Manuale operativo del portale dati Open Access OE - DataRep

Release 1.0

INGV Osservatorio Etneo - UFSOIT

Indice

1	Intro	Introduzione				
	1.1	Come pubblicare sul portale DataRep				
2	Proc	cesso di caricamento dati sul portale DataRep				
	2.1	Crezione della scheda descrittiva dal Metadata Editor				
	2.2	Compilazione dei <i>Metadati</i>				
	2.3	Selezione dell'elenco di file da pubblicare				
	2.4	Specificare e caricare un <i>dataset</i> di tipo Serie Temporale				
	2.5	Validazione della scheda e assegnazione del DOI (Ufficio Gestione Dati)				
	2.6	Pubblicazione sul portale DataRep				



Questo documento rappresenta un manuale operativo di utilizzo del Portale Dati Open Access **DataRep**. Indice:

Indice 1

2 Indice

CAPITOLO 1

Introduzione

Il repository **DataRep** fornisce uno strumento che permette, al personale del *Osservatorio Etneo* dell'Instituto Nazionale di Geofisica e Vulcanologia, di pubblicare datasets rispettando le politiche del pradigma **Open Access**.

Questo strumento, si occupa di creare automaticamente le pagine web: *landing pages* che ospiteranno sia il dataset che i metadati ad esso associati, rendendoli pubblicamente fruibili.

A ciascun dataset è associato un identificativo univoco detto **DOI**: Digital Object Identifier, la *landing page*, creata dal portale alla pubblicazione di un dataset, garantirà la risolvibilità del DOI.

1.1 Come pubblicare sul portale DataRep

Il processo di pubblicazione di un dataset su **DataRep** coinvolge due entità:

- 1. il **Metadata Editor** (link) collegato al **Registro Dati** istituzionale e gestito dall''**Ufficio Gestione Dati** dell''**INGV** (link2);
- 2. il portale **DataRep** (link3) dove i dataset vengono resi disponibili nel rispetto delle **FAIR** policies dettate dal paradgima Open Access.

Entrando nel dettaglio delle entità coinvolte nel processo di pubblicazione, il Metadata Editor, si occupa:

- della gestione centralizzata dei metadati che descrivono i datasets
- pubblica automaticamente la scheda del dataset sul portale DataRep
- carica i dati sul portale DataRep.

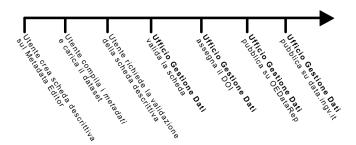
Il portale **DataRep** è un servizio web che offre alla comunità scientifica un mezzo per condividere e conservare i risultati scientifici e che ne consente la *riproducibilità* ed il *riutilizzo*:

- associa automaticamente i metadati della pubblicazione generati dal Metadata Editor alle landing pages;
- implementa la risolvibilità del DOI associato alla publicazione;
- permette l'esportazione secondo vari schemi di metadati tra cui quello, definito da **DataCite** (link4);

Manuale operativo del portale dati Open Access OE - DataRep, Release 1.0								

Processo di caricamento dati sul portale DataRep

Il processo da seguire per pubblicare *dataset* sul repository **DataRep** coinvolge due entità: il *Metadata Editor* ed il *DataRep*, ed è descritto per linee generali nella seguente figura e nei seguenti punti:



Utente

- crea la **scheda descrittiva** sul *Metadata Editor*
 - compila i *metadati*
 - seleziona i *file* che compongono il *dataset* da pubblicare
 - effettua il caricamento del dataset
- richiede la validazione della scheda

Ufficio Gestione Dati (UGD)

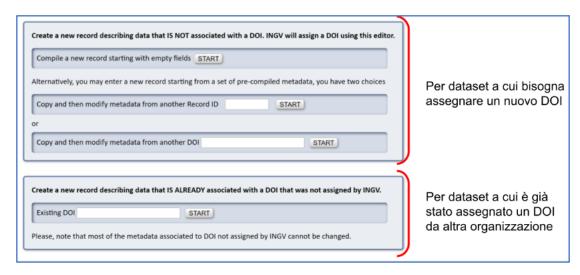
- procede alla validazione della scheda
- verifica se assegnare il *DOI*

• pubblica su DataRep

2.1 Crezione della scheda descrittiva dal Metadata Editor

L'utente si autentica sul portale del Metadata Editor utilizzando il proprio account istituzionale (nome.cognome@ingv.it come per l'accesso alla posta elettronica) al fine di **creare** una scheda descrittiva del *dataset*:

- 1. Il primo passo consiste nella *creazione* di una nuova scheda descrittiva in **bozza Drafts**, scegliendo se iniziare da una *scheda vuota* o dalla *copia* di una scheda già esistente da completare;
- 2. A questo punto il sistema consente di inizializzare una nuova bozza, permettendo all'utente di scegliere se creare un nuovo *record* che descrive il *dataset*, a cui: a) non è ancora stato assegnato nessun DOI, b) è già stato assegnato un DOI.



3. Il *web-form* verrà mostrato all'utente come un'anteprima e sarà compilabile solo in parte in quanto la scheda vera e propria non verrà creata nel sistema fino a quanto l'utente non avrà cliccato il bottone *«Create»* in fondo alla scheda.

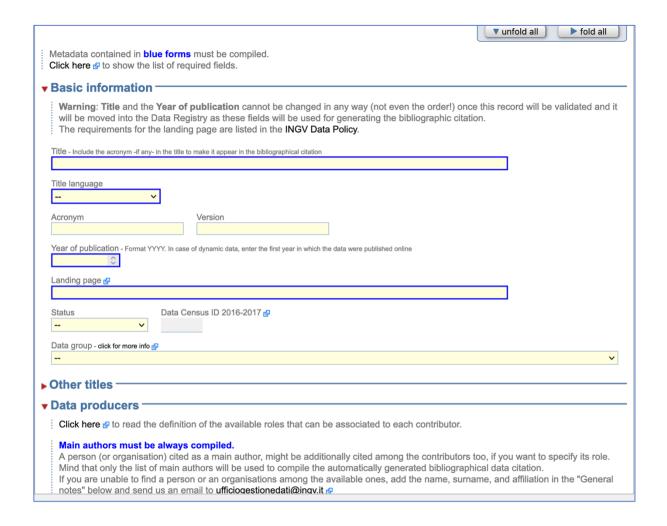
2.2 Compilazione dei *Metadati*

Scelta la modalità di creazione del nuovo record (paragrafo precedente), il sistema mostrerà un *web-form* con diversi campi da compilare che costituiscono i *metadati* caratteristici di una pubblicazione scientifica.

Solo dopo aver cliccato il bottone «Create» in fondo alla pagina la scheda sarà creata nel sistema e sarà editabile in ogni sua parte.

Quando la scheda viene creata sarà inserita nel metadata editor sotto forma di bozza (*draft*) e sarà considerata tale fino alla fine del processo di validazione, processo che verrà avviato dall'utente quando avrà terminato la compilazione della scheda.

L'utente descriverà il dataset compilando esaustivamente quanti più campi possibili tra quelli disponibili come nella scheda di esempio della seguente figura.



Nota: In particolare, il campo "Landing page" non deve essere compilato, perché questa informazione verrà restituita automaticamente dalla logica del portale DataRep non appena, verrà fatta richiesta di pubblicazione.

Nella compilazione della scheda, dovranno essere indicati anche un *referente* scientifico e uno tecnologico (le due figure possono coincidere). I referenti andranno indicati nella sezione dei metadati denominata "Info specific to INGV". Oltre ai metadati obbligatori, evidenziati in blu nella scheda (Fig. sopra), si suggerisce di compilare anche informazioni utili come la copertura temporale, geografica e soprattutto eventuali relazioni con altri dati o pubblicazioni nell'apposita sezione "Relations to other data and publications".

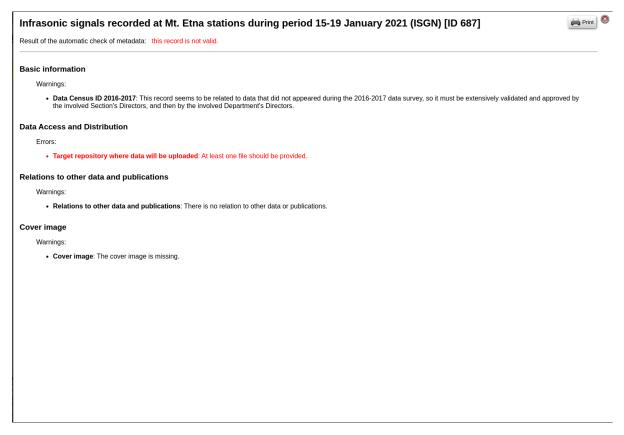
Infine, è anche possibile caricare un'immagine di copertina in fondo al *web-form* alla scritta "upload a cover image file".



Durante la compilazione dei campi è possibile verificarne la correttezza in ogni momento agendo sul bottone "Check if this record can be validated", posizionato in alto sulla form di caricamento dei metadati.



Se il processo rileva errori o campi obbligatori non compilati, tutte le anomalie saranno dettagliatamente elencate dal *Metadata Editor* e l'utente potrà procedere alla loro risoluzione.



Si consiglia di salvare frequentemente le modifiche durante la compilazione cliccando sul bottone "Save" in fondo alla scheda per evitare il rischio di perdere informazioni già inserite. La scheda può essere compilata a più riprese sia da parte dal referente scientifico che dal referente tecnologico, ma non nello stesso istante.

Se è necessario permettere ad altre persone di editare la scheda, queste possono essere aggiunte dai referenti nella sezione Additional users allowed to update metadata of this record nella sezione di metadati "Info specific to INGV".

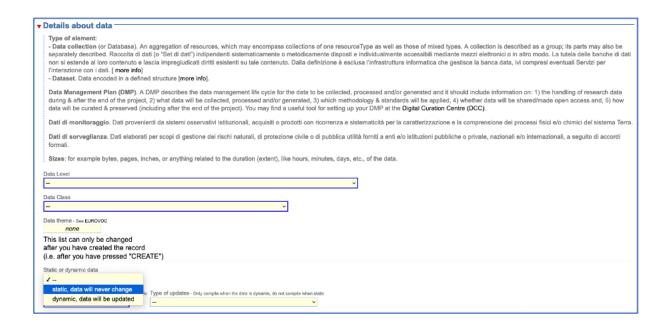
Quando la compilazione della bozza della scheda sarà ultimata e il sistema di verifica automatico non rileverà più errori, l'utente potrà avviare il processo di validazione cliccando sul bottone "Start the validation process" all'inizio della scheda.

2.3 Selezione dell'elenco di file da pubblicare

A questo punto l'utente aggiunge l'elenco dei file da caricare su DataRep:

Il Metadata Editor mostrerà le funzioni di caricamento sul portale DataRep quando le due condizioni seguenti saranno soddisfatte:

1. Il dataset deve essere dichiarato statico nella sezione di metadati "Details about data". Se necessario, in futuro sarà possibile pubblicare nuove revisioni, pubblicando un nuovo dataset seguendo la stessa procedura con l'accortezza di indicare la relazione con la versione precedente nella sezione di metadati "Relations to other data and publications". Nell'immagine in basso un esempio di selezione del tipo in statico.



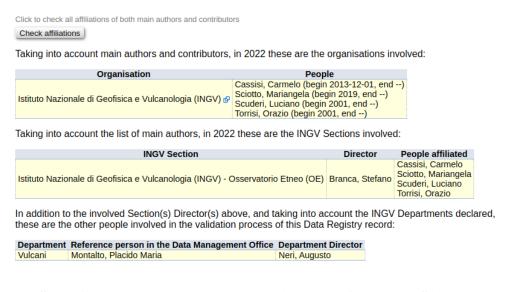
Nota: Con statico si intende che il dataset non verrà più aggiornato dopo la sua pubblicazione.

2. Almeno una persona elencata tra i "Main authors" della sezione di metadati "Data producers" deve essere affiliata all'Osservatorio Etneo.



Nota: Le funzioni per caricare dati su DataRep saranno mostrate solo dopo che l'elenco degli autori sarà compilato.

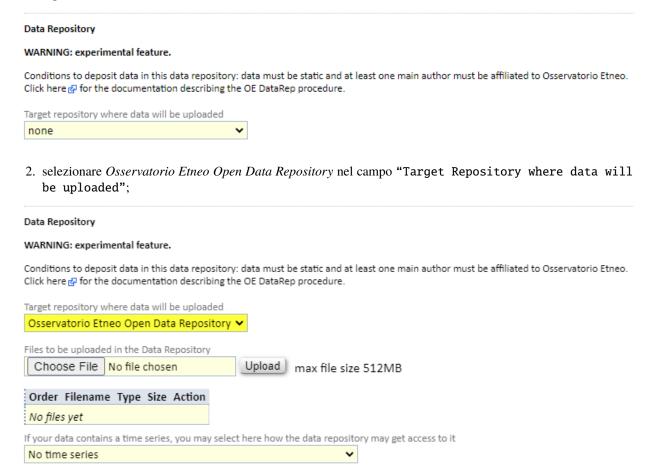
E" possibile verificare in ogni momento le affiliazioni di ciascun autore cliccando sul bottone *Check affiliations* sotto l'elenco degli autori.



Ciascun autore affiliato all'INGV può aggiornare direttamente le proprie informazioni (affiliazione compresa) modificando il proprio profilo nella sezione "People" del Metadata Editor. Le modifiche apportate ai profili personali saranno monitorate.

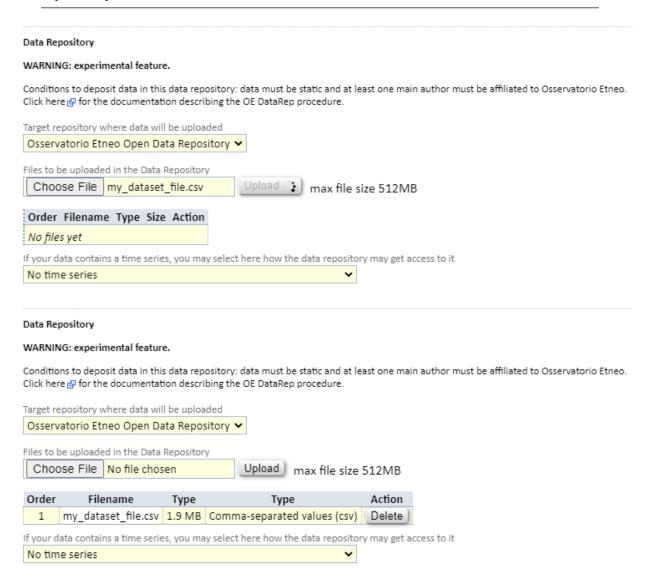
A questo punto è possibile procedere alla selezione dei file da pubblicare su DataRep. Per fare ciò:

1. espandere la sezione "Data Access and Distribution" di metadati;



3. effettuare il *caricamento* dei file che compongono il *dataset* direttamente sul *repository* OEDataRep attraverso il *form* di *upload*;

Nota: Se il *dataset* da pubblicare è una "Time Series" (serie temporale), seguire le istruzioni aggiuntive nel sottoparagrafo Specificare e caricare un dataset di tipo Serie Temporale, prima di pasare al punto 4.



4. infine cliccare sul bottone SAVE in fondo alla scheda per salvare la Bozza creata.

2.4 Specificare e caricare un dataset di tipo Serie Temporale

Se si sta caricando un dataset di tipo Serie Temporale seguire i seguenti passi aggiuntivi di configurazione:

- 1. Nel menu a tendina di selezione dedicato, scegliere tra le voci:
 - #. "Time Series from a CSV file", qualora la serie temporale è presente su uno dei file precedentemente caricati in formato CSV; Il sistema rileverà la presenza di una serie temporale all'interno di un file di tipo *csv*.
 - #. Time Series from TSDSystem (GUID required), verrà richiesto di fornire un identifico *GUID* univoco precedentemente creato dall'istanza TSDSystem (identifica una serie temporale già caricata precedentemente sul database TSDSystem da associare a questo record).
- 2. Nel *form* successivo, è possibile specificare tutti i *metadati* aggiuntivi necessari per la descrizione del *dataset* di tipo *Serie Temporale*; questi dettagli sono necessari per la *preview* grafica sulla *landing page* una volta che il dataset verrà pubblicato (vedi fase successiva validazione scheda) sul portale OEDataRep.

Avvertimento: Requisito **mandatorio** per il corretto caricamento della serie temporale sul backend del TSDSystem, è che la colonna contentente i "timestamp" deve essere nominata «**time**».

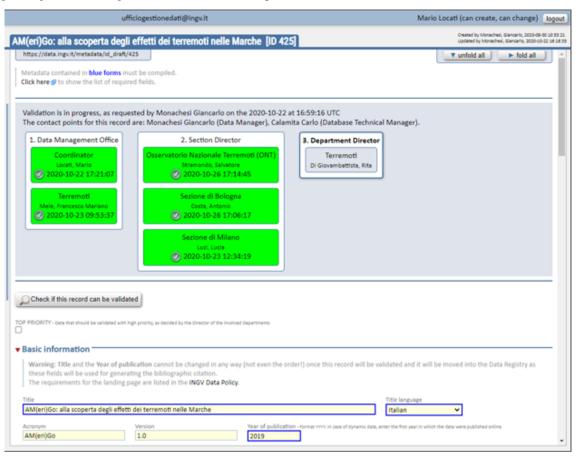
Data Repository Conditions to deposit data in this data repository; data must be static and at least one main author must be affiliated to Osservatorio Etneo. Click here @ for the documentation describing the OE DataRep procedure. Target repository where data will be uploaded OEDataRep, the Osservatorio Etneo Open Data Repository > Files to be uploaded in the Data Repository Choose File No file chosen Upload max file size 400MB Filename Type Order Action 1 my_time_series_data.xlsx 10.2 KB Microsoft Excel (xlsx) Delete 2 my_description_file.pdf 1.41 MB Portable Document Format (pdf) Delete 3 my_time_series_data.csv 1.78 KB Comma-separated values (csv) Delete If your data contains a time series, you may select here how the data repository may get access to it Time series from a CSV file my_time_series_data.csv Checks performed on the first 100 rows of the CSV files \checkmark The separator is ', (comma)' \checkmark The header is present in the first line ✓ Found 5 columns √ The first columns contains a valid Date Time ✓ The other 4 columns contains numerical values use this file for plotting a time series ♥ Description This is my time series data file Start time * 2017-01-02 03:36:13.530 Start time * 2017-01-02 03:36:13.530 End time * 2017-12-03 23:34:11.200 End time * 2017-12-03 23:34:11.200 Field name * MwPost2017 Datatype * float 🕶 Value Aggregate Gain threshold threshold Value 1 MwPost2017 Description Field name * MwInsPost2017 none 🕶 Datatype * float V Value 2 ✓ campoFinaleFloat none ✓ Description Field name * MwIntero Datatype * integer ➤ Value 3 Description Field name * campoFinaleFloat Datatype * float V Description Sampling required field

2.5 Validazione della scheda e assegnazione del DOI (Ufficio Gestione Dati)

Quando la compilazione della scheda sarà completata si potrà procedere alla **validazione** cliccando sul bottone "Start the validation process".

Avvertimento: in caso di problemi, l'Ufficio Gestione Dati si occuperà di contattare l'utente per eventuali correzioni o integrazioni.

La figura seguente mostra gli stati di avanzamento del processo di validazione.



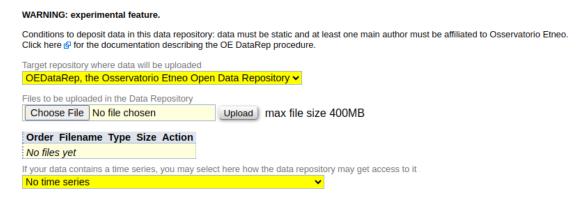
La validazione potrebbe arrivare fino a tre gradi di approvazione a seconda delle caratteristiche del dataset e degli autori indicati:

- il primo grado di validazione è in capo all" *Ufficio Gestione Dati*, nello specifico il coordinatore e un referente per ciascuno dei Dipartimenti indicati;
- il secondo grado è in capo ai *Direttori delle Sezioni* coinvolte identificate automaticamente dal Metadata Editor in base alle affiliazioni disponibili nella sezione "People" corrispondenti agli autori elencati;
- il terzo grado di validazione è in capo ai *Direttori dei Dipartimenti* indicati in "INGV Department" nella sezione di metadati "Info specific to INGV".

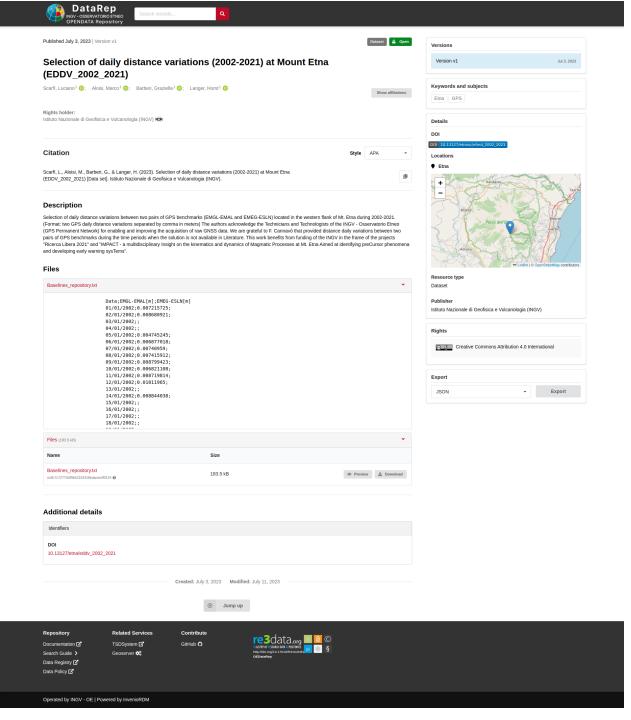
Nota: se il dataset è una nuova versione di un dataset già validato e pubblicato in precedenza, si attiverà automaticamente solo il primo grado di validazione.

2.6 Pubblicazione sul portale DataRep

L'ultimo step, nel processo di upload di un dataset sul portale DataRep consiste nel creare il record sul repository. L'interfaccia del *Metadata Editor* mette a disposizione, nella sezione "9. Data Access and Distribution", la possibilità di selezionare il repository DataRep. Una volta selezionato il repository, l'interfaccia permette di caricare i file che appartegono alla pubblicazione, come mostrato nella figura sottostante.



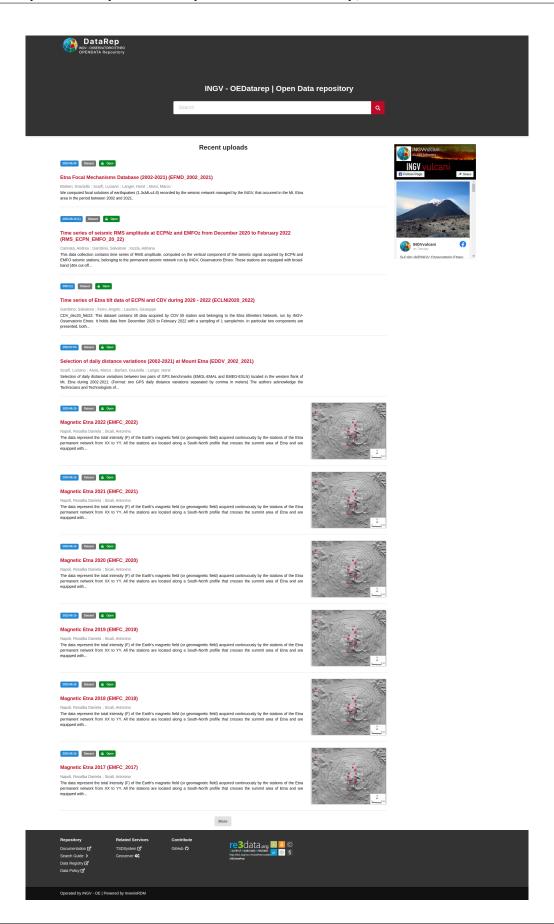
La landing-page creata dal Portale DataRep è mostrata nella figura sotto



In particolare è possibile notare come la pagina creata conterrà:

- un estratto dei metadati inseriti sul Metadata Editor,
- il DOI assegnato alla pubblicazione,
- il dataset,
- la possibilità di esportare i metadati nei diversi formati supportati dal portale.

Una email notifica gli amministratori del portale che una nuova pubblicazione è stata effettuata ed è disponibile; sarà loro compito accettarla per renderla visibile tra i contenuti recenti visualizzati nella home-page di DataRep.



Capitolo 2. Processo di caricamento dati sul portale DataRep