
Abbracciare il Futuro. Fare Scienza Aperta oggi

Dati della ricerca: deposito accesso riuso

Ferrara 11 Giugno 2024

Paolo Del Fiore
Istituto Oncologico Veneto IOV-IRCCS
Padova



Come si definiscono i dati della ricerca

Risultati
esperimenti
digitale
scienza

informazioni
ricerca

Research data is any information that has been collected, observed, generated or created to validate original research findings. Although usually digital, research data also includes non-digital formats such as laboratory notebooks and diaries. Leeds University
<https://library.leeds.ac.uk/info/14062/research-data-management/61/research-data-management-explained>

Sono le **informazioni**, in qualsiasi formato (digitale e/o cartaceo, numerico, descrittivo, audio o video), raccolte e utilizzate durante un'attività di ricerca, necessarie a validare i risultati raggiunti.
<https://www.sns.it/it/guida/la-gestione-dei-dati-della-ricerca>

Data, in its simplest form, refers to information that can be collected, stored, and analyzed. It can take various forms, including numbers, text, images, and more. Virtually anything can be data: measurements, simulations, books, transactions, diaries, musical scores, recipes, X-Rays, historical logs, algorithms, geographies. It can be common to think of data as only hard numbers or categorical data, but data does include so much more. <https://geo-data-support.sites.uu.nl/what-is-research-data/>

The recorded information (regardless of the form or the media in which it may exist) necessary to support or validate a research project's observations, findings or outputs, or which is required for legal or regulatory compliance.
<https://researchdata.ox.ac.uk/research-data-management-policy-full-text>

Tutte le informazioni, in qualsiasi formato, create o utilizzate secondo protocolli o metodologie definiti nell'ambito di una specifica attività di ricerca. A titolo meramente esemplificativo, sono da intendersi dati della ricerca i risultati (positivi o negativi) di tutti gli esperimenti rilevanti per la ricerca stessa, le osservazioni, le esperienze, le fonti edite e/o inedite, i riferimenti bibliografici, i software e i codici, i testi, gli oggetti, del presente o del passato, raccolti o creati in formato digitale e/o cartaceo. I dati della ricerca possono essere espressi sotto forma di simbolo, numero, testo, immagine, audio, video, etc., e possono essere grezzi o elaborati.
<https://www.polito.it/sites/default/files/2024-02/PolicyDatiRicerca%202024.pdf>

Research data is any information that has been collected, observed, generated or created to validate original research findings. Research data may be arranged or formatted in such a way as to make it suitable for communication. Data interpretation and processing comes in many formats, both digital and physical.
<https://guides.library.unisa.edu.au/ResearchDataManagement/WhatIsData>

... una definizione...

Representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship

Digital curation centre (DCC) <https://www.dcc.ac.uk/>

Rappresentazioni
osservazioni
entità
evidenza
fenomeni
studio ricerca

Perché è importante **depositare** i dati della ricerca?

L'accesso ai dati della ricerca scientifica favorisce il progresso scientifico, la riproducibilità e la validazione delle ricerche, riduce le duplicazioni e incrementa la trasparenza

<https://openscience.unife.it/depositare-i-dati-della-ricerca-1/depositare-i-dati-della-ricerca>

Ricerca
Finalizzata

Programmazione
triennale IRCSS

Rendicontazione

PRIN
Programmi
europei

H2020

SIR

Open
science

Ricerca corrente
interoperabilità

Deposito da **discrezionale** a **necessario**

discrezionale consapevolezza che dati attivi soggetti a modifiche nel corso della ricerca andavano adeguatamente **strutturati** e **annotati** per garantire una corretta **conservazione** e consentire una facile scelta sulla **pubblicazione** **archiviazione** o **scarto**

necessario al fine di rendicontare e validare i prodotti della ricerca (programma europeo H2020, Ricerca Finalizzata, SIR, PRIN e la Programmazione Triennale 2022-2024) rendere **accessibili** i dati della ricerca attraverso il **deposito** e la condivisione dei dati all'interno della comunità scientifica.

Descrizione e punti salienti della procedura di **deposito**

RUOLI E RESPONSABILITA' DEL DEPOSITO

SELEZIONE E PREPARAZIONE DATI

ANONIMIZZAZIONE E PROTEZIONE DATI SENSIBILI

SCELTA DEL REPOSITORY PUBBLICO

ACCESSO

RIUSO

Ruoli e responsabilità del deposito

Il **Principal Investigator** (PI) di un Progetto di Ricerca (studio) è la persona responsabile dell'intero processo di **selezione estrazione anonimizzazione deposito** e **condivisione** dei dati. Generalmente ricopre anche una posizione rilevante all'interno della pubblicazione scientifica (first author, last author, corresponding author) legata allo studio . Il processo può essere svolto anche da una persona incaricata dal PI, generalmente uno dei collaboratori principali (co-PI)

I ricercatori, in particolare coloro che usufruiscono di finanziamenti pubblici, sono invitati a redigere un piano di gestione dei dati della ricerca (data management plan) e a prendere tutte le misure necessarie per favorire la condivisione dei dati e il loro riuso secondo i principi identificati dall'acronimo FAIR (findable, accessible, interoperable, reusable)

Selezione e preparazione dei dati I.

Aspetti Legali , Protezione dati sensibili, minimizzazione

Il/La PI verifica se ha il diritto di depositare i dati dello studio, considerando eventuali:

- Accordi contrattuali esistenti (es: studi profit sponsorizzati da case farmaceutiche)
- Diritto di terzi (es: nell'ambito di studi multicentrici o collaborazioni)
- Obblighi di riservatezza (es: studi profit)
- Approvazioni del comitato etico
- Requisiti di finanziamento (es: alcuni enti finanziatori richiedono espressamente il deposito dei dati)

Per quanto riguarda la protezione dei dati personali, il/la PI verifica che il dataset rispetti le prescrizioni che seguono:

- il dataset non deve contenere dati direttamente riconducibili agli/alle interessati/e: nome e cognome, codice fiscale, data di nascita, dati di contatto, numero di cartella e qualsiasi dato che possa contribuire ad una identificazione diretta.
- il dataset deve contenere soltanto i dati strettamente necessari alla ricerca che si intende pubblicare, nel rispetto del principio di minimizzazione.
- il/la PI provvederà, inoltre, alla pseudoanonimizzazione dei dati contenuti nel dataset da depositare

Selezione e preparazione dei dati IIa.

Preparazione dei dati

A seconda della tipologia di studio, i dati presentarsi in diverse forme, possono essere costituiti da file separati nel formato in cui vengono scaricati da uno strumento (per esempio i file di output di un sequenziatore), da immagini, o possono essere raccolti in una matrice organizzata (dataset).

Si consiglia di rendere i formati dei files il più semplice possibile, si veda ad esempio la pagina della MIT Libraries (<https://libraries.mit.edu/data-management/store/formats/>) per indicazioni riguardo la scelta dei formati in cui depositare i files

I files devono essere conservati in un archivio atto al compito, ovvero:

- accessibile solo da personale autorizzato
- disponibile solo da rete aziendale tramite connessione protetta
- dotato di un'adeguata protezione da attacchi e data breach
- logicamente organizzato in modo preciso e funzionale
- con backup periodici pianificati

Selezione e preparazione dei dati IIb.

Preparazione dei dati

Quando i dati delle ricerche sono raccolti in un dataset, questo dovrà essere 'pulito' e 'ben annotato', in particolare il/la PI dovrà:

- effettuare una revisione sistematica di duplicati, errori e dati incompleti;
- etichettare i dati in modo coerente (ad es: denominazione dei file, nomi delle variabili, codifica dei dati mancanti);
- fornire informazioni sufficienti su ciascuna variabile per consentire la replica dei risultati presentati nella pubblicazione associata al dataset.

Scelta del repository

Dopo aver preparato i files o il dataset , il PI sceglierà un appropriato repository in cui depositare i dati purché sia di pubblico accesso.

I dati della ricerca possono essere depositati in archivi pubblici specifici della disciplina e riconosciuti dalla comunità scientifica (vedasi l'elenco fornito da Nature:

<https://www.nature.com/sdata/policies/repositories#general>

In alternativa, il PI potrà scegliere un **repository generalista** per il deposito dei dati, a questo riguardo le varie istituzioni di ricerca hanno individuato degli archivi open access di riferimento (lo IOV ha scelto Zenodo).

È **raccomandabile** effettuare l'operazione di **deposito** prima della submission del manoscritto ad una rivista in modo da poter inserire nel manoscritto anche il DOI che identifica in modo univoco e permanente il dataset. È comunque possibile eseguire l'operazione di deposito dei dati anche successivamente, a pubblicazione edita.

Scelta del repository pubblico

Zenodo e gli altri...

I repository di dati sono specializzati nella conservazioni di dati e metadati. I repository generalisti come [Zenodo](#), sono repository polivalenti adatti a tutti i prodotti della ricerca (compresi presentazioni, poster, immagini e dunque dati, codice e ovviamente articoli). Di solito in assenza di un repository di riferimento si fa uso di questo tipo di archivi generalisti. [Zenodo](#) per esempio non richiede l'uso di formati, dimensioni, licenze o altro, e ha metadati generici (<https://open-science.it/it/article?rpk=67443>).

Repository Name	Information on fees/costs	Size limits	Integrated with Scientific Data's manuscript submission system	Re3data / FAIRsharing entry
Dryad Digital Repository	\$120 USD for first 20 GB, and \$50 USD for each additional 10 GB	None stated	Yes ✓	view FAIRsharing entry
figshare	100 GB free per Scientific Data manuscript.	1 TB per dataset	Yes ✓ - To qualify for the 100 GB of free storage, data must be uploaded to figshare via our submission system. Download instructions.	view FAIRsharing entry
Harvard Dataverse	Contact repository for datasets over 1 TB	2.5 GB per file, 10 GB per dataset	No	view re3data entry
Open Science Framework	Free of charge	5 GB per file, multiple files can be uploaded	No	view FAIRsharing entry
Zenodo	Donations towards sustainability encouraged	50 GB per dataset	No	view re3data entry
Science Data Bank	Free of charge	8 GB per file, no limit to dataset size	No	view FAIRsharing entry



Guida all'inserimento dei metadati

Zenodo permette di riservare un identificativo univoco (DOI) al dataset; il DOI può essere utilizzato dal PI all'interno della pubblicazione da inviare all'editore per citare correttamente il dataset nell'articolo.

...di registrare una community istituzionale dove i ricercatori possono depositare i dati secondo linee guida interne condivise con la Direzione Scientifica (*Procedura Raccolta, deposito e condivisione dei dati della ricerca prodotti dall'Istituto Oncologico Veneto , Padova 9 Ottobre 2023*), il PI Per depositare richiederà di registrare i dati da depositare all'interno della Community, gestita da un moderatore interno istituzionale (**GSRD** - Gruppo supporto Raw Data IOV).



Guida all'inserimento dei metadati

Il **GSRD** riceverà una e-mail da **Zenodo** in cui viene notificata la richiesta di inserimento di un dataset all'interno della **Community** e ne consentirà l'inclusione purché siano soddisfatti i requisiti seguenti:

- il PI deve essere personale IOV o in convenzione
- il PI deve concedere al moderatore l'accesso ai RD perché possa verificare che i file caricati in Zenodo siano conformi ai requisiti di cui sopra
- i devono essere completi e devono descrivere in modo esaustivo le variabili dello studio Metadati

Se una di queste caratteristiche viene a mancare, il moderatore contatta il/la PI per richiedere la correzione delle informazioni e, quando queste saranno corrette, il dataset caricato dal/dalla **PI** in **Zenodo** sarà accettato all'interno della **Community**.

1

Per visualizzare la **community IOV** vai direttamente al link <https://zenodo.org/communities/iov/>

Oppure vai su <https://zenodo.org>, clicca su *Communities* e digita «Veneto Institute of Oncology» all'interno del campo dedicato, quindi clicca su *view*

In questa community sono presenti i dataset dell'Istituto Oncologico Veneto, depositati su Zenodo.

2

Clicca su *New upload* per procedere al caricamento del dataset



3

Qui puoi eseguire il login, oppure registrarti a Zenodo se non hai ancora un'utenza



Per procedere con il caricamento di un dataset, puoi seguire queste indicazioni:

Files:

Scegli uno o più file da caricare (*Choose files*).
Dopo clicca su *Start upload*

Communities:

Digita nel campo dedicato la parola «veneto» e seleziona la community *Veneto Institute of Oncology*

The screenshot shows the 'New upload' page. At the top, there are instructions and a 'Files' dropdown menu. Below this is a large area with the text 'Drag and drop files here' and a 'Choose files' button (1). To the right of this area are two buttons: 'Choose files' (2) and 'Start upload' (3). Below the main area is a 'Communities' section with a search bar (4) and a list of communities. The 'VENETO INSTITUTE OF ONCOLOGY Padova - Italy' community is highlighted with a blue oval.

Attenzione: è obbligatorio compilare i campi indicati con l'asterisco *

Digital Object Identifier:

Il DOI verrà assegnato automaticamente da Zenodo, puoi lasciare questo spazio vuoto. Se invece vuoi che ti venga riservato un DOI (ad esempio, vuoi citare il dataset all'interno del paper che stai scrivendo), puoi cliccare su *Reserve DOI*.

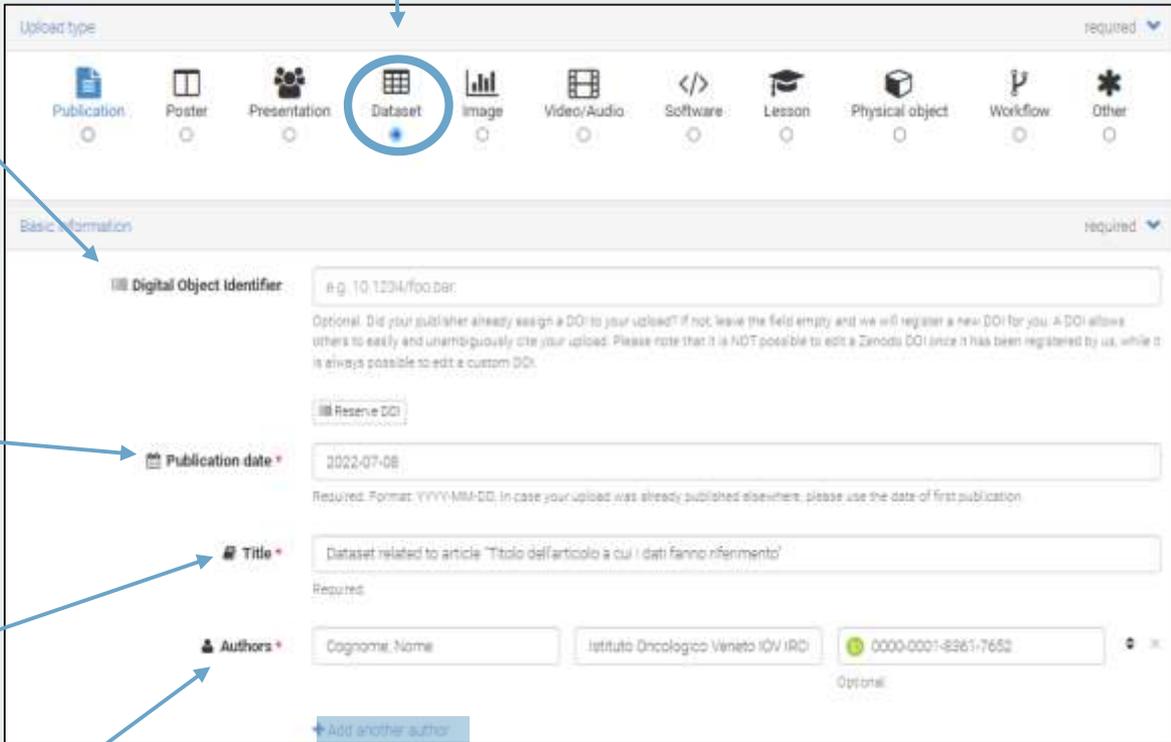
Publication date:

Indica la data di pubblicazione del dataset, puoi lasciare quella proposta automaticamente

Title:

Indica il titolo del dataset. Ad esempio: «Dataset related to article XYZ» oppure «Raw data related to article XYZ»

Upload type:
Seleziona la tipologia *Dataset*



The screenshot shows the Zenodo upload form. The 'Upload type' section has a dropdown menu set to 'required' and a row of icons for different file types: Publication, Poster, Presentation, Dataset (circled in blue), Image, Video/Audio, Software, Lesson, Physical object, Workflow, and Other. Below this is the 'Basic information' section, also set to 'required'. It contains several fields: 'Digital Object Identifier' with a placeholder 'e.g. 10.1234/foo.bar' and a 'Reserve DOI' button; 'Publication date' with a date picker set to '2022-07-08' and a 'Required' note; 'Title' with a text input containing 'Dataset related to article "Titolo dell'articolo a cui i dati fanno riferimento"' and a 'Required' note; and 'Authors' with a text input for 'Cognome, Nome', a dropdown for 'Istituto Oncologico Veneto IOV IRCC' and ORCID, and a text input for '0000-0001-8361-7652' with an 'Optional' note and an 'Add another author' button.

Authors:

Indica il tuo nominativo (Cognome, Nome) e se vuoi quello degli/delle autori/autrici dello studio, con affiliazione e ORCID.

Description:

Descrivi brevemente il contenuto del dataset. Trovi alcuni esempi di descrizioni [qui](#)

Version:

Se lo desideri, puoi indicare la versione del dataset caricato. A volte i dataset vengono aggiornati (caricate versioni successive alla prima, aggiornate), in questo caso è opportuno indicarlo.

Language:

Puoi digitare il nome della lingua utilizzata nel documento (english, italian, french...) e selezionare l'idioma dal menù a tendina oppure indicarne la forma abbreviata (eng, ita, fr...)

Keywords:

Puoi indicare le parole chiave rappresentative del dataset. Per aggiungere più di un termine dovrai cliccare su *+Add another keyword*

The image shows a screenshot of the Zenodo dataset upload form. The form is divided into several sections, each with a label and a corresponding input field. The sections are: Description, Version, Language, Keywords, and Additional notes. The Description field is a large text area with a rich text editor toolbar. The Version field is a text input with a dropdown arrow. The Language field is a text input with a dropdown arrow. The Keywords field is a text input with a dropdown arrow and a '+ Add another keyword' button. The Additional notes field is a text input with a dropdown arrow. The form is titled 'Description' and has a 'Required' label. The 'Additional notes' field is labeled 'Optional'.

Additional notes:

Puoi indicare altre informazioni che ritieni rilevanti per la descrizione del dataset e che non hai potuto indicare nei campi precedenti. Ad esempio la tua mail per essere contattato/a oppure il riferimento a progetti di enti finanziatori che non sono ancora tracciati in Zenodo (RC, RF, 5 x 1000...)

Accesso ai dati

Non tutti i dati sono adatti alla condivisione

Tuttavia, anche se i dati non possono essere condivisi apertamente o immediatamente, ciò non significa che non possano essere condivisi affatto.

Ragioni per non rendere i dati disponibili al pubblico

- natura stessa dei dati (ad esempio, i dati sono riservati o altrimenti sensibili)
- natura del processo di ricerca (ad esempio, i ricercatori potrebbero essere ancora al lavoro sull'analisi primaria o potrebbero essere intenzionati a richiedere un brevetto).

Anche se i dati sono molto sensibili e sarebbe necessario un accordo di condivisione dei dati personalizzato per qualsiasi riutilizzo, è meglio, se possibile, che questo processo sia mediato da un archivio specializzato, piuttosto che affidare la responsabilità ai singoli ricercatori (archivi dispongono di procedure per inviare le richieste ai creatori dei dati)

Accesso ai dati licenze di utilizzo

License required ▾

Access right *

-  Open Access
-  Embargoed Access
-  Restricted Access
-  Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

Open Access è liberamente accessibile a chiunque abbia una connessione internet, senza restrizioni

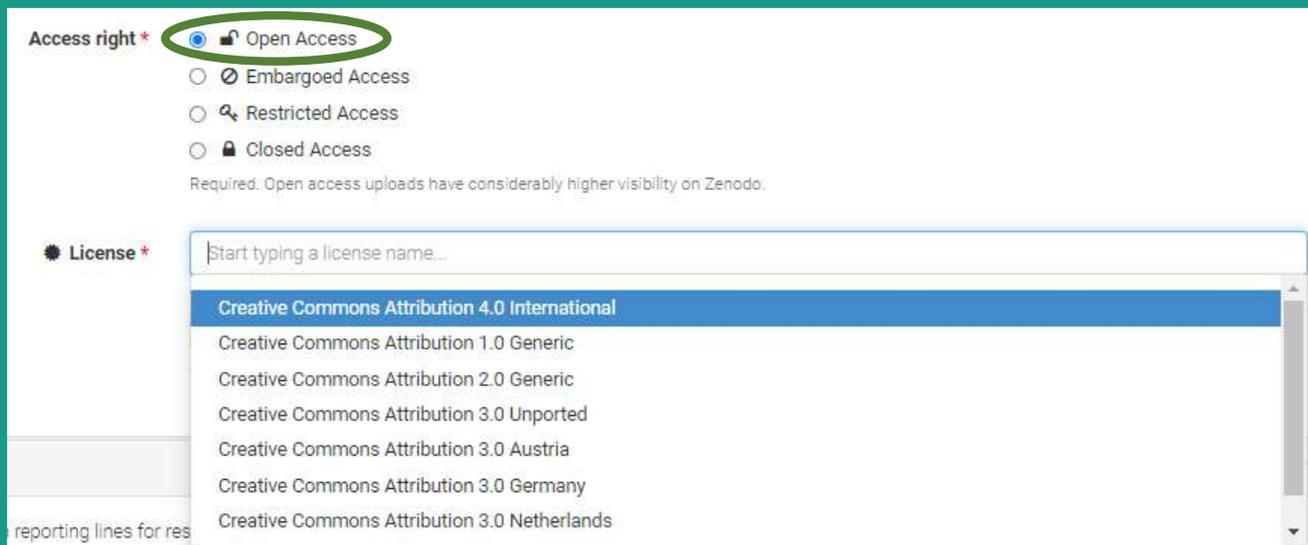
Embargoed Access è accessibile dopo un periodo di embargo

Restricted Access è accessibile solo su autorizzazione

Closed Access non accessibile a terzi

Open access

Open Access il dataset caricato è liberamente accessibile a chiunque abbia una connessione internet, senza restrizioni; scegliendo la licenza *Creative Commons Attribution 4.0 International*. In questo modo chi utilizzerà questo dataset sarà tenuto/a a riconoscere una menzione di paternità al dataset che hai condiviso.



The screenshot shows a form with two main sections: 'Access right' and 'License'. The 'Access right' section has four radio button options: 'Open Access' (selected and circled in green), 'Embargoed Access', 'Restricted Access', and 'Closed Access'. Below these is a note: 'Required. Open access uploads have considerably higher visibility on Zenodo.' The 'License' section has a search input field with the placeholder text 'Start typing a license name...'. A dropdown menu is open below the input, listing several Creative Commons Attribution licenses, with 'Creative Commons Attribution 4.0 International' highlighted in blue. The list includes: 'Creative Commons Attribution 1.0 Generic', 'Creative Commons Attribution 2.0 Generic', 'Creative Commons Attribution 3.0 Unported', 'Creative Commons Attribution 3.0 Austria', 'Creative Commons Attribution 3.0 Germany', and 'Creative Commons Attribution 3.0 Netherlands'. At the bottom left, there is a partial label 'reporting lines for res'.

Embargoed Access

Embargoed Access il dataset caricato è accessibile dopo un periodo di embargo; si dovrà indicare anche la data in cui il dataset diventerà accessibile e il tipo di licenza applicata (ad es Creative Commons Attribution 4.0 International, come per l'opzione Open Access)

Access right *

Open Access

Embargoed Access

Restricted Access

Closed Access

Required: Open access uploads have considerably higher visibility on Zenodo.

Embargo date

2023-06-23

Required only for Embargoed Access uploads. Format: YYYY-MM-DD. The date your upload will be made publicly available in case it is under an embargo period from your publisher.

License *

Creative Commons Attribution 4.0 International

Required: Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the *Other* licenses available (*Other (Open)*, *Other (Attribution)*, etc.). The supported licenses in the list are harvested from opendefinition.org and spdx.org. If you think that a license is missing from the list, please contact us.

Restricted access

Restricted Access il dataset caricato non è visibile/utilizzabile da terzi ed è accessibile solo previa tua autorizzazione. Nel campo **Conditions** vanno specificate le condizioni in base alle quali si concede agli/alle utenti l'accesso al dataset. Gli/le interessati dovranno spiegare il motivo della loro richiesta, e il creatore dei dati potrà decidere se concedere o meno l'accesso al dataset.

Access right *

- Open Access
- Embargoed Access
- Restricted Access
- Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

Conditions *



Specify the conditions under which you grant users access to the files in your upload. User requesting access will be asked to justify how they fulfil the conditions. Based on the justification, you decide who to grant/deny access. You are not allowed to charge users for granting access to data hosted on Zenodo.

Closed access

Closed Access il dataset caricato non è visibile/utilizzabile da terzi. Non devono essere specificate ulteriori informazioni.

Access right *

- Open Access
- Embargoed Access
- Restricted Access
- Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

Riuso

cosa significa.....

Utilizzo di dati della ricerca per scopi diversi da quelli per cui sono stati originariamente raccolti. Il riutilizzo dei dati è particolarmente importante in ambito scientifico, in quanto consente a diversi ricercatori di analizzare e pubblicare risultati basati sugli stessi dati indipendentemente l'uno dall'altro. Il **Riuso** è una componente chiave dei principi **FAIR**

Il **Riuso** dei dati, o analisi dei dati secondari, è l'analisi di dati esistenti raccolti da altri individui o istituzioni per un nuovo scopo di ricerca. Può riferirsi a dati statistici e quantitativi o a dati descrittivi e qualitativi

Riuso

caratteristiche dei dati

I dati della ricerca devono essere gestiti secondo i principi **FAIR** (Findable, Accessible, Interoperable, Reusable), per rendere la conoscenza facile da rintracciare, farla circolare e favorire l'innovazione.

Per rendere i dati Findable / Rintracciabili:

- assegnazione al dataset di un identificatore persistente (es. DOI, handle, ecc.)
- i metadati devono essere esaustivi, accurati e indicizzati dai motori di ricerca

Per rendere i dati Accessible / Accessibili:

- i metadati devono essere pubblici, visibili e indicizzabili anche se i dati non sono ad accesso aperto

Per rendere i dati Interoperable / Interoperabili:

- devono essere disponibili in formati aperti o molto diffusi
- i metadati devono seguire schemi standard riconosciuti
- devono essere presenti link con altre risorse collegate ai dati (ad es. pubblicazioni, rapporti tecnici, ecc.)

Per rendere i dati Re-usable / Riutilizzabili:

- i dati devono essere descritti in modo facile e comprensibile
- ai dati depositati deve essere attribuita una licenza con possibilità di riutilizzo

Riuso

Quali sono i vantaggi ?

Mantiene la continuità della ricerca

Fa risparmiare tempo e risorse

Favorisce nuove collaborazioni l'interdisciplinarietà

Aumenta la visibilità della ricerca

Espone i dati a nuovi strumenti, metodi e approcci

Contribuisce alla comunità scientifica e alla società

Riuso conclusioni

Sebbene uno degli obiettivi principali della condivisione dei dati della ricerca sia quello di promuovere il riutilizzo, i dati **FAIR** non sono sempre aperti.

Se la loro ricerca è facilitata da ricchi metadati e vengono forniti chiari dettagli sul processo di richiesta di accesso, allora anche un set di dati sensibili che non può essere reso disponibile apertamente può raggiungere un alto grado di **FAIRness**

Grazie per l'attenzione !

Credits:

Giuliana Prevedello e Valentina Bozzato - *Gestione Documentazione Scientifica IOV*

(Testi, Riferimenti Bibliografici e Normativi)

Giampaolo Moscardi - *UOC Chirurgia oncologica dei tessuti molli, del peritoneo e dei melanomi*

(Grafica e Formattazione)

Paolo Del Fiore

Soft-Tissue Peritoneum and Melanoma Surgical Oncology Unit

Veneto Institute of Oncology IOV-IRCCS Padova

 <https://orcid.org/0000-0002-2862-8014>

