Biblioteca Dario Nobili
Consiglio Nazionale delle **Ricerche**
Area Territoriale di Ricerca di Bologna

dalia

Dati Aperti bibLIoteca Area bologna

Manuale d'uso

2023

Biblioteca Dario Nobili, CNR Bologna - Via Gobetti 101, 40129 Bologna, Italia
Email:dalia-info@area.bo.cnr.it Sito web: https://dalia-bo.cnr.it

CREDITS

Manuale d'uso DALIA: Agosto 2023
A cura di Gabriela Carrara (Biblioteca Dario Nobili - Area Territoriale della Ricerca CNR di Bologna), Giorgio Resci, Alessandro Frezzini, Elisabeth De Maio (INKODE SOC. COOP).

Sommario

Questo manuale è dedicato agli utenti di DALIA Data Repository e riguarda l'uso della sua interfaccia web, basata su piattaforma CKAN e personalizzata, per organizzare, pubblicare e trovare i dati.

Biblioteca Dario Nobili
Consiglio Nazionale delle Ricerche
Area Territoriale di Ricerca di Bologna

- 2023 Biblioteca Dario Nobili del CNR di Bologna.
Questo documento è stato prodotto utilizzando il software libero LaTeX, scritto da Leslie Lamport, mantenuto e sviluppato dal LaTeX3 Project e distribuito sotto la LaTeX Project Public License (LPPL).

Autore della conversione in LaTeX: **Gabriela Carrara**

Indice

1 COS'È DALIA **5**

2 ARCHIVI, DATASET E RISORSE **7**
2.1 Metadati 7
2.2 Risorse 7

3 UTENTI, ARCHIVI E AUTORIZZAZIONI **8**

4 REGISTRAZIONE E ACCESSO **8**

5 USARE DALIA PER L'INSERIMENTO DEI DATI **8**
5.1 Aggiungere un nuovo dataset 8
5.2 Compilazione dei metadati 9
5.2.1 Dati minimi necessari 10
5.2.2 Dati aggiuntivi facoltativi 11
5.3 Inserimento delle risorse 12

Elenco delle figure

1 Home page di DALIA (https://dalia-bo.cnr.it) 5
2 Logo piattaforma CKAN 6
3 Schema dell'organizzazione dati all'interno di DALIA 7
4 Creare un dataset dal link Dati 9
5 Creare un dataset selezionando l'archivio desiderato 9
6 Creare un dataset selezionando l'archivio desiderato 10
7 Aggiungere una risorsa 11
8 Schermata di errore per assenza di risorse inserite 12
9 Aggiungere una nuova risorsa 13

1 COS'È DALIA

DALIA (Dati Aperti della bibLIoteca d'Area di bologna) è un archivio dati digitali e del relativo catalogo dei metadati in cui confluiscono i dati digitali che La Biblioteca Dario Nobili dell'Area Territoriale del CNR di Bologna ha prodotto e/o acquisito nel corso delle attività e dei progetti portati avanti durante gli ultimi 20 anni.
I dati digitali prodotti dalla Biblioteca Dario Nobili (BDN) appartengono a due grandi categorie:

- alcuni sono in formato digitale "nativo" come documenti elettronici, filmati, webinar, presentazioni, mappe concettuali, etc.
- altri sono ancora solo in formato analogico e necessitano di essere trasformati in digitale come alcune pubblicazioni pre-1995, le cartine, i libri storici, etc.

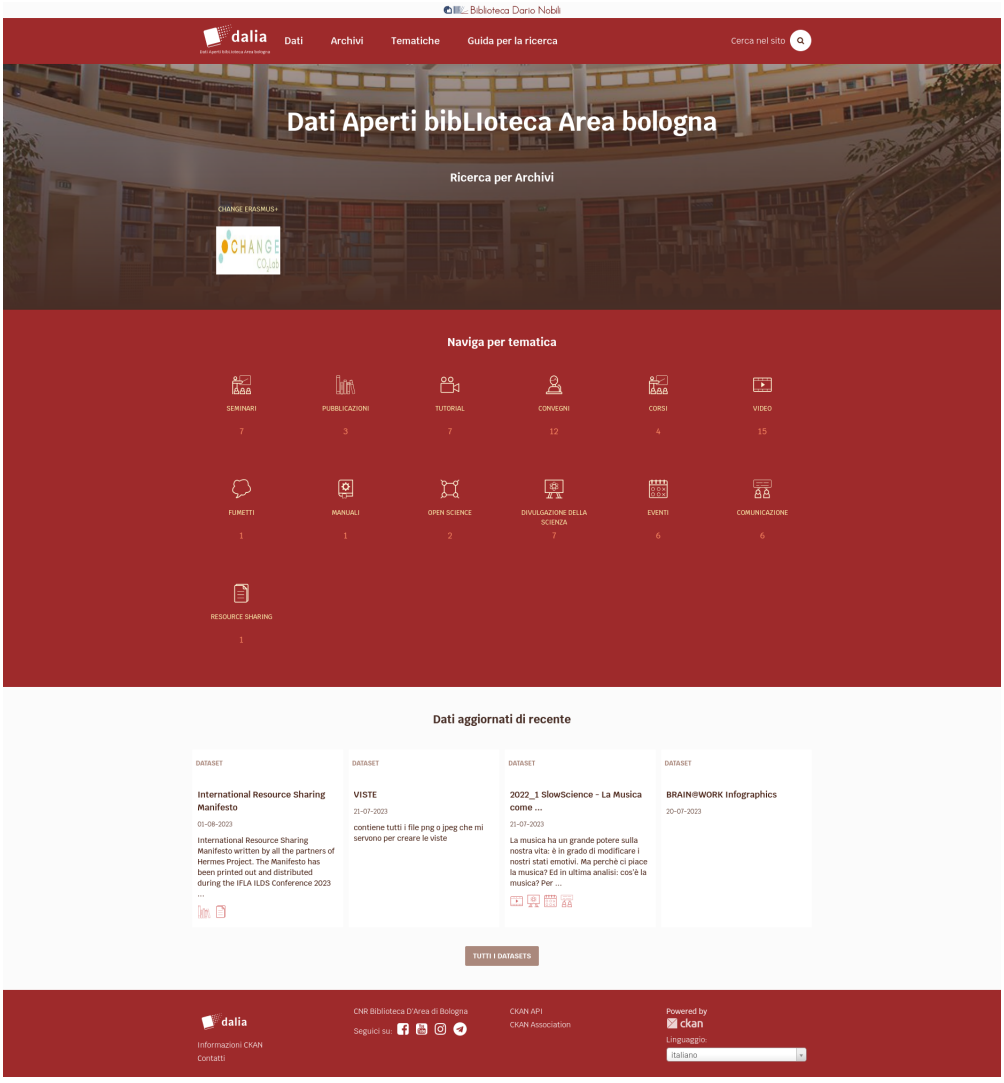

Figura 1: Home page di DALIA (https://dalia-bo.cnr.it)

Questo insieme di dati costituisce il patrimonio di informazioni della BDN e, per poter essere condiviso in futuro, necessita di essere organizzato, conservato e reso fruibile a chiunque possa manifestare interesse. Per questo abbiamo deciso di costruire un data repository, inizialmente solo per uso interno e successivamente come servizio agli utenti della BDN.
DALIA rappresenta il risultato delle seguenti fasi:

- Organizzazione dati e data storage;
- Gestione dati e metadati;
- Distribuzione dati in conformità alla nostra data-policy;
- Accessibilità e sicurezza di conservazione a garanzia del riutilizzo futuro dei dati (in base ai principi Open e FAIR data)

DALIA si basa su una piattaforma CKAN [1] che è stata opportunamente riadattata alle esigenze della nostra biblioteca.

Figura 2: Logo piattaforma CKAN

[1] CKAN acronimo di Comprehensive Knowledge Archive Network (`http://ckan.org/`) è la piattaforma leader mondiale per i portali di dati open-source. CKAN è open e sotto la licenza the GNU Affero General Public License (AGPL) v3.0 consultabile qui http://www.fsf.org/licensing/licenses/agpl-3.0.html. Rappresenta una soluzione software completa e pronta all'uso che rende accessibili e utilizzabili i dati fornendo strumenti per ottimizzarne la pubblicazione, la ricerca e l'utilizzo (inclusa l'archiviazione e la disponibilità di solide API). CKAN si rivolge alle organizzazioni che pubblicano dati aperti (governi nazionali e locali, aziende ed istituzioni) e che desiderano renderli accessibili a tutti. Questa piattaforma è usata da governi e gruppi di utenti in tutto il mondo per gestire una vasta serie di portali di dati di enti ufficiali e di comunità, tra cui portali per governi locali, nazionali e internazionali, come data.gov.uk nel Regno Unito e publicdata.eu dell'Unione Europea, dados.gov.br in Brasile, portali di governo dell'Olanda e dei Paesi Bassi, oltre a siti di amministrazione cittadine e municipali negli USA, nel Regno Unito, Argentina, Finlandia e altri paesi.

2 ARCHIVI, DATASET E RISORSE

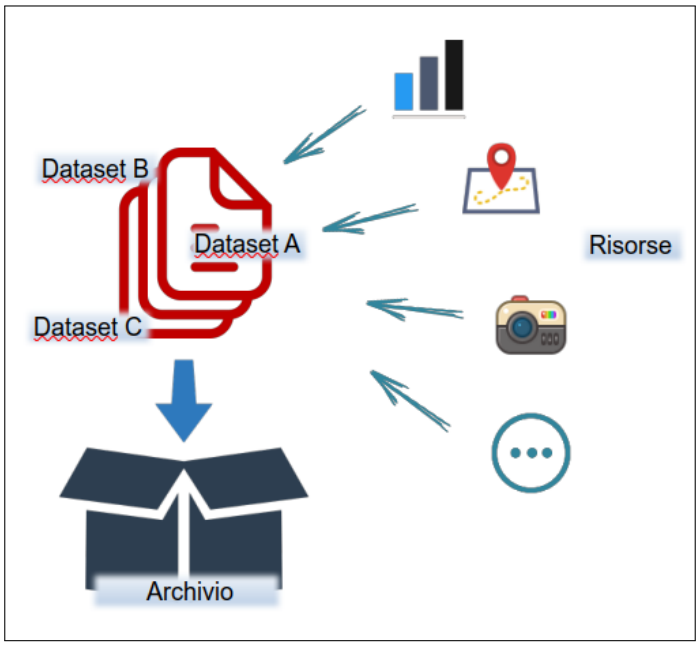

Figura 3: Schema dell'organizzazione dati all'interno di DALIA

All'interno di DALIA esistono numerose entità chiamate **Archivi** a cui appartengono i dati (ergo ogni dato appartiene ad uno specifico archivio)
I dati sono pubblicati in unità chiamate ***dataset***.
Un dataset è costituito da un insieme di dati come ad esempio le statistiche sulla criminalità di una regione, le cifre di spesa di un dipartimento governativo, le letture della temperatura di diverse stazioni meteorologiche oppure come nel nostro caso specifico l'insieme dei prodotti digitali realizzati nel corso di un progetto o di un convegno.
Quando gli utenti cercano dei dati, i risultati della ricerca sono singoli insiemi di dati ovvero dataset.
Un **dataset** o set di dati contiene due cose:

- Informazioni o "**metadati**" sui dati.
- Un certo numero di "**risorse**", che contengono i dati stessi.

2.1 Metadati

I metadati sono descrizioni dei dati, del loro significato, modalità di acquisizione etc. I metadati aiutano gli utenti a navigare tra le informazioni e favoriscono l'indicizzazione dei dataset sui motori di ricerca. Rappresentano in sostanza tutte quelle informazioni che permettono una riutilizzazione futura dei dati. Sono metadati il titolo e l'editore di un documento, la data, i formati in cui è disponibile, la licenza con cui è stato rilasciato, le modalità di acquisizione, le caratteristiche della strumentazione e così via.

2.2 Risorse

Una risorsa è un file di dati. Come detto in precedenza un dataset può contenere un numero qualsiasi di risorse e, cosa importante, non ci sono limitazioni sul formato dei file.
Una risorsa può essere un foglio CSV o Excel, un file XML, PDF, file immagine, dati collegati in formato RDF, ecc.
Le risorse possono essere caricate internamente al data repository oppure memorizzate semplicemente come link che permettano di accedere alla risorsa stessa depositata in un altro punto del web. Per

esempio risorse diverse possono contenere i dati di anni diversi, o possono contenere gli stessi dati in formati diversi. Il gestore del singolo archivio agisce in perfetta autonomia sull'organizzazione dei propri dati.

3 UTENTI, ARCHIVI E AUTORIZZAZIONI

In DALIA, come in tutte le piattaforme CKAN, non è necessario registrare un account utente e fare login per la *consultazione* e il *download* dei dati se questi ultimi sono *pubblicati.*

Diviene necessario avere le credenziali di accesso se si devono inserire e pubblicare dei dati o se si vogliono visualizzare dati contenuti nell'archivio in modalità privata e non pubblica.

Come detto in precedenza ogni set di dati appartiene ad un solo "archivio".

DALIA può ospitare un numero qualsiasi di archivi che vengono gestiti dall'amministratore della piattaforma per l'assetto generale (ambiente, tematiche, creazione utenti etc).
Per l'inserimento e descrizione dei dati all'interno di ogni singolo archivio vengono creati, dall'amministratore di DALIA, utenti aventi ruoli diversi a seconda del livello di autorizzazione necessario.

I ruoli all'interno di un archivio sono i seguenti:

- **Amministratore**: può aggiungere/modificare ed eliminare i dataset, oltre a gestire i membri dell'archivio.
- **Curatore**: può aggiungere e modificare i dataset, ma non può gestire i membri dell'archivio.
- **Membro**: può visualizzare i dataset privati dell'archivio, ma non può aggiungerne di nuovi.

4 REGISTRAZIONE E ACCESSO

La registrazione è necessaria per la maggior parte delle funzioni di pubblicazione e per la modifica e l'aggiornamento dei metadati e non ultima la manutenzione e l'aggiornamento dei dataset.
Non è necessaria per cercare e scaricare dati.
L'amministratore di DALIA provvederà a creare un ID utente a cui verranno richiesti i seguenti dati:

- *Nome utente* – scegliere un nome utente utilizzando solo lettere, numeri, caratteri - e _.
  Ad esempio, "pippo" o "pippo_pluto88".
- *Nome completo* – nome da visualizzare nel profilo utente.
- *Indirizzo e-mail* – non sarà visibile agli altri utenti.
- *Password* – una volta scelta, inserire la stessa password in entrambe le caselle

E' inoltre possibile aggiungere un file immagine o un link al profilo utente.

5 USARE DALIA PER L'INSERIMENTO DEI DATI

5.1 Aggiungere un nuovo dataset

Una volta selezionato l'archivio di appartenenza è possibile accedere alla schermata "Crea set di dati" in due modi.

- Selezionare il link "Dati" visibile nella parte superiore di qualsiasi pagina. Da qui selezionare il pulsante "Aggiungi un dataset" illustrato in figura.

Figura 4: Creare un dataset dal link Dati

- In alternativa, selezionare il link "Archivi" nella parte superiore di una pagina; selezionare la pagina dell'archivio che dovrebbe possedere il nuovo set di dati.

  A condizione di essere membri di questo archivio, è possibile selezionare il pulsante "Aggiungi un Dataset" come illustrato in figura.

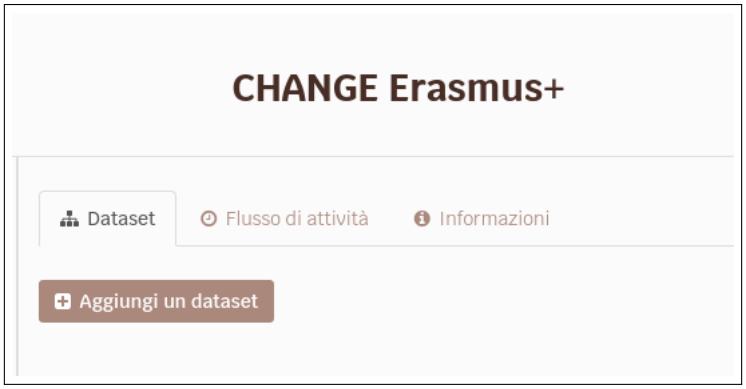

Figura 5: Creare un dataset selezionando l'archivio desiderato

IMPORTANTE:
I dataset possono essere creati in forma **privata** o **pubblica**.
Quando sono in forma privata i dataset e tutte le risorse in essi contenute sono visibili solo ed esclusivamente all'amministratore del progetto ed a tutti i membri associati a quel progetto dotati di account, e naturalmente all'amministratore di DALIA.
Dal momento in cui il dataset viene reso pubblico è visibile sulla rete da tutti gli utenti collegati ad internet.

5.2 Compilazione dei metadati

La compilazione dei metadati di un dataset è fondamentale per permettere ad altre persone di comprendere la natura del dataset, che tipo di dati (risorse) siano presenti e con che specificità.

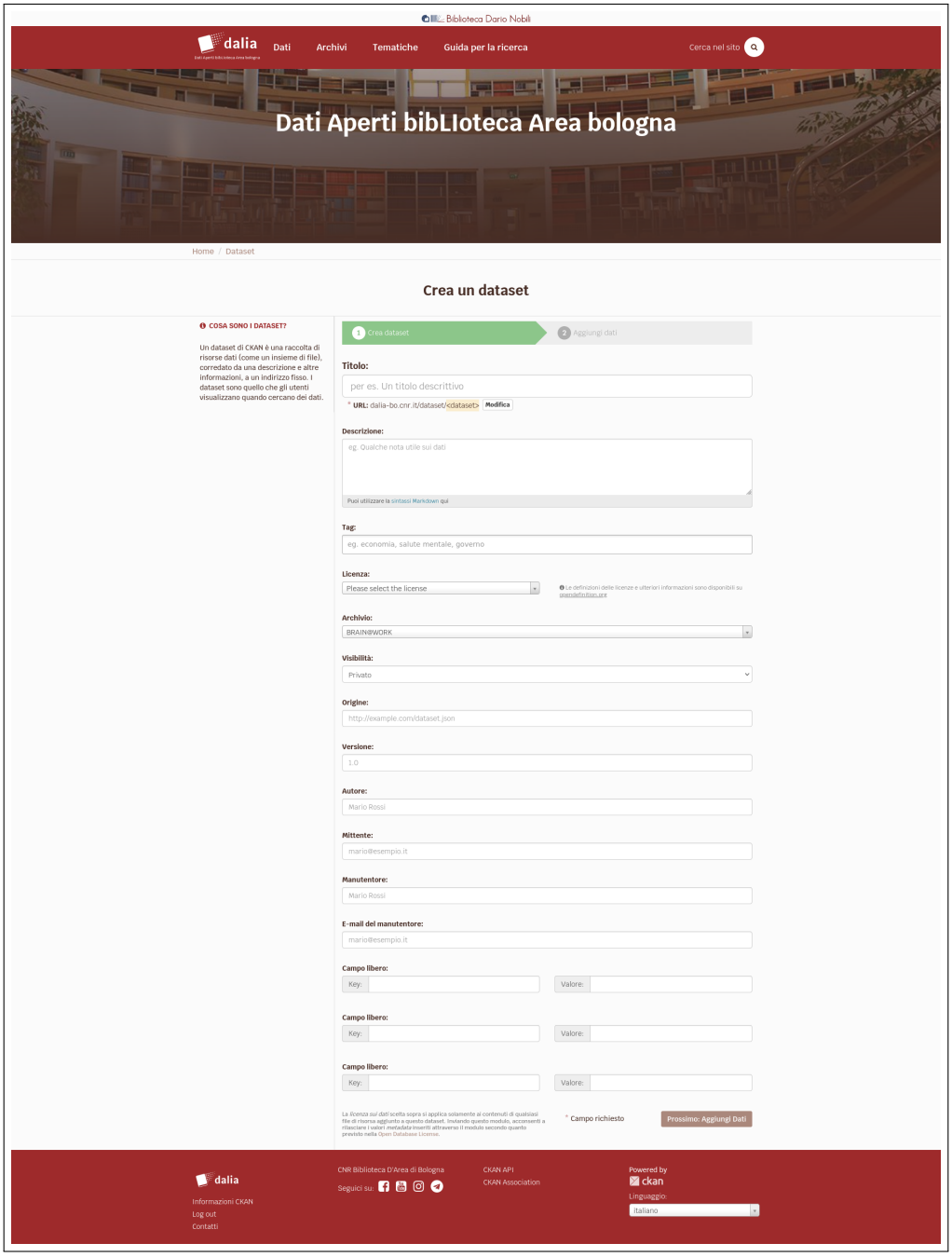

Figura 6: Creare un dataset selezionando l'archivio desiderato

Il sistema propone di default la compilazione di un numero minimo di informazioni ma possono esserne aggiunte altre in funzione delle caratteristiche dei dati immessi o delle necessità descrittive.

5.2.1 Dati minimi necessari

Per una corretta creazione dei dataset sono necessarie le seguenti informazioni

- **Titolo** - Questo titolo sarà unico in tutto il data repository, quindi può non essere breve ma specifico. Ad esempio, "2010 - Perugia - VI Convegno NILDE" è meglio di "Convegno NILDE".
- **Descrizione** - È possibile aggiungere qui una descrizione estesa del dataset, della sua natura, del suo scopo, includendo informazioni come la provenienza dei dati e qualsiasi informazione che le persone devono sapere quando utilizzano i dati.

- **Tags** - Qui si possono aggiungere tag specifici che aiutano a trovare i dati e a collegarli con altri dati correlati. Per inerire più tag basta premere il tasto ¡invio¿ tra un tag e l'altro. Se si inserisce un tag in modo errato è possibile utilizzare il pulsante di cancellazione per rimuoverlo prima di salvare il set di dati.
- **Licenza** - E' importante includere le informazioni sulla licenza in modo che gli utenti sappiano come possono utilizzare i dati. Questo campo è costituito da un menù a tendina in cui poter scegliere la licenza desiderata. Nella policy di DALIA la licenza scelta è CC-BY-SA. I firmatari dell'accordo di utilizzo del servizio DALIA accettano tale licenza.
- **Archivio** - E' presente un menu a tendina per scegliere quale archivio dovrà essere proprietario dei dati. Assicurarsi che l'impostazione predefinita scelta sia quella corretta prima di procedere. (Probabilmente la maggior parte degli utenti farà parte di un'unico archivio, in tal caso il sistema sceglierà di default archivio di appartenenza dell'utente).
- **Visibilità** - In questo campo è possibile stabilire se i dati siano pubblici (visibili da chiunque) o privati (visibili solo all'interno dei gestori e membri dell'archivio).

5.2.2 Dati aggiuntivi facoltativi

I seguenti dati pur essendo facoltativi ai fini della creazione dei dataset possono comunque essere importanti.

- **Autore** - Il nome della persona o dell'archivio responsabile della produzione dei dati.
- **Email dell'autore** - un indirizzo di posta elettronica dell'autore, a cui inviare le domande sui dati.
- **Manutentore e e-mail del manutentore** - I dati di riferimento del gestore di DALIA.
- **Campi liberi** - Se si desidera che il set di dati abbia un altro campo, è possibile aggiungere qui il nome e il valore del campo. Ad esempio, "Anno di pubblicazione". Si noti che se c'è un campo aggiuntivo che è necessario per un gran numero di set di dati, contattare l'amministratore del sito per modificare lo schema predefinito e i moduli dei dataset.

Dopo aver compilato le informazioni di questa pagina, selezionare "Salva e aggiungi altro". (In alternativa, selezionare "Precedente" per eliminare le informazioni compilate).

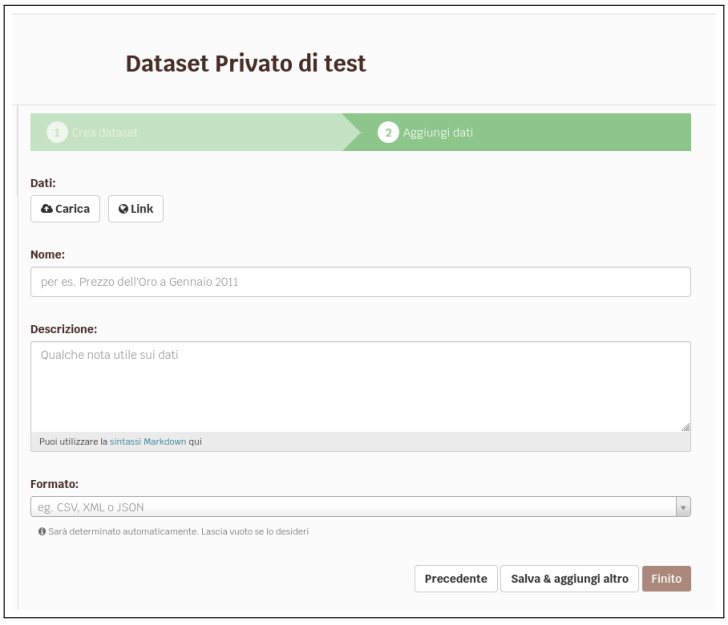

Figura 7: Aggiungere una risorsa

NOTA:
Ogni dataset deve contenere almeno una risorsa per essere creato. In caso contrario si avrà la seguente schermata di errore (fig. 8)

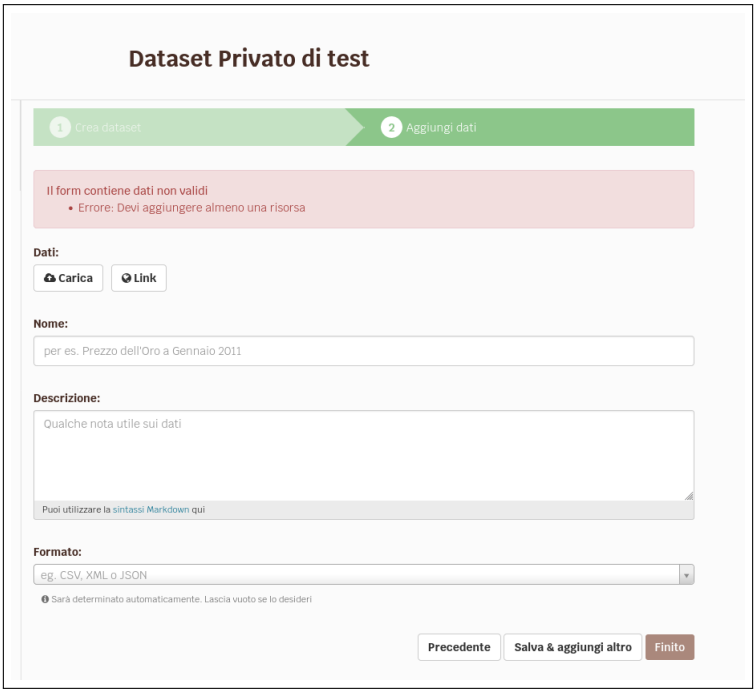

Figura 8: Schermata di errore per assenza di risorse inserite

5.3 Inserimento delle risorse

La schermata che si raggiunge cliccando "Aggiungi Dati" (vedi fig. 7) è quella in cui è possibile aggiungere una o più risorse che contengono i dati relativi al dataset che si sta creando.
Si può scegliere di inserire nell'archivio un file o un link per la risorsa di dati nel caso sia presente in rete e selezionare la scelta appropriata nella parte superiore dello schermo:

- Se si fornisce un link ai dati, come ad esempio "http://example.com/mydata.csv", selezionare "Link".
- Se i dati da aggiungere a DALIA si trovano in un file presente sul proprio computer, selezionare "Carica".
  DALIA fornisce un file-browser per selezionarlo.

Nella pagina è possibile inserire ulteriori informazioni che, nonostante siano facoltative, è buona pratica compilare.

- Nome - il nome della risorsa, i.e. "xxxxxx". All'interno di un dataset non possono esserci risorse con lo stesso nome.
- Descrizione - una breve descrizione della risorsa.
- Formato - Indicare il formato del file risorsa es: CSV, XLS, JSON, PDF, etc.

Cliccando sul bottone "Fatto" il dataset viene creato. Se vanno inserite più risorse (file o link) all'iterno di un dataset selezionare il bottone "Salva e aggiungi altro" e verrà visualizzata una nuova schermata di inserimento dati.

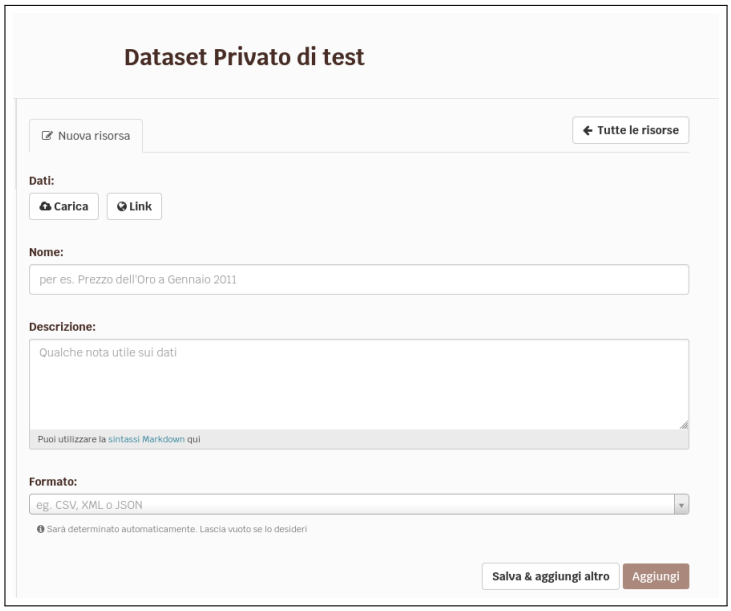

Figura 9: Aggiungere una nuova risorsa

Cliccando sul bottone "Aggiungi" la risorsa viene inserita nel dataset.
La schermata di fig. 9 è quella che compare anche quando in un secondo tempo si vogliano aggiungere nuove risorse a dataset esistenti.