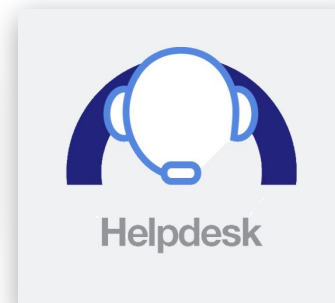
↓ Get EC report (CSV)

↑ 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: Scolexplorer
- Help Desk Nazionale: NOAD (noad-it@openaire.eu)



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](#)
- [Cascading Grants RDA Europe 4.0](#):
 - Early Career
 - Experts
 - Adoption Grants



**I dati in
Horizon2020
vanno resi il più aperti
possibile per
permettere il riuso e
favorire
innovazione**

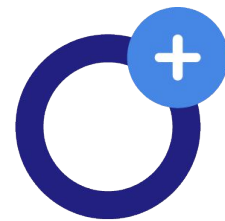


Grazie!

elena.giglia@unito.it
emma.lazzeri@isti.cnr.it



RESEARCH DATA ALLIANCE
EUROPE



OpenAIRE

