

RAPPORTI TECNICI INGV

Prototipo di un portale
per la distribuzione dei dati
dell'Osservatorio Etneo
basato sulla piattaforma Zenodo



Direttore Responsabile

Valeria DE PAOLA

Editorial Board

Milena MORETTI - Editor in Chief (milena.moretti@ingv.it)

Raffaele AZZARO (raffaele.azzaro@ingv.it)

Christian BIGNAMI (christian.bignami@ingv.it)

Viviana CASTELLI (viviana.castelli@ingv.it)

Rosa Anna CORSARO (rosanna.corsaro@ingv.it)

Luigi CUCCI (luigi.cucci@ingv.it)

Domenico DI MAURO (domenico.dimauro@ingv.it)

Mauro DI VITO (mauro.divito@ingv.it)

Marcello LIOTTA (marcello.liotta@ingv.it)

Mario MATTIA (mario.mattia@ingv.it)

Nicola PAGLIUCA (nicola.pagliuca@ingv.it)

Umberto SCIACCA (umberto.sciacca@ingv.it)

Alessandro SETTIMI (alessandro.settimi1@istruzione.it)

Andrea TERTULLIANI (andrea.tertulliani@ingv.it)

Segreteria di Redazione

Francesca DI STEFANO - Coordinatore

Rossella CELI

Robert MIGLIAZZA

Barbara ANGIONI

Massimiliano CASCONI

Patrizia PANTANI

Tel. +39 06 51860068

redazione@ingv.it

REGISTRAZIONE AL TRIBUNALE DI ROMA N.174 | 2014, 23 LUGLIO

© 2014 INGV Istituto Nazionale

di Geofisica e Vulcanologia

Rappresentante legale: Carlo DOGLIONI

Sede: Via di Vigna Murata, 605 | Roma



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

RAPPORTI TECNICI INGV

Prototipo di un portale per la distribuzione
dei dati dell'Osservatorio Etneo
basato sulla piattaforma Zenodo

*Etneo Observatory data distribution web
platform prototype based on Zenodo*

Mario Torrisi¹, Fabrizio Pistagna¹, Marcello D'Agostino¹, Mario Locati², Placido Montalto¹

¹INGV | Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania - Osservatorio Etneo

²INGV | Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Milano

Accettato 16 giugno 2022 | Accepted 16 June 2022

Come citare | How to cite Torrisi M., Pistagna F., D'Agostino M., Locati M. e Montalto P., (2022). Prototipo di un portale per la distribuzione dei dati dell'Osservatorio Etneo basato sulla piattaforma Zenodo. Rapp. Tec. INGV, 457: 1-26, <https://doi.org/10.13127/rpt/457>

In copertina Schema logico che descrive la intercomunicazione tra i processi che compongono un deployment completo di Zenodo | Cover Logical diagram describing the intercommunication between the processes that make up a complete Zenodo deployment

457

INDICE

Riassunto	7
<i>Abstract</i>	7
Introduzione	7
1. Datarep: un repository open data basato su Zenodo	9
1.1 Obiettivi della personalizzazione del software zenodo	9
1.2 Processo di caricamento dati su datarep	9
1.3 Utilizzo delle rest api	15
1.3.1 Autenticazione (Authentication)	16
1.3.2 Deposito (Deposition)	17
1.3.3 Deposito di File (Deposition files)	17
1.3.4 Azioni di Deposito (Deposition Actions)	18
2. Datarep: metodologia di implementazione	18
3. Datarep: architettura software	19
4. Conclusione e sviluppi futuri	21
Glossario	21
Sitografia	22

Riassunto

Il modello *Open Data* prevede che le pubblicazioni di risultati scientifici da parte di enti o istituti di ricerca siano liberamente accessibili e fruibili da chiunque, con l'unico vincolo di citarne la fonte. Questo approccio ha spinto gli enti di ricerca a pubblicare i propri prodotti scientifici tramite un'operazione definita "deposito", su portali web auto-gestiti e auto-mantenuti che rappresentano delle banche dati istituzionali secondo le direttive del paradigma *Open Access*. Tra le operazioni chiave che l'Osservatorio Etneo (OE) svolge, vi è la raccolta di dati provenienti sia dalle strumentazioni di acquisizione dati disseminate sul territorio di competenza, sia da misure ed elaborazioni dei dati raccolti; tali dati vengono catalogati in insiemi, definiti *datasets*, in base a valori discreti quali ad esempio: il tipo di fenomeno esaminato, il suo intervallo di tempo e la sua geolocalizzazione. Seguendo le direttive *Open Access* per la pubblicazione di tali *dataset*, l'Osservatorio Etneo adotta una implementazione delle suddette banche dati, altresì dette *repository* di dati o archivi digitali. Tale implementazione, chiamata DataRep e descritta in questo documento, è basata sul portale Zenodo, un *software open-source* per l'implementazione di un *repository* multidisciplinare mantenuto dal CERN. Il lavoro descritto in questo documento è frutto della collaborazione dell'Osservatorio Etneo e dell'Ufficio Gestione Dati, ed in particolare dell'interazione tra i software DataRep - implementato dall'OE personalizzando ed adattando il software Zenodo - e del Metadata Editor, strumento ufficiale per la pubblicazione scientifica da parte dell'ente. DataRep permette di ottenere una *landing-page* automaticamente associata e risolvibile tramite DOI pubblico per ogni *dataset* pubblicato tramite il Metadata Editor.

Abstract

The Open Data model requires that scientific publications results, produced by research organizations or institutes must be freely accessible and usable by anyone, with the only constraint of citing the source. This approach has driven research institutions to publish their scientific products through an operation called "deposit", on self-managed and self-maintained web portals that represent institutional databases according to the Open Access directives. Among the key operations carried out by the INGV Osservatorio Etneo (OE), there is the gathering and collection of data from the data-acquisition instruments scattered throughout its competence area; these data cataloged in sets, called datasets, on the basis of discrete values such as: examined phenomenon, its time interval and its geolocation. Following the Open Access directives for the publication of these datasets, the OE adopts an implementation of the aforementioned databases, also known as data repositories or digital archives. The implementation described in this document, called DataRep, is based on the Zenodo portal: an open-source software for the implementation of a multidisciplinary repository maintained by CERN. This document describes the result of the collaboration between the OE and the INGV Ufficio Gestione Dati, specifically the interaction between the DataRep software - implemented by the OE by customizing and adapting the Zenodo software - and the Metadata Editor, the official tool for publishing scientific products adopted by INGV. DataRep allows to automatically obtain a landing page associated to each dataset, resolvable through the public DOI published through the Metadata Editor.

Keywords Dati-Aperti; Archivio digitale; Accesso-Pubblico | *Open-Data; Repository; Open-Access*

Introduzione

L'analisi dei dati relativi ai distretti vulcanici di competenza dell'Osservatorio Etneo è alla base di tutte le attività di sorveglianza, monitoraggio e ricerca della sezione etnea del INGV.

I dati vengono catalogati in insiemi, definiti **datasets**, in base a valori discreti quali, ad esempio, il tipo di fenomeno esaminato, il suo intervallo di tempo e la sua geolocalizzazione [1]. La pubblicazione di tali **datasets** come risorse fruibili dalla comunità scientifica rappresenta un aspetto essenziale dell'attività svolta all'interno dell'Osservatorio Etneo dell'Istituto Nazionale di Geofisica e Vulcanologia (INGV-OE). Questi **datasets**, una volta accertata l'aderenza alle linee guida della Politica dei Dati istituzionale [2], vengono pubblicati secondo le direttive dell'**Open Data** [3], ovvero liberamente accessibili a chiunque con il solo vincolo per gli utenti di citarne la fonte. La pubblicazione di **Open Data** implica anche la gestione di banche dati sempre accessibili rispettando la modalità di pubblicazione denominata **Open Access** [4]. Un esempio è il portale **Zenodo** [5], un *repository* multidisciplinare mantenuto dal CERN che consente ai ricercatori di condividere e preservare i risultati della ricerca in qualsiasi dimensione e formato.

Ogni ente o istituto di ricerca può pubblicare i propri risultati scientifici seguendo modelli gestionali e infrastrutturali diversi: in base alle caratteristiche dei dati trattati e alle proprie esigenze, rendendoli fruibili attraverso la pubblicazione degli stessi su portali *web* auto-gestiti e auto-mantenuti tramite un'operazione di pubblicazione che viene definita **deposito** [6].

L'INGV dispone dal 2005 di un proprio archivio istituzionale chiamato Earth-Prints [7] che permette al personale di archiviare e rendere pubblicamente accessibili documenti come articoli scientifici, rapporti tecnici, presentazioni, poster, ecc...

Se per quanto riguarda le pubblicazioni l'INGV è dunque autonomo, per l'archiviazione di dati frutto della ricerca, l'Ente non ha ancora una piattaforma istituzionale e ogni gruppo di lavoro gestisce in autonomia ed in maniera eterogenea la pubblicazione dei propri **dataset**. Con l'approvazione di una Politica dei Dati [8] e l'istituzione di un Ufficio Gestione Dati [9], dalla fine del 2018 l'INGV sta avviando un processo atto ad avere una più efficiente gestione dei propri dati. Un primo passo concreto in questo ambito è avvenuto alla fine del 2020 con l'inaugurazione del Portale dei Dati Aperti istituzionale [10]. Questo portale svolge, al momento, la sola funzione di rendere pubblicamente consultabile il Registro Dati, la catalogazione e descrizione dei dati pubblicati. In gergo informatico, questa tipologia di portale può essere definita: "*Metadata Catalogue*"¹.

Al fine di indagare e sperimentare la realizzazione di uno strumento per l'archiviazione dei dati, si è scelto di prendere in considerazione il *repository Zenodo*: una piattaforma software per l'archiviazione e pubblicazione di prodotti della Ricerca di tipo *general-purpose*, ovvero non legata a una specifica tipologia di prodotto o disciplina scientifica. La piattaforma è sviluppata e mantenuta dal CERN [11] in collaborazione con l'iniziativa europea OpenAire [12], la quale si occupa di fornire le linee guida per il suo sviluppo.

Zenodo permette ai ricercatori di archiviare, secondo il paradigma *Open Access*, rapporti scientifici, presentazioni, immagini, **datasets**, **software**, ed ogni altro tipo di oggetto digitale relativo alle attività di ricerca. Ad ogni sottomissione viene associato un **Digital Object Identifier**² (**DOI**) che rende facilmente "*citabile*" il prodotto pubblicato [13].

Questo documento presenta un prototipo di archivio - *repository* - di dati denominato **DataRep**, frutto di un'implementazione adattata alle esigenze di archiviazione dei **dataset** prodotti presso l'Osservatorio Etneo basata sul software Zenodo i cui obiettivi sono:

¹ Un *Metadata Catalogue* è un servizio per la memorizzazione, l'accesso e l'interrogazione delle informazioni descrittive (metadati) di oggetti digitali.

² Il Digital Object Identifier (acronimo DOI, in italiano "Identificatore di un oggetto digitale") è uno standard che consente l'identificazione duratura e univoca di oggetti di qualsiasi tipo all'interno di una rete digitale, e l'associazione ad essi dei relativi dati di riferimento - i metadati - secondo uno schema strutturato ed estensibile.

- permettere al ricercatore di pubblicare in modo semplice dei *dataset* rispettando le linee guida della politica dei dati istituzionale;
- offrire la possibilità di accedere facilmente ai dati pubblicati dal personale tramite uno strumento efficace e semplice da usare.

1. DataRep: un repository Open Data basato su Zenodo

1.1 Obiettivi della personalizzazione del software Zenodo

L'obiettivo principale del repository DataRep è quello di fornire uno strumento che permetta al personale di pubblicare i propri dati creando automaticamente pagine web - *landing-page* [14] - e renderle accessibili conformemente alle direttive degli *Open Data*. A ciascun dataset pubblicato sarà associato un codice identificativo DOI, a sua volta associato alla pagina web creata automaticamente da DataRep, in modo da garantire così la “risolvibilità”³ del DOI.

Il processo di pubblicazione dei risultati scientifici prevede che il ricercatore proceda alla descrizione e caricamento dei dati tramite uno strumento web denominato **Metadata Editor** [15] collegato al **Registro Dati** istituzionale e gestito dall'Ufficio Gestione Dati dell'INGV. Il Metadata Editor si occupa della gestione centralizzata dei *metadati* che descrivono i dati e ne permette l'esportazione secondo vari schemi di metadati, tra cui quello definito da **DataCite**⁴ [16].

La Politica dei Dati istituzionale prevede il rispetto dei principi **FAIR** [17] dei dati: *Findable*, *Accessible*, *Interoperable*, *Reusable*. La metadattazione esaustiva e i servizi di accesso ai metadati gestiti tramite il Metadata Editor soddisfano la “*Findability*”. Le capacità di archiviazione ed i servizi per l'accesso ai dati offerti da DataRep soddisfano la “*Accessibility*”. La scelta del formato di codifica dei file nel rispetto dei principi degli *Open Data* (che non implicano l'uso di uno specifico software commerciale per l'uso dei dati) soddisfa il requisito di “*Interoperability*”. Infine, la licenza di tipo Creative Commons, ammessa dalla Politica istituzionale dei Dati, soddisfa la “*Reusability*”.

DataRep permette sia di associare in maniera automatica i metadati della pubblicazione generati dal Metadata Editor alle relative *landing-pages*, sia di rendere accessibile quanto pubblicato attraverso la risolvibilità del codice DOI che reindirige automaticamente gli utenti alla pagina web (*landing-page*) che ospita i dati ed i metadati associati.

1.2 Processo di caricamento dati su DataRep

Il processo da seguire per pubblicare i propri *dataset* nel repository DataRep è descritto per linee generali in Figura 1, e coinvolge due entità: il Metadata Editor ed il DataRep.

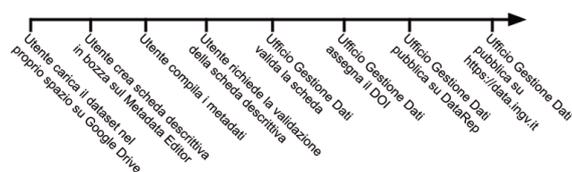


Figura 1 Panoramica del processo di pubblicazione di dati su DataRep.

Figure 1 Overview of the process of publishing data on DataRep.

³ Permettere il libero accesso alla risorsa attraverso un protocollo di comunicazione standard, ad es. Web.

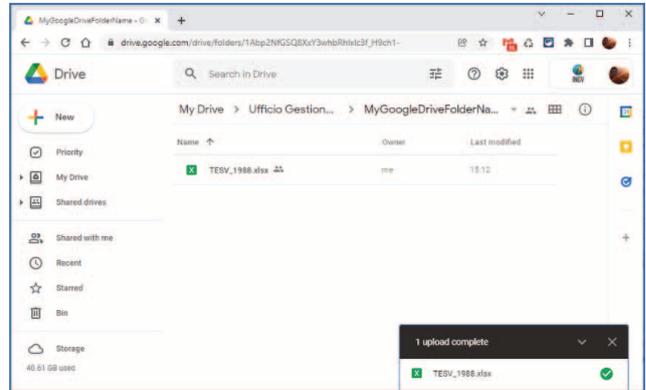
⁴ DataCite è un'organizzazione internazionale no-profit il cui scopo è quello di migliorare la citazione dei dati al fine di facilitare l'accesso agli stessi attraverso Internet.

Di seguito viene riportata una descrizione dettagliata dei singoli passi richiesti all'utente durante il processo di caricamento:

1. Inizialmente l'utente carica il *dataset* all'interno di uno spazio dati personale e privato sul cloud di Google Drive (Figura 2) (il dominio @ingv viene gestito dal servizio *business* di Google workspace) [18].

Figura 2 Tramite il proprio spazio personale istituzionale sul Google Drive, l'utente carica il dataset da pubblicare sul portale DataRep.

Figure 2 Through their own Google Drive cloud space, users upload datasets to be published on DataRep.



2. Successivamente l'utente rende accessibili gli stessi *files* del *dataset* mediante la funzionalità "Condividi" e selezionando l'opzione "Chiunque abbia il link" generando univocamente un *link* per ciascun *file* che sarà utilizzato dal Metadata Manager per l'operazione di *upload* su DataRep (un *dataset* può essere composto da uno o più file a cui corrispondono di conseguenza più *link*) (Figura 3 e Figura 4). I dati devono essere caricati in un formato aperto [19] tenendo conto che lo scopo è quello della fruibilità degli stessi.

Figura 3 Si inizia il processo di condivisione del file tramite l'apposita funzionalità.

Figure 3 The sharing process starts through the appropriate functionality.

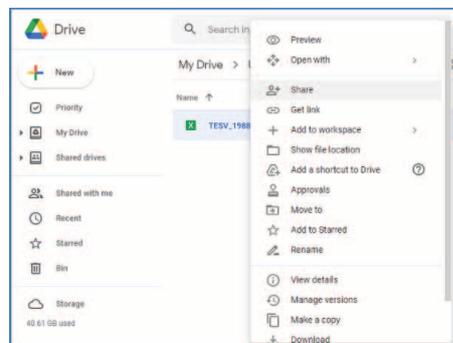
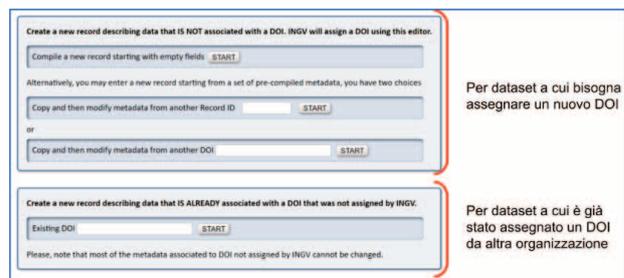


Figura 4 Si ottiene un link univoco per la condivisione del file, che successivamente verrà utilizzato dal Metadata Editor per recuperare il dato dal Drive.

Figure 4 You get a unique sharing link to the file, which will later be used by the Metadata Editor to retrieve the data from the Drive.



3. Ottenuto il *link* di condivisione, l'utente si autentica sul portale del Metadata Editor utilizzando il proprio account istituzionale (nome.cognome@ingv.it) al fine di compilare e sottomettere una scheda descrittiva del *dataset*. L'utente può scegliere se (Figura 5):
 - a. creare una nuova scheda;

- b. creare una nuova scheda copiando e modificando i metadati associati ad una scheda già esistente, specificando il suo identificativo;
- c. creare una nuova scheda copiando e modificando i metadati associati ad una scheda già esistente, specificando il suo DOI assegnato dall'INGV;
- d. creare una nuova scheda copiando e modificando i metadati associati ad una scheda già esistente, specificando il suo DOI assegnato da un'altra organizzazione.

Figura 5 Scelta tra le modalità di creazione di una nuova scheda nel Metadata Editor.

Figure 5 New tab creation choices into the Metadata Editor.

Le prime tre opzioni si riferiscono alla creazione di una scheda descrittiva di un *dataset* a cui l'INGV assegnerà un nuovo DOI, previa valutazione della scheda stessa.

4. Scelta la modalità di creazione della nuova scheda, secondo le opzioni descritte al punto precedente, il sistema mostra un *web form* con diversi campi da compilare che costituiscono i *metadati* caratteristici di una pubblicazione scientifica. L'utente descrive il *dataset* compilando esaurientemente quanti più campi possibili tra quelli disponibili nella scheda (Figura 6). In particolare, il campo "Landing page" non deve essere compilato, se il *dataset* sarà caricato sul *repository* DataRep (Figura 10), perché questa informazione verrà restituita automaticamente dalla logica del portale non appena, tramite l'utilizzo delle sue API, verrà fatta richiesta di caricare una coppia dato/metadato (paragrafo 1.3).

- a. 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 (Figura 6), 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.

Figura 6 Porzione del form web del Metadata Editor per la compilazione di un nuovo record.

Figure 6 New record web-form portion.

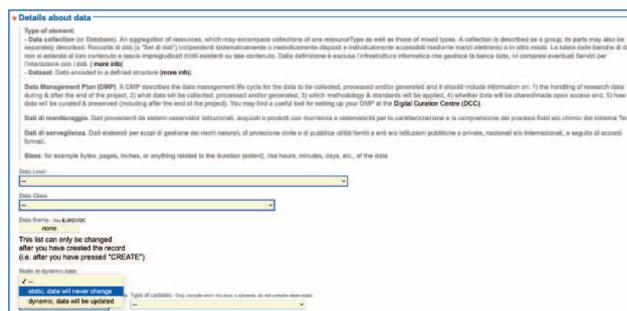
- b. Durante la compilazione dei campi è possibile verificarne la correttezza in ogni momento agendo sul bottone “Check if this record can be validated”. Se il processo rileva errori o incompletezze, queste saranno dettagliatamente elencate dal Metadata Editor.
- c. Solo quando l’utente avrà cliccato il bottone “Create” in fondo alla scheda, essa sarà compilabile in ogni sua parte. Il *dataset* non verrà pubblicato quando la scheda verrà creata o salvata, ma solo quando espressamente richiesto dall’utente e comunque dopo un processo di validazione così come previsto dalla Politica dei Dati istituzionale [20].

5. A questo punto l’utente aggiunge l’elenco dei file da caricare su DataRep:

- a. Il Metadata Editor mostrerà le funzioni di caricamento su DataRep quando queste due condizioni sono soddisfatte:
 - i. il dataset deve essere dichiarato **statico** (Figura 7) nella sezione di metadati “Details about data”. Con statico si intende che il dataset non verrà mai più aggiornato dopo la sua pubblicazione. 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”.

Figura 7 Selezione “statica” del dataset da caricare sul portale DataRep.

Figure 7 “Static” dataset selection to upload on DataRep.



- ii. almeno una persona elencata tra i “Main authors” della sezione di metadati “Data producers” deve essere affiliata all’Osservatorio Etneo (Figura 8). Le funzioni per caricare dati su DataRep saranno mostrate solo dopo che l’elenco degli autori sarà compilato. È 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.

Figura 8 Dettaglio della selezione del “Main author” e dei rispettivi “Contributors”.

Figure 8 “Main author” and related “Contributors” detailed view.



b. Si elencano i file da caricare su DataRep nella sezione di metadati “Data Access and Distribution”. I file vanno indicati sotto forma di link a Google Drive (Figura 9) e vanno aggiunti cliccando il bottone “Modify the list”. Per verificare che i dati siano caricabili su DataRep cliccare sul bottone “Check Google Drive files” (Figura 10) (Figura 11).

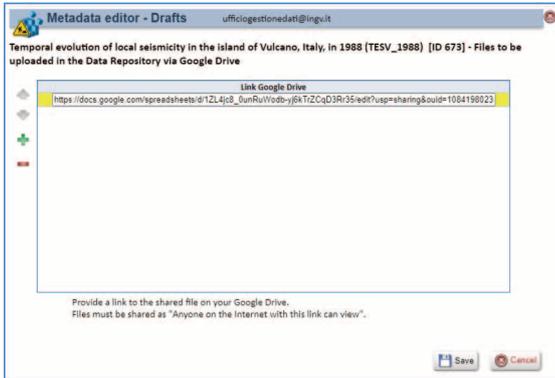


Figura 9 Dettaglio per la gestione dei link relativi ai file caricati sul Drive di Google.

Figure 9 Detailed view for the management and insertion of the links to the uploaded files on the Google Drive.

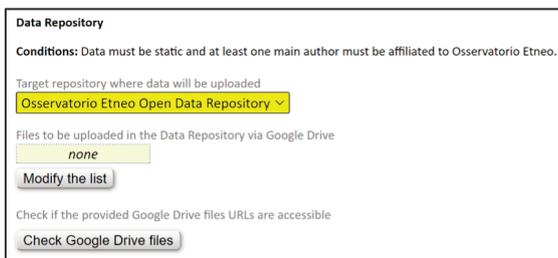


Figura 10 Durante la fase di inserimento dei link per ogni file componente il dataset che verrà caricato sul portale DataRep è possibile verificarne la presenza sul Drive ed i permessi di accesso.

Figure 10 During the insertion of the links for each component file of the dataset that will be loaded on the DataRep portal, it is possible to check its presence on the Drive and the access permissions.

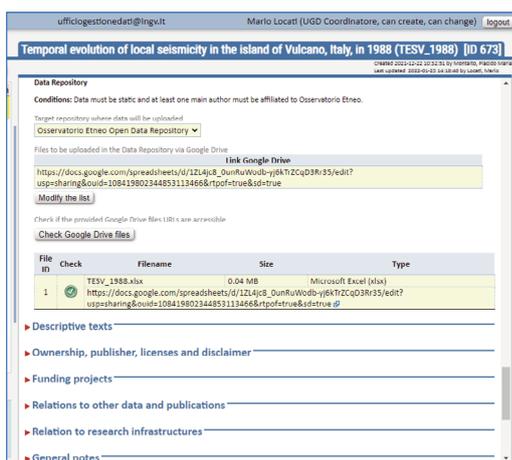


Figura 11 Il dettaglio dei file reperibili ai link del Google Drive precedentemente inseriti. Tale dettaglio è visionabile nella stessa scheda del Metadata Editor e modificabile tramite il bottone “Modify the list”, mentre con il bottone “Check Google Drive files”, viene effettuata una verifica che il file remoto esista e sia accessibile.

Figure 11 The details of the files available on Google Drive links. This detail can be viewed in the same tab of the Metadata Editor and can be modified using the “Modify the list” button, while with the “Check Google Drive files” button, a check is made that the remote file exists and is accessible.

6. Quando la compilazione della scheda sarà completata si potrà procedere alla validazione cliccando sul bottone “Start the validation process”; in caso di problemi, l'Ufficio Gestione Dati si occuperà di contattare l'utente per eventuali correzioni o integrazioni.

La figura Figura 12 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:

- a. il primo grado di validazione è in capo all'Ufficio Gestione Dati, nello specifico il coordinatore e un referente per ciascuno dei Dipartimenti indicati;
- b. 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;
- c. il terzo grado di validazione è in capo ai Direttori dei Dipartimenti indicati in "INGV Department" nella sezione di metadati "Info specific to INGV".
- d. 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.

7. Il Metadata Editor si interfacerà automaticamente con DataRep per creare una landing-page (Figura 13) e caricare sia i dati presenti sul Google Drive sfruttando le API di Zenodo (paragrafo 1.3), che associare i metadati compilati precedentemente in fase di compilazione bozza. L'esempio in Figura 13 mostra un dataset⁵ pubblicato dall'OE utilizzato per eseguire i test di pubblicazione con la piattaforma DataRep. Al momento della pubblicazione del dataset i Referenti Scientifico e Tecnologico associati riceveranno notifica (Figura 14). La figura Figura 15 riassume in maniera schematica e sequenziale l'intero processo descritto in questo paragrafo.

Figura 12 Gli stati di avanzamento del processo di validazione, visionabili sempre attraverso la pagina della scheda di pubblicazione.
Figure 12 The validation process progress state, which can always be viewed on the publication form page.

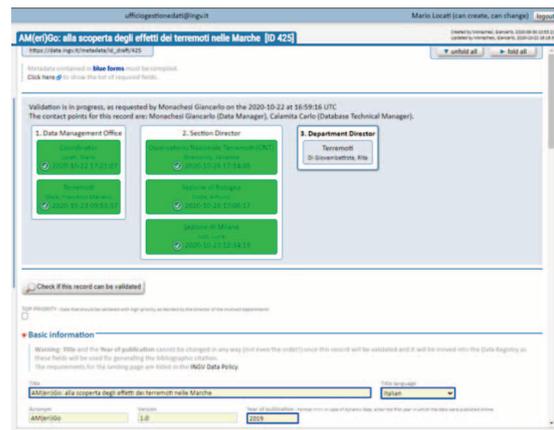
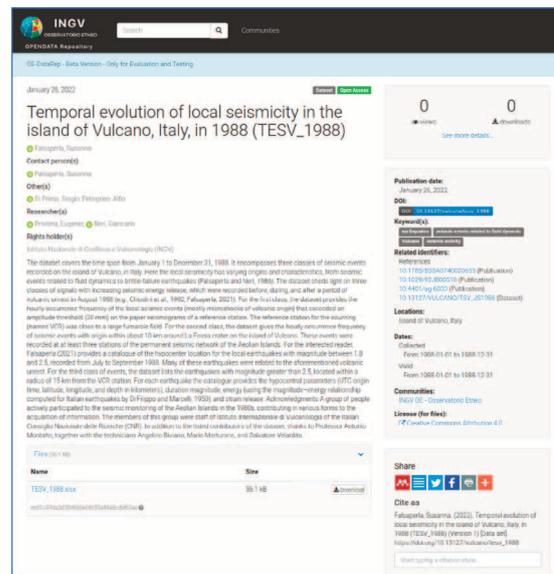


Figura 13 La landing-page creata all'interno del portale DataRep.
Figure 13 The landing page created within the DataRep portal.



⁵ Falsaperla S., (2022). *Temporal evolution of local seismicity in the island of Vulcano, Italy, in 1988 (TESV_1988)* (Version 1) [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/vulcano/tesv_1988

In Figura 13 è mostrato un esempio di *landing-page* creata automaticamente sul portale DataRep dall'operazione di *upload* del Metadata Editor, il quale carica sia i metadati descrittivi del *dataset*, compilati attraverso il *form web* messo a disposizione, che il dataset stesso, caricato in precedenza sullo spazio di *staging* temporaneo su Google Drive. In ciascuna *landing-page*, associata ad un *record* caricato sul portale DataRep, vengono mostrati:

- il titolo associato alla pubblicazione;
- gli autori e le relative affiliazioni;
- la descrizione del record.

La *landing-page* offre la possibilità di scaricare il dataset ed inoltre, a seconda del formato del dataset caricato, può presentare un'anteprima dello stesso. Nella colonna di destra della *landing-page* del *record*, vengono mostrate le varie opzioni per la citazione della pubblicazione, un dettaglio delle versioni pubblicate, il DOI assegnato ed un estratto dei metadati associati, quali, ad es.: *keywords*, *geolocations*, *licenses*, ecc...

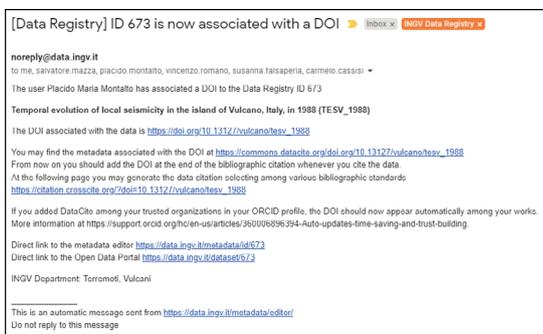


Figura 14 Un esempio di notifica ricevuta dai Responsabili Tecnologico e Scientifico di una nuova avvenuta pubblicazione.

Figure 14 A notification example received by both Technological and Scientific managers after a new publication.

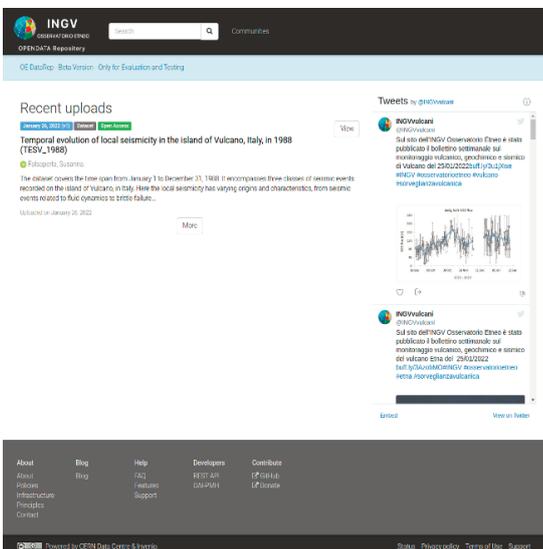


Figura 15 La pubblicazione tra gli uploads recenti nella home-page del portale DataRep.

Figure 15 The publication will be shown on the DataRep home page.

1.3 Utilizzo delle REST API

Il processo di pubblicazione di un *dataset* all'interno di DataRep prevede la creazione di un *deposito* [6]. Questa operazione avviene grazie all'**interazione** tra il Metadata Editor e DataRep. Il software Zenodo, su cui quest'ultimo è basato, permette di esporre delle API⁶ di tipo REST⁷

⁶ Con API, acronimo di *Application Programming Interface*, nel gergo informatico si indica un insieme di procedure (in

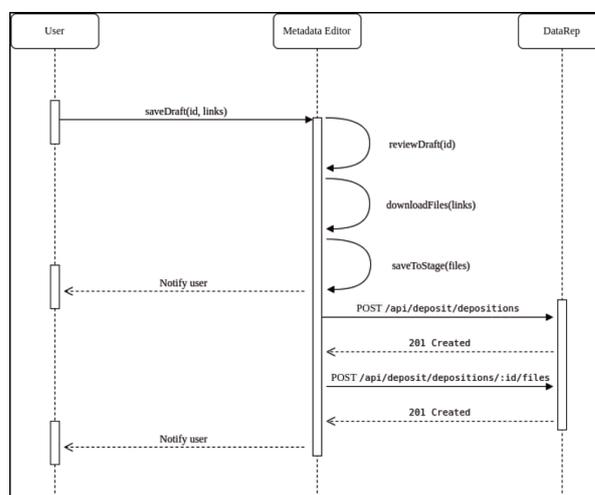
[21][22] che supportano:

- l'operazione di **Deposito**, quando attraverso il Metadata Editor l'utente effettua un *upload* del dato e dei metadati associati e, successivamente, la pubblicazione del *dataset*;
- la ricerca e la consultazione dei **Record** registrati e gestiti dal *backend* Zenodo di DataRep;
- le operazioni *download/upload* di file.

L'esito delle richieste eseguite verso le REST API viene notificato alla *business-logic* del Metadata Editor mediante l'utilizzo di codici standard *HTTP Response Status Code* [22]. Il Metadata Editor interpreta tali codici e intraprende l'operazione associata, ad esempio: in caso di errore, notificare l'utente e riprendere dall'ultimo stato salvato precedente, mentre in caso di un *response code* di successo, andare avanti con le altre operazioni previste.

Di seguito un elenco delle REST API esposte dal portale DataRep, proprie del *backend* del software Zenodo, ed utilizzate dalla componente di interazione del Metadata Editor.

Figura 16 Workflow sequenziale di esecuzione per la pubblicazione di un dataset sul portale DataRep.
Figure 16 Sequential diagram workflows for the publication of a dataset to DataRep.



1.3.1 Autenticazione (Authentication)

Tutte le richieste alle API di DataRep devono essere autenticate su una connessione HTTPS⁸; le richieste su protocollo HTTP, quindi senza cifratura, falliranno. L'autenticazione al portale è gestita tramite OAuth 2.0 [23].

Su DataRep è stato creato un utente che identifica l'Ufficio Gestione Dati dell'INGV (questo è attualmente l'unico utente presente sul portale). Per tale utente è stato creato un **Personal Access Token** necessario per autorizzare le chiamate verso le risorse REST del portale DataRep, effettuate dal Metadata Editor. Questo *Access Token* deve essere incluso in tutte le richieste all'interno di un *HTTP Authorization request header*⁹. Durante la creazione del *token* vengono specificati gli

genere raggruppate per strumenti specifici) atte all'espletamento di un dato compito; spesso tale termine designa le librerie software di un linguaggio di programmazione.

⁷ REST API *Representational state transfer* (REST) è uno stile architetturale per sistemi distribuiti, rappresenta un sistema di trasmissione di dati su HTTP senza ulteriori livelli. I sistemi REST non prevedono il concetto di sessione, ovvero sono *stateless*.

⁸ In telecomunicazioni e informatica l'*HyperText Transfer Protocol* (HTTP) over *Secure Socket Layer* (HTTPS), anche noto come HTTP over TLS, HTTP over SSL e HTTP Secure è un protocollo per la comunicazione sicura attraverso una rete di computer utilizzato su Internet.

⁹ Il campo *HTTP Authorization request header*, in una richiesta HTTP, contiene le informazioni necessarie ad autorizzare l'utilizzo di una risorsa esposta da un server.

OAuth scopes¹⁰, ovvero le autorizzazioni ad eseguire determinate operazioni (ad es. gli scopes: *deposit:write* e *deposit:actions*, per permettere la creazione di un deposito e la sua modifica).

1.3.2 Deposito (Deposition)

La risorsa esposta dalle API di *deposition* (vedi Tabella 1) è principalmente usata per le operazioni di upload e modifica dei *metadati* di una **risorsa di deposito** in Zenodo. A tale risorsa, verrà associato uno spazio di archiviazione univoco e dedicato alla memorizzazione del dato e di tutte le informazioni associate ad esso necessarie alla rappresentazione logica sulla *landing-page*.

La risorsa REST esposta dal portale DataRep consente un completo ciclo di operazioni di tipo CRUD [24] - *create, retrieve, update, delete* - del dato; quelle utilizzate in questo ambito sono riportate in Tabella 1.

	HTTP Request	URL	Header	Scopes	Response
CREATE	POST	/api/deposit/depositions	Content-Type: application/json Authorization: Bearer <ACCESS_TOKEN>	deposit:write	Code: 201 Created Body: a deposition resource.
UPDATE	PUT	/api/deposit/depositions/: id	Content-Type: application/json Authorization: Bearer <ACCESS_TOKEN>	deposit:write	Code: 201 Created Body: a deposition resource.

Tabella 1 La risorsa REST deposito, per la creazione e modifica del deposito sul portale DataRep.

Table 1 REST Deposit resource to create and modify a deposit on DataRep.

1.3.3 Deposito di File (Deposition files)

La risorsa *Deposition file* viene usata per effettuare il caricamento (*uploading*) e la modifica (*editing*) dei file associati ad un precedente deposito *deposition*. Le API REST impiegate in questa fase sono riportate in Tabella 2.

	HTTP Request	URL	Header	Scopes	Success Response
CREATE	POST	/api/deposit/depositions/: id/files	Content-Type: multipart/form-data Authorization: Bearer <ACCESS_TOKEN>	deposit:write	Code: 201 Created Body: a deposition resource.

Tabella 2 La risorsa REST deposition file per l'upload dei file sul relativo deposito.

Table 2 The REST deposition file resource to upload a file on the appropriate deposit.

¹⁰ Gli OAuth Scopes sono un meccanismo implementato in OAuth 2.0 [23] per limitare l'accesso di un'applicazione all'account di un utente. Un'applicazione può richiedere uno o più ambiti, il token di accesso rilasciato all'applicazione sarà limitato agli ambiti concessi.

1.3.4 Azioni di Deposito (Deposition Actions)

La risorsa *Deposition Actions* (Tabella 3) permette di eseguire azioni specifiche sul deposito precedentemente creato quali, nel caso in esame, l'effettiva **pubblicazione del deposito** (*publish*) che rende pubblica la *landing-page* sul portale e la sua eventuale modifica (*edit*).

	HTTP Request	URL	Header	Scopes	Success Response
PUBLISH	POST	/api/deposit/depositions/:id/actions/publish	Content-Type: multipart/form-data Authorization: Bearer <ACCESS_TOKEN>	deposit:actions	Code: 202 Accepted Body: a deposition resource.
EDIT	POST	/api/deposit/depositions/:id/actions/edit	Content-Type: multipart/form-data Authorization: Bearer <ACCESS_TOKEN>	deposit:actions	Code: 201 Created Body: a deposition resource.

Tabella 3 La risorsa REST deposition actions usate, per la pubblicazione e modifica della landing-page su DataRep.
Table 3 The deposition actions REST resource to publish and edit a landing-page on DataRep.

2. DataRep: metodologia di implementazione

Per l'implementazione del *repository* dei *dataset* basato sul software Zenodo, sono state seguite le sue linee guida ufficiali relative all'installazione o **deployment** [25]. Con il termine *deploy* (*deployment* o *deploying*) ci si riferisce al processo di installazione dei *software* che partecipano alla composizione del sistema secondo i moderni *pattern* dettati dall'ingegneria del *software* (in particolare quello che prevede l'utilizzo di *Micro Services* [26] [27]).

Le linee guide prevedono una duplice installazione del *software*:

- una versione di sviluppo (**Development**), necessaria per le modifiche e per le aggiunte di moduli supplementari;
- una versione denominata di produzione (**Production**), che rappresenta l'implementazione in uso del *repository* Zenodo.

La strategia adottata dall'INGV-OE per il *deployment* del *repository* **DataRep** prevede un approccio suddiviso in tre fasi. In particolare:

- **Fase 1 Deployment** del *software* Zenodo, seguendo le sopra citate linee guida, per creare un ambiente di sviluppo al fine di familiarizzare con l'ambiente e con le API di programmazione esposte dal *backend*. Questo *deployment* è in esecuzione su un'istanza virtuale - *Virtual Machine* (VM) - disponibile presso il centro di calcolo dell'INGV-OE all'interno del *cluster*¹¹ basato su *hypervisor*¹² Hyper-V di Microsoft [28] [29]. Durante questa fase sono state sviluppate funzionalità supplementari quali, ad esempio, l'integrazione con il Metadata Editor all'interno di un ambiente di test sicuro. Inoltre sono state apportate alcune modifiche grafiche all'interfaccia utente - implementata con il *framework* Javascript AngularJS [30] - personalizzandola in modo da rispondere ai requisiti del caso d'uso del Portale Dati dell'Osservatorio Etneo.

¹¹ In informatica un *computer cluster*, o più semplicemente un cluster (dall'inglese grappolo), è un insieme di computer connessi tra loro tramite una rete telematica, con lo scopo di distribuire un'elaborazione molto complessa tra i vari computer.

¹² L'*hypervisor*, noto anche come monitor della macchina virtuale, è un processo che crea e gestisce le macchine virtuali.

- **Fase 2 Deployment** di due istanze di Zenodo al fine di ottenere un'installazione adatta alla produzione, distribuita tramite *Docker containers* [31], su VMs in Alta Disponibilità¹³ o *High Availability* [32] [33] [34] all'interno del *cluster* del CED¹⁴ dell'INGV-OE. In questo modo si mira ad avere sia un ambiente di pre-produzione per testare le modifiche e le componenti sviluppate, sia un ambiente di produzione che rappresenta ad oggi la prima implementazione pubblica della piattaforma DataRep raggiungibile al link <https://oedatarep.ct.ingv.it>.
- **Fase 3** Contestualmente alla Fase 2, è iniziata una fase di ricerca e sviluppo per studiare la fattibilità riguardo la configurazione, l'installazione e messa in produzione di un *cluster* basato su *Kubernetes* [35] [36] [37] [38], atto ad ospitare come primo caso d'uso il *software* Zenodo impiegato nell'implementazione di DataRep. La nuova infrastruttura basata su *Kubernetes* servirà come *host* per il *porting* dell'istanza Zenodo, precedentemente istanziata su VMs (Fase 2), al fine di garantire una Alta Disponibilità - HA - dei servizi istanziati su *software containers* di tipo *Docker*. Si prevede una installazione di un *cluster* *Kubernetes* di tipo locale, ospitata presso il CED dell'Osservatorio Etneo.

3. DataRep: architettura *software*

L'architettura *software* di DataRep, comprende diverse componenti che interagiscono tra loro al fine di implementare un *repository* dati accessibile tramite una standard *web-application*. Per garantire l'isolamento di questi componenti *software* dal sistema operativo *host*, essi vengono eseguiti tramite *Docker*¹⁵ *containers*.

L'orchestrazione¹⁶ di questi *containers*, viene gestita tramite l'utilizzo dello strumento per applicazioni *multi-container* *Docker Compose* [39].

La definizione delle caratteristiche dei *Docker containers* (*services*), oltre che la loro configurazione (variabili d'ambiente), viene descritta all'interno di uno specifico file *YAML* [40]. *Docker Compose* ne interpreta il contenuto e trova le descrizioni necessarie per le risorse logiche da assegnare ad ogni *container*, quali ad esempio: i volumi da usare all'interno di un *container*, le porte *TCP* da esporre a livello *host*, ecc... Infine, è ruolo del *Docker Compose* eseguire i servizi seguendo la gerarchia di processi descritta nel file *YAML* che rispecchia il grafico riportato in Figura 17. Tale schema, facente parte della documentazione ufficiale del *software* *Invenio*, rappresenta graficamente la disposizione dei servizi sopra citati all'interno di un *deployment* di DataRep.

L'elenco di seguito descrive i servizi che compongono il sistema:

- **Load balancers:** un servizio (o più) di *load-balancing* per gestire le richieste *HTTP* in entrata, implementato dal *software* **HAProxy** [41].
- **Web servers:** un insieme di *web servers* che implementano il *frontend* per l'applicazione *uWSGI*¹⁷ di **Flask** [42] fornito da uno (o più) istanze del *software* **Nginx** [43] che opera da *upstream server*¹⁸.

¹³ Viene definita Alta disponibilità (in inglese *High Availability*) o HA, l'insieme tutte le disposizioni che puntano a garantire la disponibilità di un servizio, cioè assicurare il buon funzionamento di un servizio 24/7.

¹⁴ Un CED Centro Elaborazione Dati o in inglese *data center*, è un'unità organizzativa all'interno di un'organizzazione aziendale che coordina e mantiene le apparecchiature ed i servizi di gestione dei dati, ovvero l'infrastruttura informatica a servizio dell'organizzazione stessa.

¹⁵ *Docker* è un progetto *open-source* che automatizza il processo di *deployment* di applicazioni all'interno di contenitori *software*, fornendo un'astrazione aggiuntiva grazie alla virtualizzazione a livello di sistema operativo.

¹⁶ Si definisce orchestrazione (in ambito informatico) la configurazione, la gestione e il coordinamento automatico di sistemi informatici, applicazioni e servizi. L'orchestrazione è una metodologia che aiuta i team IT a gestire più facilmente attività e flussi di lavoro complessi.

¹⁷ *uWSGI* è un'applicazione *software* usata per servire *web application* scritte in *Python* in combinazione con un *web server*, ad es. *Nginx* che offre supporto nativo al protocollo *Web Server Gateway Interface* *WSGI*, per la comunicazione tra *web server* e *web applications*.

- **Application servers:** che ospitano il *core* dell'applicazione, ovvero la logica che implementa il *repository* di dati basata sul framework Invenio, scritta con l'ausilio del *micro web framework Flask*.
- **Distributed task queue:** un set di *workers Celery* [44] [45] per l'esecuzione di *task* asincroni; un'istanza del software **Flower** [46] che rappresenta un'interfaccia web per il monitoraggio dei workers.
- **Database:** un database RDBMS¹⁹ basato su **PostgreSQL** [47].
- **Search engine:** un'istanza del motore di ricerca **ElasticSearch** [48].
- **Message queue:** un'istanza del *message broker RabbitMQ* [49].
- **Cache system:** un'istanza del *key-value store Redis* [50].

La richiesta di accesso alle risorse del portale da parte di un *client*, sia per la visualizzazione *web* che l'inserimento/modifica dei contenuti tramite API, viene presa in carico dal *load-balancer* HAProxy che si occuperà di inoltrarle ai *web-servers* (Nginx). Ricevute le richieste, diventa responsabilità del *web-server* quella di inoltrarle a sua volta alla *business-logic* dell'applicazione DataRep in Python uWSGI, la quale esegue i moduli *core* del *framework* Invenio su cui DataRep si basa.

Lo stato del sistema, comprensivo di ogni azione effettuata sul portale (visualizzazione pagine, scaricamento dati, accesso alle risorse tramite API), nonché i dettagli strutturati di archiviazione dei dati (*path* locale al dato archiviato, identificativo univoco del *record* associato e relativi metadati) vengono memorizzati all'interno di un database relazionale di tipo PostgreSQL. Il dato ed i metadati associati ad un *record* vengono memorizzati come documento JSON²⁰ all'interno del database.

Una *cache* delle richieste viene gestita tramite il componente invenio-cache di Zenodo che fa utilizzo di un *key-value store* residente in memoria di tipo **Redis**.

L'interoperabilità tra le componenti software interne di una istanza di DataRep, viene implementata tramite un sistema di *message-broker* con la gestione asincrona delle code delle operazioni di tipo **AMQP**²¹ implementato dal software **RabbitMQ** e dai nodi *workers* (implementati dal software **Celery**), i quali "consumano" i messaggi pubblicati nella coda portando a termine le operazioni richieste. Queste ultime riguardano principalmente la gestione delle indicizzazioni dei contenuti (dati e metadati associati) per le future ricerche strutturate effettuate tramite il motore **ElasticSearch**, che è utilizzato da Zenodo come motore di ricerca sottostante. Quest'ultimo è infatti *JSON-based*, e quindi pienamente compatibile con l'architettura di memorizzazione del record in formato JSON all'interno del database sopra descritto.

¹⁸ Con *upstream server* ci si riferisce ad un server che fornisce servizi ad un altro server, ovvero un server che gerarchicamente posizionato più in alto rispetto ad altri servers.

¹⁹ Un RDBMS – *Relational Database Management System*, è, letteralmente, un sistema per la gestione di database relazionali. Un sistema di gestione di database impartisce istruzioni al sistema operativo per facilitare l'accesso ai dati e strutturarne correttamente l'archiviazione.

²⁰ In informatica, nell'ambito della programmazione *web*, JSON acronimo di *JavaScript Object Notation*, è un formato adatto all'interscambio di dati fra applicazioni *client/server*.

²¹ *Advanced Message Queuing Protocol* (AMQP) è uno standard aperto che definisce un protocollo a livello applicativo per il *message-oriented middleware*, che garantisce funzionalità di messaggistica, accordamento, routing (con paradigmi punto-punto e pubblicazione-sottoscrizione), affidabilità e sicurezza.

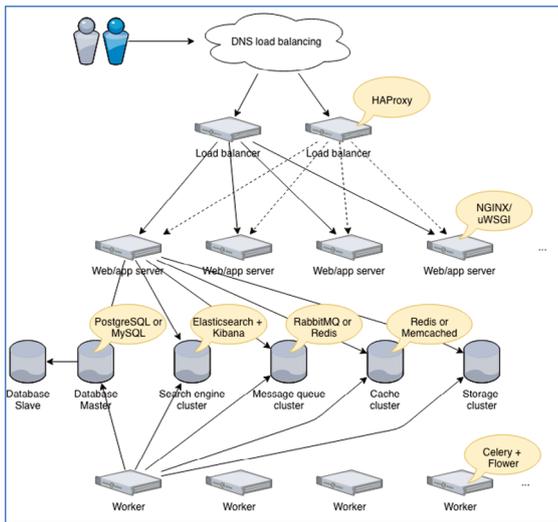


Figura 17 Schema logico che descrive la intercomunicazione tra i processi che compongono un deployment completo di Zenodo. Ogni servizio nella figura viene eseguito all'interno di uno specifico Docker container.

Figure 17 Logical diagram describing the intercommunication between the processes that make up a complete Zenodo deployment. Each service in the figure run into a Docker container.

4. Conclusione e sviluppi futuri

Il portale dati DataRep è attualmente online e raggiungibile seguendo la URL <https://oedatarep.ct.ingv.it>.

Come già descritto al paragrafo 2.1, attualmente l'insieme dei servizi *software* che compongono DataRep vengono eseguiti su una *Virtual Machine* all'interno di un *hypervisor* Hyper-V del Centro di Calcolo dell'Osservatorio Etneo, che è ridondata a livello di istanza ma non a livello di servizi *software*. Seguendo le fasi operazionali descritte al suddetto paragrafo, da un punto di vista operativo si prevede di scalare il *deployment* in modo da garantire un'alta disponibilità dei servizi forniti nei *container* a livello di *runtime*. L'evoluzione naturale per avere una gestione dei servizi *software* in alta disponibilità è quello di passare ad un sistema di orchestrazione di *container* come Kubernetes [35] [36] [37] [38].

Inoltre, per una totale integrazione del prototipo DataRep con i servizi *software* attualmente in uso presso l'Osservatorio Etneo, si prevede l'integrazione con *repository* di dati *raw* e l'integrazione con la piattaforma TSDSystem [51].

Glossario

Dataset Un *dataset* è una collezione di dati, più comunemente costituisce un insieme di dati strutturati in forma relazionale, cioè corrisponde al contenuto di una singola tabella di base di dati, oppure ad una singola matrice di dati statistici, in cui ogni colonna della tabella rappresenta una particolare variabile, e ogni riga corrisponde ad un determinato membro del dataset in questione. Il termine dataset può essere usato anche più genericamente, per indicare i dati in un insieme di tabelle strettamente connesse, relative ad un particolare esperimento o evento. (Fonte Wikipedia)

Metadato Un metadato - letteralmente "(dato) per mezzo di un (altro) dato" - è un'informazione che descrive un insieme di dati. Un esempio tipico di metadati è costituito dalla scheda del catalogo di una biblioteca, la quale contiene informazioni circa il contenuto e la posizione di un libro, cioè dati riguardanti più dati che si riferiscono al libro. (Fonte Wikipedia)

Repository Un repository, in informatica, è un ambiente di un sistema informativo, in cui vengono gestiti i metadati, attraverso tabelle relazionali; l'insieme di tabelle, regole e motori di calcolo tramite cui si gestiscono i metadati prende il nome di metabase. Si tratta di un ambiente che può essere implementato attraverso numerose piattaforme hardware e sistemi di gestione

delle basi di dati seguendo uno stile architeturale basato su sottosistemi software che accedono e modificano una singola struttura dati chiamata appunto repository. (Fonte Wikipedia)

REST API Representational state transfer (REST) è uno stile architeturale per sistemi distribuiti. Il termine REST rappresenta un sistema di trasmissione di dati su HTTP senza ulteriori livelli. I sistemi REST non prevedono il concetto di sessione, ovvero sono stateless. L'architettura REST si basa su HTTP. Il funzionamento prevede una struttura degli URL ben definita che identifica univocamente una risorsa o un insieme di risorse e l'utilizzo dei metodi HTTP specifici per il recupero di informazioni (GET), per la modifica (POST, PUT, PATCH, DELETE). (Fonte Wikipedia)

Sitografia

- [1] Geolocalizzazione, <https://it.wikipedia.org/wiki/Geolocalizzazione>
- [2] Principi della Politica dei Dati dell'INGV, <https://data.ingv.it/docs/principi/index.html>
- [3] Open Data - Dati Aperti, https://it.wikipedia.org/wiki/Dati_aperti
- [4] Open Access, https://it.wikipedia.org/wiki/Open_access
- [5] Zenodo - Research. Shared, <https://www.zenodo.org/>
- [6] Deposito istituzionale, https://it.wikipedia.org/wiki/Deposito_istituzionale
- [7] INGV, INGV - Earth-prints, <https://www.earth-prints.org/>
- [8] Documenti relativi alla Politica dei Dati dell'INGV, <https://istituto.ingv.it/it/ufficio-gestione-dati#documenti>
- [9] Ufficio Gestione Dati INGV, <https://istituto.ingv.it/it/50-archivi-e-banche-dati/1760-ufficio-gestione-dati.html>
- [10] Portale Dati Aperti - INGV, <https://data.ingv.it>
- [11] CERN Home, <https://home.cern/>
- [12] OpenAIRE, <https://www.openaire.eu/>
- [13] DOI - Digital object identifier, https://en.wikipedia.org/wiki/Digital_object_identifier
- [14] Landing Page, https://en.wikipedia.org/wiki/Landing_page
- [15] INGV Metadata Editor, <https://data.ingv.it/metadata>
- [16] DataCite Metadata Schema, <https://schema.datacite.org/>
- [17] FAIR Principles - GO FAIR, <https://www.go-fair.org/fair-principles/>
- [18] Cloud storage for Work and Home - Google Drive, <https://drive.google.com/>
- [19] Formato aperto, https://it.wikipedia.org/wiki/Formato_aperto
- [20] Principi della Politica dei Dati dell'INGV, <https://data.ingv.it/docs/attuazione/index.html#procedura-di-inserimento-di-elementi>
- [21] Developers | Zenodo, <https://developers.zenodo.org/>
- [22] HTTP response status codes, <https://developer.mozilla.org/en-US/docs/Web/HTTP/Status>
- [23] OAuth 2.0, <https://oauth.net/2/>
- [24] Create, read, update and delete, https://en.wikipedia.org/wiki/Create,_read,_update_and_delete
- [25] Installation - Zenodo 3.0.0 documentation, <https://zenodo.readthedocs.io/en/latest/installation.html>
- [26] Microservices, <https://en.wikipedia.org/wiki/Microservices>
- [27] What are microservices?, <https://microservices.io/>
- [28] Hyper-V Microsoft - Wikipedia.org, <https://en.wikipedia.org/wiki/Hyper-V>
- [29] Hyper-V Technology Overview - Microsoft Docs, <https://docs.microsoft.com/en-us/windows-server/virtualization/hyper-v/hyper-v-technology-overview>
- [30] AngularJS - Superheroic JavaScript MVW Framework, <https://angularjs.org/>
- [31] Empowering App Development for Developers | Docker, <https://www.docker.com/>

- [32] High availability - Wikipedia.org, https://en.wikipedia.org/wiki/High_availability
- [33] What is High Availability? - DigitalOcean.com, <https://www.digitalocean.com/community/tutorials/what-is-high-availability>
- [34] High Availability Book Solutions - Dell Labs, <https://i.dell.com/sites/doccontent/business/solutions/operating-systems/en/Documents/high-availability-solutions.pdf>
- [35] Kubernetes." kubernetes.io, <https://kubernetes.io/>
- [36] Scalable web application on Kubernetes." cloud.ibm.com, <https://cloud.ibm.com/docs/solution-tutorials?topic=solution-tutorials-scalable-webapp-kubernetes>
- [37] Deploying a scalable web application with Docker and Kubernetes, <https://medium.com/better-practices/deploying-a-scalable-web-application-with-docker-and-kubernetes-a5000a06c4e9>
- [38] Synthetic procedure to replicate the INFN Open Access Repository (based on Invenio v3 and Zenodo), <https://www.openaccessrepository.it/record/73184>
- [39] Overview of Docker Compose, <https://docs.docker.com/compose>
- [40] The Official YAML Web Site, <https://yaml.org/>
- [41] HAProxy - The Reliable, High Performance TCP/HTTP Load Balancer, <http://www.haproxy.org/>
- [42] Flask | The Pallets Projects, <https://palletsprojects.com/p/flask/>
- [43] NGINX | High Performance Load Balancer, Web Server, & Reverse Proxy, <https://www.nginx.com/>
- [44] Workers Guide - Celery 5.0.5 documentation, [docs.celeryproject.org, https://docs.celeryproject.org/en/stable/userguide/workers.html](https://docs.celeryproject.org/en/stable/userguide/workers.html)
- [45] How to run periodic tasks in Celery, <https://antoniodimariano.medium.com/how-to-run-periodic-tasks-in-celery-28e1abf8b458>
- [46] Flower - Celery monitoring tool, <https://flower.readthedocs.io/en/latest/>
- [47] PostgreSQL." PostgreSQL: The world's most advanced open source database, <https://www.postgresql.org/>
- [48] Free and Open Search: The creators of Elasticsearch, ELK & Kibana | Elastic, <https://www.elastic.co/>
- [49] Messaging that just works - RabbitMQ, <https://www.rabbitmq.com/>
- [50] Redis, <https://redis.io/>
- [51] TSDSystem - INGV, <http://tsd.ct.ingv.it/tsdws/masterView.php>
- [52] What is Kibana? | Elastic, <https://www.elastic.co/what-is/kibana>
- [53] Overview of Docker Compose, <https://docs.docker.com/compose/>
- [54] Web Server Gateway Interface - Wikipedia, https://en.wikipedia.org/wiki/Web_Server_Gateway_Interface
- [55] Invenio Framework v3, <https://invenio.readthedocs.io/en/latest/>

QUADERNI di GEOFISICA

ISSN 1590-2595

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/quaderni-di-geofisica.html/>

I QUADERNI DI GEOFISICA (QUAD. GEOFIS.) accolgono lavori, sia in italiano che in inglese, che diano particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari che necessitano di rapida diffusione nella comunità scientifica nazionale ed internazionale. Per questo scopo la pubblicazione on-line è particolarmente utile e fornisce accesso immediato a tutti i possibili utenti. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi. I QUADERNI DI GEOFISICA sono presenti in "Emerging Sources Citation Index" di Clarivate Analytics, e in "Open Access Journals" di Scopus.

QUADERNI DI GEOFISICA (QUAD. GEOFIS.) welcome contributions, in Italian and/or in English, with special emphasis on preliminary elaborations of data, measures, and observations that need rapid and widespread diffusion in the scientific community. The on-line publication is particularly useful for this purpose, and a multidisciplinary Editorial Board with an accurate peer-review process provides the quality standard for the publication of the manuscripts. QUADERNI DI GEOFISICA are present in "Emerging Sources Citation Index" of Clarivate Analytics, and in "Open Access Journals" of Scopus.

RAPPORTI TECNICI INGV

ISSN 2039-7941

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html/>

I RAPPORTI TECNICI INGV (RAPP. TEC. INGV) pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico come manuali, software, applicazioni ed innovazioni di strumentazioni, tecniche di raccolta dati di rilevante interesse tecnico-scientifico. I RAPPORTI TECNICI INGV sono pubblicati esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi.

RAPPORTI TECNICI INGV (RAPP. TEC. INGV) publish technological contributions (in Italian and/or in English) such as manuals, software, applications and implementations of instruments, and techniques of data collection. RAPPORTI TECNICI INGV are published online to guarantee celerity of diffusion and a prompt access to published data. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

MISCELLANEA INGV

ISSN 2039-6651

http://istituto.ingv.it/it/le-collane-editoriali-ingv/miscellanea-ingv.html

MISCELLANEA INGV (MISC. INGV) favorisce la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV. In particolare, MISCELLANEA INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc. La pubblicazione è esclusivamente on-line, completamente gratuita e garantisce tempi rapidi e grande diffusione sul web. L'Editorial Board INGV, grazie al suo carattere multidisciplinare, assicura i requisiti di qualità per la pubblicazione dei contributi sottomessi.

MISCELLANEA INGV (MISC. INGV) favours the publication of scientific contributions regarding the main activities carried out at INGV. In particular, MISCELLANEA INGV gathers reports of scientific projects, proceedings of meetings, manuals, relevant monographs, collections of articles etc. The journal is published online to guarantee celerity of diffusion on the internet. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

Coordinamento editoriale

Francesca DI STEFANO
Istituto Nazionale di Geofisica e Vulcanologia

Progetto grafico

Barbara ANGIONI
Istituto Nazionale di Geofisica e Vulcanologia

Impaginazione

Barbara ANGIONI
Patrizia PANTANI
Massimiliano CASCONI
Istituto Nazionale di Geofisica e Vulcanologia

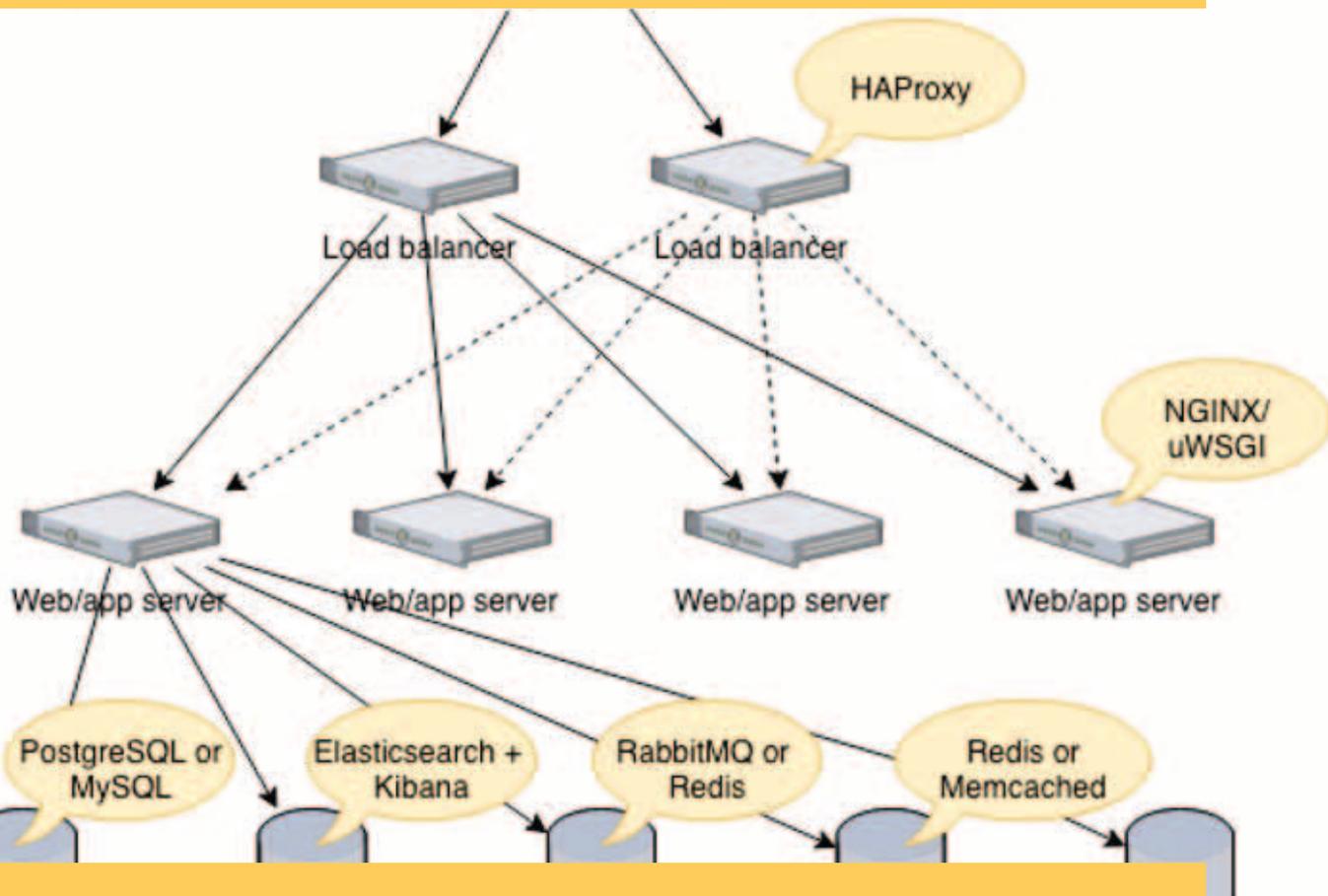
©2022

Istituto Nazionale di Geofisica e Vulcanologia
Via di Vigna Murata, 605
00143 Roma
tel. +39 06518601

www.ingv.it



Creative Commons Attribution 4.0 International (CC BY 4.0)



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA



DNS load balancing