

Il ruolo del dato aperto e della formazione nell'infrastruttura di ricerca GeoSciences-IR

Maria Pia Congi, Valentina Campo, Carlo Cipolloni, Paolo Lino Manganello

Istituto Superiore per la Protezione e la Ricerca Ambientale – Dipartimento per il Servizio Geologico d'Italia, Via Vitaliano Brancati 60, 00144 Roma
(mariapia.congi@isprambiente.it)

Abstract. La produzione e la diffusione dei dati aperti stanno crescendo in maniera esponenziale e, di conseguenza, il beneficio ed il valore delle risorse così distribuite viene riconosciuto da tutti i livelli di utenza. Su questo tema, le infrastrutture di ricerca hanno un ruolo chiave per l'assunzione del paradigma dell'Open Science.

In questo contesto, il progetto GeoSciences-IR, di cui ISPRA è capofila, mira a creare un'infrastruttura di ricerca (CRI) che consenta la condivisione e l'utilizzo dei dati forniti da 13 università e 3 istituti di ricerca, partners di progetto.

Il target principale del progetto è la Rete Italiana dei Servizi Geologici (RISG), che rappresenta la rete di coordinamento tra il Servizio Geologico d'Italia, parte dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), ed i Servizi Geologici Regionali.

Parole chiave: Progetto GeoSciences-IR, Direttiva INSPIRE, Principi FAIR

GeoSciences-IR

Nell'Infrastruttura di ricerca saranno gestiti diversi tipi di prodotti: dataset, metadati, servizi, vocabolari, pubblicazioni, software e applicazioni, e-learning. I dataset e i prodotti della formazione, per esempio, coprono diverse tematiche: mappatura e modellazione geologica e geotematica, frane, sinkholes, mitigazione del rischio idrogeologico, rilevamento satellitare ed in situ, faglie attive e capaci, attività mineraria sostenibile, uso e consumo del suolo, ecc.

Per tutti i prodotti che entreranno a far parte dell'IR GeoSciences è obbligatorio il rispetto dei principi FAIR (Findable, Accessible, Interoperable, Reusable).

Il Work Package 6 (WP6) ha il compito di raccogliere ed analizzare tutti i dataset prodotti dalle unità operative dei WP tematici (2-3-4-5) ed individua il processo di armonizzazione da seguire, insieme al WP8 che ne deve validare il rispetto dei principi FAIR. Per raggiungere questi obiettivi conduce diverse azioni: analisi del modello dei dati di partenza e del contenuto informativo, analisi del modello dei dati di destinazione, creazione di tabelle di mappatura. Il risultato è la progettazione del modello di dati finale (attraverso l'arricchimento semantico), la trasformazione del dataset, la validazione (validatore INSPIRE) e la condivisione. I metadati dei dataset e dei servizi seguono lo stesso flusso di validazione.

Per gestire i diversi tipi di dati geospaziali è stato deciso di utilizzare il formato GeoPackage, formato aperto e basato su standard sviluppato dall'Open Geospatial Consortium (OGC).

Una delle principali sfide all'interno del progetto è rappresentata dunque dall'armonizzazione dei dataset che verranno condivisi, nel rispetto dei modelli dati definiti dalla Direttiva INSPIRE, insieme all'armonizzazione semantica sulla base di vocabolari standard (ad esempio, INSPIRE, GeoSciML).

Il progetto GeoSciences-IR: uso di vocabolari e codelist

Una delle attività più rilevanti che viene svolta all'interno del progetto è quella relativa all'analisi dei vocabolari e/o codelist disponibili nell'ambito delle scienze della terra.

Per tutti i dataset è stata avviata l'analisi dei termini ivi contenuti ai fini dell'individuazione dell'esatta corrispondenza del significato attribuito.

Questo passaggio è di fondamentale importanza per fornire agli utenti dataset omogenei con lo stesso contenuto informativo minimo a qualsiasi livello.

Tutti i termini non attualmente disponibili in alcun vocabolario standard saranno descritti e successivamente proposti all'Agenzia per l'Italia Digitale (AgID), che gestisce il Registro INSPIRE Italia, ai fini della pubblicazione e condivisione a livello nazionale.

E-learning

Un'altra attività fondamentale propria dell'Infrastruttura di Ricerca riguarda l'implementazione di una piattaforma di e-learning integrata e flessibile. Considerando che l'Infrastruttura di Ricerca dovrà essere mantenuta e aggiornata per i 10 anni successivi alla fine del progetto, prevista a settembre 2025, si è deciso di dotarla di una propria "sezione web" per accedere sia ai contenuti prodotti per scopi prevalentemente divulgativi/promozionali sia a quelli didattici. Tale piattaforma di e-learning sarà sviluppata sul modello Moodle e sarà accessibile dall' HUB della Cloud Research Infrastructure (CRI) GeoSciences-IR.

Sarà reso possibile l'accesso a video tutorial, video lezioni, corsi asincroni e webinar.

Il target principale dell'attività saranno i Servizi Geologici Regionali, ma anche tutti gli utenti interessati ai temi delle scienze della terra.

Ringraziamenti

L'attività di ricerca descritta in questo lavoro è finanziata dall'Unione Europea – Programma NextGenerationEU - Missione 4 “Istruzione e Ricerca” - Componente 2 “Dalla ricerca all'impresa” – Linea di investimento 3.1 “Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione” - Progetto IR0000037 – GeoSciences-IR.

Riferimenti bibliografici

1. Battistoni P., Cipolloni C., Congi M. P., Sebillo M.M.: L'architettura dell'infrastruttura di ricerca GeoSciences IR. Atti ASITA 2024 (2024).
2. Cipolloni C., Congi M.P., Minelli A.: Un processo FAIR per la gestione dei dati e dei prodotti dell'infrastruttura GeoSciencesIR, GIT 2024 Proceedings on Proceedings (2024).
3. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al.: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.