

Il percorso di adozione del paradigma Open Science all'INGV

Mario Locati, Francesco Mariano Mele, Placido Montalto, Vincenzo Romano, Valentino Lauciani, Roberto Vallone, Stefano Cacciaguerra

Istituto Nazionale di Geofisica e Vulcanologia (INGV)

Abstract. L'Istituto Nazionale di Geofisica e Vulcanologia (INGV) svolge attività di ricerca scientifica nell'ambito delle Scienze della Terra e attività all'interno del sistema nazionale di Protezione Civile grazie all'uso delle sue numerose reti di monitoraggio multiparametrico dislocate sull'intero territorio nazionale. L'INGV intende adottare il paradigma dell'Open Science, ovvero un approccio alla Ricerca basato su una più veloce ed efficiente condivisione dei dati per favorire nuovi scenari multidisciplinari inesplorati. In questo contesto, l'INGV si è dotato di una Politica dei Dati che predispone gli strumenti gestionali necessari all'attuazione dell'Open Science e dei principi FAIR e demanda il coordinamento di queste attività all'Ufficio Gestione Dati. Ogni azione è intrapresa considerando i processi scientifici, tecnologici e gestionali che gravitano intorno al dato, per crearlo, mantenerlo, condividerlo, conservarlo, tracciarne la provenienza e l'uso, e infine misurarne l'impatto

Keywords. Open Science, Data Policy, Open Data, FAIR Data, EOSC

Introduzione

Lo scambio di dati nell'ambito delle Scienze della Terra è parte integrante di questa disciplina dato che i fenomeni naturali ignorano i confini amministrativi. Da sempre le comunità attive nel settore condividono informazioni beneficiando dell'evoluzione tecnologica perché i dati raccolti, analizzati e condivisi crescono in quantità e sono sempre più complessi. Già negli anni '80 del secolo scorso INGV, allora chiamato ING, realizzò la prima rete di sismica attiva con qualche decina di stazioni su scala mediterranea. Oggi INGV gestisce reti di monitoraggio multi-parametrico con diverse centinaia di stazioni dislocate sull'intero territorio nazionale e si avvale inoltre di dati da altre reti di monitoraggio gestite da altre organizzazioni come il Dipartimento di Protezione Civile o reti satellitari gestite dall'Agenzia Spaziale Europea e dalla NASA.

La standardizzazione dei processi per condividere i dati dell'INGV è una sfida complessa che copre varie aree disciplinari condotte in una struttura distribuita geograficamente su sette sezioni principali (Milano, Bologna, Pisa, Roma, Napoli, Palermo e Catania) e ventuno altre sedi distaccate sparse in tutta Italia.

È importante sottolineare come le informazioni raccolte sui fenomeni naturali osservati siano estremamente preziose perché per loro stessa natura uniche e irripetibili. È quindi fondamentale adottare soluzioni di conservazione dei dati con un orizzonte temporale di lungo periodo.

La ricerca condotta all'INGV adotta il paradigma dell'Open Science (Nielsen 2011) dal

momento che uno dei cardini della propria attività è la condivisione il più possibile libera, semplice, veloce ed efficiente della conoscenza, sfruttando le moderne tecnologie di comunicazione. La condivisione delle informazioni deve permettere a chiunque un semplice utilizzo dei dati nel maggior numero possibile di ambiti scientifici a potenziale beneficio di tutta la comunità.

La fase preliminare

Per attuare i principi dell'Open Science servono soluzioni concrete che possano essere adottate dai ricercatori minimizzando eventuali impatti negativi sulle loro attività quotidiane (Meadows 2018). Servono pratiche efficaci che permettano di migliorare il lavoro dei ricercatori, ad esempio per meglio mettere a frutto l'enorme e complesso patrimonio di conoscenza oggi disponibile ma spesso di difficile accesso e uso. A prescindere dagli strumenti adottati, il loro successo passa da un inquadramento amministrativo-gestionale che ne stabilisca finalità, tempistiche, costi, ruoli e responsabilità la cui evoluzione potrà comportare il cambiamento di processi gestionali consolidati.

L'adozione dei principi dell'Open Science è ovunque in fase sperimentale e per la sua attuazione è difficile trovare sia esempi pratici da esaminare, sia l'ingaggio di personale qualificato. Al fine di avviare una sperimentazione, INGV si è avvalso di figure professionali già presenti nell'istituto che hanno seguito un percorso auto-formativo recuperando informazioni e documentazione dalle fonti più disparate di ambito scientifico, normativo e informatico.

Fin dal 2012-2013 all'INGV si iniziò a discutere di quali fossero le possibili strade da percorrere, discussioni condotte sia all'interno dell'istituto, sia con altre organizzazioni europee nell'ambito di progetti come FP7 EUDAT e FP7 European Plate Observing System (EPOS). Queste discussioni culminarono nel 2013 con l'attivazione del servizio di assegnazione di codici DOI rilasciati da DataCite grazie a una convenzione con la Conferenza dei Rettori delle Università Italiane (CRUI). Nonostante l'identificazione dei dati tramite DOI fosse disponibile, mancava tuttavia una strutturazione delle procedure scientifiche e amministrative per la validazione delle richieste dei colleghi che volevano condividere i propri dati. Emerse la necessità di dotarsi di una formalizzazione delle procedure da un lato, e dall'altro di avere una visione istituzionale dei principi che avrebbero dovuto guidare la condivisione dei dati per il prossimo futuro. Nel 2015 il Presidente INGV costituì un gruppo multidisciplinare composto da ricercatori, informatici e legali con il compito di redigere una Politica dei Dati istituzionale.

Una delle prime attività effettuate fu la mappatura del panorama normativo a livello nazionale, europeo e internazionale. Parallelamente si fece un censimento dei dati scientifici disponibili presso ciascuna sezione, in modo da avere contezza sia della varietà, sia della quantità di dati in gioco.

La redazione della Politica dei Dati durò dal 2015 al 2018 e si compone di tre documenti approvati dal Consiglio di Amministrazione dell'INGV. Un primo documento venne approvato nel 2016 e sancisce i Principi generali che si volevano perseguire. Un secondo documento venne approvato nel 2017 e tratta i principi dell'Open Access rivolti principal-

mente alle pubblicazioni scientifiche. Un terzo documento approvato nel 2018 descrive infine le modalità di attuazione di strumenti della Politica dei Dati che permettono la distinzione tra pubblicazioni di articoli scientifici e dati (Fig. 1).

Alcuni elementi presenti nella Politica dei Dati dell'INGV traggono spunto da indicazioni presenti in documenti redatti in vari ambiti della Ricerca e della Pubblica Amministrazione, sia a livello nazionale che internazionale. Ad esempio la costituzione di un "Ufficio Gestione Dati" che si occupi dell'attuazione della Politica dei Dati è stato ispirato dal "Gruppo di lavoro Open Data" descritto nelle Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo pubblico, a cura dell'Agenzia Digitale per l'Italia, la cui redazione è in continua evoluzione e aperta a nuovi contributi. L'adozione di un "Registro Dati" (<https://data.ingv.it>), che tra le funzioni svolte assegna anche identificativi univoci DOI secondo le specifiche DataCite, ha preso spunto da diversi cataloghi di metadati, come: quelli diffusi in ambito delle Direttive Europee INSPIRE che trattano la circolazione dei dati territoriali e dalle indicazioni fornite dalla Research Data Alliance. L'adozione di una catalogazione di dati basata su livelli di elaborazione e l'adozione di licenze di tipo Creative Commons è frutto di suggerimenti maturati nell'ambito dell'ERIC EPOS, l'European Plate Observing System e della sua Politica dei Dati.

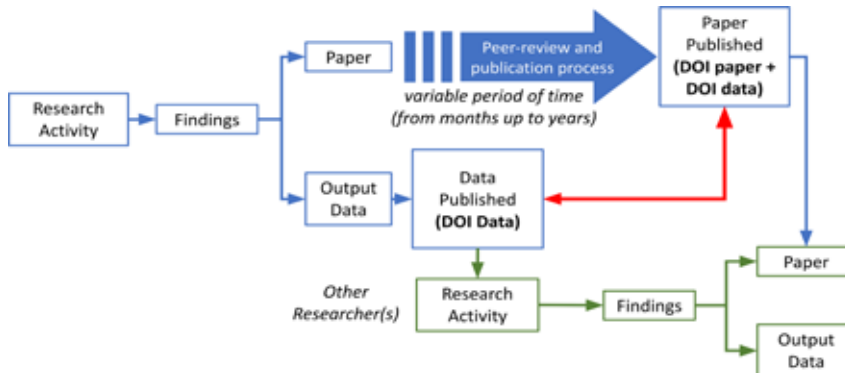


Fig. 1
Condivisione dei
risultati della Ricerca
tramite la separazione
tra la pubblicazione di
articoli e dati scientifici

La fase attuativa

L'attuazione della Politica dei Dati è iniziata nel novembre 2018 ed è in capo all'Ufficio Gestione Dati i cui componenti dedicano parte del proprio tempo a supporto di queste attività istituzionali. L'Ufficio è costituito da un Coordinatore, da un Referente per ciascuna delle tre aree scientifiche dell'istituto (i.e. Terremoti, Vulcani e Ambiente), e da personale informatico esperto nella gestione di infrastrutture digitali e di banche dati. Le attività dell'Ufficio sono pianificate su base annuale e vertono sulla progressiva adozione e diffusione dei principi FAIR (Findable, Accessible, Interoperable, Reusable) dei dati (Wilkinson et al. 2016), a cui si tende ad aggiungere il concetto di Reproducibility.

Lo strumento cardine per l'attuazione della Politica dei Dati è il Registro Dati, definibile in termini informatici come un "catalogo di metadati". Questo strumento è stato progettato e realizzato per interfacciarsi con i sistemi informativi dell'istituto come quello del personale e dei progetti di ricerca finanziati da altre organizzazioni. I metadati in esso contenuti

non seguono uno specifico standard, si tratta infatti di un'aggregazione di informazioni con standard diversi. Al momento sono stati sviluppati moduli di formattazione ed esportazione di metadati per supportare:

- lo standard DataCite per l'assegnazione di codici DOI;
- lo standard ISO 19115/19139 per la pubblicazione di metadati nel Repertorio Nazionale di Dati Territoriali;
- il profilo nazionale dei metadati DCAT-AP secondo le indicazioni dell'Agenzia per l'Italia Digitale.

Il Registro Dati mira a semplificare la "Findability" dei dati, in cui oltre alla descrizione dei dati condivisi, sono chiarite le regole di accesso e riutilizzabilità grazie all'apposizione di Licenze Creative Commons. Dal novembre 2020 il contenuto del Registro Dati è consultabile all'indirizzo <https://data.ingv.it> ed è basato sul software Open Source CKAN (<https://ckan.org/>).

Si mira a raggiungere l'interoperabilità dei dati e dei servizi di accesso ai dati tramite la progressiva adozione di servizi web che rispettino standard disciplinari come FDSN, International Federation of Digital Seismograph Networks, standard generici come OGC, Open Geospatial Consortium e le Direttive INSPIRE per i dati territoriali.

Le attività condotte per l'attuazione della Politica dei Dati dell'INGV vengono anche concretamente supportate dalla partecipazione attiva dell'Ente a iniziative quali la San Francisco Declaration on Research Assessment (DORA) sottoscritta nel 2018, che mira all'evoluzione dei sistemi di valutazione dei risultati della Ricerca, la partecipazione a iniziative di coordinamento nazionale come ICDI (Italian Computing and Data Infrastructure) coordinato dal GARR, o la partecipazione diretta alla costruzione dell'EOSC, l'European Open Science Cloud.

Riferimenti Bibliografici

Meadows, A. (2018), Enter once, reuse often, *Revista ORL*, 9(4), 301-302. <https://doi.org/10.14201/orl.18547>

Nielsen M., (2011), *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press, 280 pp. <http://press.princeton.edu/titles/9517.html>

Wilkinson M. D., Dumontier M., Aalbersberg Ij. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L. B., Bourne P. E. et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3(1), <https://doi.org/10.1038/sdata.2016.18>

Autori

Mario Locati mario.locati@ingv.it

Tecnologo all'INGV dal 2002. Si occupa della gestione di banche dati di terremoti e con questi partecipa a progetti nazionali ed europei. Promuove i dati aperti poiché si occupa di prodotti frutto dell'integrazione di dati da diverse fonti. Nel 2013 permette l'attivazione del servizio DataCite per l'assegnazione di codici DOI ai dati istituzionali. "Master in Management of Research, Innovation and Technology" presso il Politecnico di Milano nel 2017. Dal 2018 coordina l'Ufficio Gestione Dati.

Francesco Mariano Mele franco.mele@ingv.it

Laureato in fisica, è stato ricercatore del ING-INGV per oltre 36 anni, responsabile del CED ING, dei sistemi di calcolo real-time per il Servizio di Sorveglianza Sismica e del Bollettino Sismico. Si è occupato di calcolo di parametri ipocentrali, tomografia su scala locale e regionale, analisi di sequenze sismiche. È stato membro del Comitato Direttivo del Centro Allerta Tsunami dell'INGV dal 2017 al 2021 e referente per la Sismologia dell'Ufficio Gestione Dati dal 2018 all'aprile 2021.

Placido Montalto placido.montalto@ingv.it

Ricercatore presso l'Osservatorio Etneo, è laureato in ingegneria informatica con PhD in ingegneria dei sistemi. Responsabile della Sala Operativa e dei servizi di Information Technology dell'Osservatorio Etneo, si occupa della progettazione e realizzazione di sistemi per l'elaborazione e l'archiviazione dei dati, della progettazione e realizzazione di software per il processamento dei dati geofisici. È co-autore di oltre 40 pubblicazioni in ambito tecnico e scientifico.

Vincenzo Romano vincenzo.romano@ingv.it

PhD presso "University of Nottingham", Ingegnere Elettronico, Primo Tecnologo presso l'INGV. Esperienza ventennale nelle aree: Fisica Ionosferica e radiopropagazione, Space weather, GNSS, Data Management e Trasferimento tecnologico. Ha partecipato ad oltre 20 progetti di Ricerca e Sviluppo, nazionali e internazionali, spesso con ruoli di responsabilità. È stato fondatore e Direttore Generale del primo spin-off INGV. È co-autore di oltre 50 articoli su riviste peer-review internazionali.

Valentino Lauciani valentino.lauciani@ingv.it

Tecnico informatico presso INGV dal 2000. Partecipa alla progettazione e allo sviluppo di molti dei software e siti web utilizzati nella Sala di Sorveglianza Sismica di Roma. Si occupa anche di Hardware e nello specifico della gestione del Cluster VMWare. Sviluppa progetti su Azure con l'uso dei container e negli ultimi anni si dedica ai backend, progettando e sviluppando Database ed API nel rispetto dei vari standard esistenti. Il bello di questo lavoro? Ogni giorno si scoprono cose nuove ;-)

Roberto Vallone roberto.vallone@ingv.it

Laurea in Scienze della Terra. Tecnologo, presso INGV dal 2017. Gestisce la piattaforma tecnologica (hardware e software) che sta alla base del database delle faglie sismogenetiche sia a livello di area mediterranea a fini scientifici e di protezione civile, sia a livello europeo in ambito EPOS ERIC per la sismologia (<https://www.seisnofaults.eu>) e la pericolosità da tsunami (<https://www.tsunamidata.org>). Dal 2019 fa parte dell'Ufficio Gestione Dati dell'Ente come referente informatico.

Stefano Cacciaguerra stefano.cacciaguerra@ingv.it

Laurea in Scienze dell'Informazione, Ph.D in Informatica e Master in "Management of Research, Innovation and Technology". Come Tecnologo INGV, si occupa del trasferimento tecnologico dell'innovazione nel campo dell'Informatica a favore della progettazione, realizzazione e gestione di nuove infrastrutture ICT per l'implementazione di applicazioni Geofisiche, Vulcanologiche e Oceanografiche per attuare a livello nazionale progetti europei come EOSC, EDI e EUROHPC. Responsabile del CED di Bologna.