

## L'utilizzo del formato GeoPackage nelle Infrastrutture di Ricerca dei progetti PNRR GeoSciences-IR e MEET

Paolo Lino Manganello, Giuseppe Castorina, Anna Maria Blumetti, Valentina Campo, Carlo Cipolloni, Deborah Maceroni, Maurizio Marino, Maria Pia Congi

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

**Abstract.** *This work mainly aims at illustrating the advantages of using GeoPackage format for the implementation of spatial databases within the Research Infrastructures of GeoSciencesIR and MEET projects. Considering that data which will become part of such Research Infrastructures have to be in accordance with FAIR principles and European Directive INSPIRE, the GeoPackage format possesses features that make its use optimal, with a view also to satisfying the Open Data paradigm. Indeed, GeoPackage is an open and standards-based format for transferring of geospatial data, and is characterized by versatile data support and large storage capacity. These features of it are very advantageous considering that a large amount of different kinds of data will have to be handled in the abovementioned Research Infrastructures. In this regard, we present two use cases of GeoPackage format for the implementation of geological databases. The first use case concerns the implementation of a Stratigraphic Database, which collects rock samples and prepared with their respective stratigraphic and paleontological analyses. The second use case relates the implementation of ITHACA Catalogue, which contains available information on the capable faults in the Italian territory and on the paleoseismological sites. The two presented use cases demonstrate how the GeoPackage format is strongly suited for the management and sharing of geospatial databases.*

**Riassunto.** Il seguente lavoro è principalmente finalizzato ad illustrare i vantaggi dell'utilizzo del formato *GeoPackage* per l'implementazione di database spaziali nelle Infrastrutture di Ricerca dei progetti *GeoSciences-IR* e *MEET*. Considerando che i dati che entreranno in tali Infrastrutture di Ricerca devono rispettare i principi *FAIR* e la Direttiva Europea *INSPIRE*, il formato *GeoPackage* possiede delle caratteristiche che rendono ottimale il suo utilizzo, nell'ottica anche di soddisfare il paradigma *Open Data*. Infatti il *GeoPackage* è un formato aperto e basato su standard per il trasferimento di dati geospaziali, ed è caratterizzato da un supporto versatile di dati e da una grande capacità di archiviazione. Queste sue caratteristiche risultano molto vantaggiose considerando che una grande quantità di dati di diversa natura dovrà essere gestita nelle Infrastrutture di Ricerca menzionate.

A tal proposito presentiamo due casi d'uso del formato *GeoPackage* per l'implementazione di database geologici. Il primo caso concerne l'implementazione di un Database Stratigrafico, il quale raccoglie campioni di roccia e preparati con le loro rispettive analisi stratigrafiche e paleontologiche. Il secondo caso d'uso riguarda l'implementazione del Catalogo ITHACA, il quale contiene le informazioni disponibili sulle faglie capaci presenti nel territorio italiano e sui siti paleosismologici. I due casi d'uso presentati dimostrano come il formato *GeoPackage* sia ottimale per la gestione e la condivisione di database geospaziali.

**Parole chiave:** *GeoPackage*, Infrastrutture di Ricerca, Principi FAIR, Direttiva INSPIRE

## 1. Introduzione

Il presente lavoro si inquadra nell'ambito dei progetti GeoSciences-IR e MEET, i quali sono entrambi finanziati dal Piano Nazionale di Ripresa e Resilienza (PNRR).

Il Progetto GeoSciences-IR ha come obiettivo quello di realizzare un'Infrastruttura di Ricerca per la Rete Italiana dei Servizi Geologici (RISG), la quale rappresenta la rete di coordinamento tra il Dipartimento per il Servizio Geologico d'Italia dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) ed i Servizi Geologici Regionali. L'obiettivo principale di questa Infrastruttura di Ricerca è dunque quello di rafforzare e sviluppare le competenze scientifiche dei Servizi Geologici Regionali relativamente a tematiche geologiche ritenute di prioritaria importanza, fornendo di conseguenza uno strumento utile per il monitoraggio del territorio [1].

Per quanto concerne invece il Progetto MEET (*Monitoring Earth's Evolution and Tectonics*), esso è coordinato dall'Istituto Nazionale di Geofisica e Vulcanologia (INGV) ed ha come scopo quello di rinnovare, implementare e creare nodi di monitoraggio della dinamica terrestre per la costruzione di reti di infrastrutture di ricerca [2].

I prodotti che entreranno a far parte delle Infrastrutture di Ricerca dei progetti GeoSciences-IR e MEET dovranno rispettare i principi FAIR (*Findability, Accessibility, Interoperability, Reuse*) e la Direttiva INSPIRE (*Infrastructure for SPatial InfoRmation in Europe*) 2007/2/CE.

I principi FAIR sono un insieme di 15 principi che si applicano a tre diverse entità: dati, metadati ed infrastrutture. Questi principi hanno ad esempio tra gli altri l'obiettivo di ottimizzare la riutilizzabilità dei dati della ricerca, aumentando l'impatto e l'efficienza della ricerca stessa [3]. A livello internazionale sono stati elaborati diversi strumenti in grado di valutare il livello di rispetto di questi principi. Tra questi a titolo di esempio ricordiamo l'interessante lavoro di Jones & Grootveld [4], che hanno prodotto una semplice *checklist* che consente un'autovalutazione del livello di *FAIRness*.

La Direttiva Europea INSPIRE, entrata in vigore il 15 maggio 2007 e recepita nell'ordinamento italiano con il D.Lgs. 32/2010, ha l'obiettivo di garantire che le infrastrutture per l'informazione territoriale degli Stati membri della Comunità Europea siano compatibili ed utilizzabili in un contesto comunitario e transfrontaliero [5]. Questa

Direttiva definisce 34 categorie tematiche di dati territoriali relative alle applicazioni ambientali e richiede la standardizzazione di ognuna di queste tematiche. Tra queste tematiche sono presenti ad esempio i siti protetti, la copertura del suolo, la geologia, l'utilizzo del territorio, gli impianti di monitoraggio ambientale, le zone a rischio naturale e le risorse minerarie.

Nell'ottica di soddisfare i principi FAIR e la Direttiva INSPIRE, si è deciso di utilizzare il formato *GeoPackage* per la gestione dei dati geospaziali che entreranno a far parte delle Infrastrutture di Ricerca dei progetti GeoSciences-IR e MEET [6].

## 2. Formato *GeoPackage*

Il formato *GeoPackage* possiede delle caratteristiche che lo rendono ottimale ad essere utilizzato per la gestione di dati geospaziali nelle Infrastrutture di Ricerca, nell'ottica anche di soddisfare il paradigma *Open Data*.

Il *GeoPackage* è infatti un formato aperto e basato su standard, sviluppato dall'*Open Geospatial Consortium* (OGC), per la condivisione ed il trasferimento di informazioni geospaziali [7].

Questo formato è caratterizzato da un supporto versatile di dati, che consente di archiviare efficacemente in un database SQLite differenti tipologie di dati, sia spaziali (dati vettoriali, dati raster) che non-spaziali (attributi). Un'altra fondamentale caratteristica di questo formato è la sua grande capacità di archiviazione, la quale consente di gestire in maniera ottimale dataset molto estesi. Considerando che nelle Infrastrutture di Ricerca menzionate dovrà essere gestita una grande quantità di dati di diversa natura, l'utilizzo del formato *GeoPackage* risulta dunque particolarmente utile a causa di queste sue caratteristiche vantaggiose.

Il *GeoPackage* supporta inoltre l'uso diretto, il che si esplicita nel fatto che in esso l'accesso e l'aggiornamento dei dati avviene in un formato di archiviazione "nativo", senza la necessità di ricorrere a conversioni intermedie di formato.

Altre caratteristiche importanti di questo formato sono il fatto che esso può essere utilizzato in tutte le piattaforme, l'interoperabilità e la portabilità. Infine, un'altra sua caratteristica vantaggiosa risiede nel fatto che esso supera la limitazione, propria degli *shapefile*, di 10 caratteri per i nomi dei campi delle tabelle e di 255 caratteri per gli attributi testuali.

## 3. Casi d'uso

In questa sezione presentiamo due casi d'uso del *GeoPackage encoding* per l'implementazione di database geologici, entrambi di proprietà del Dipartimento per il Servizio Geologico d'Italia (ISPRA).

Il primo caso d'uso riguarda l'implementazione di un nuovo Database Stratigrafico, il quale raccoglierà campioni di roccia e preparati con le loro rispettive analisi stratigrafiche e paleontologiche.

Il secondo caso d'uso concerne invece l'implementazione del già esistente Catalogo ITHACA, il quale contiene le informazioni disponibili sulle faglie capaci presenti in territorio italiano e sui siti paleosismologici.

### 3.1 Database Stratigrafico

Il Database Stratigrafico ISPRA sarà costituito da tre layer geometrici:

- *Sample\_Registry* (layer puntuale), il quale contiene tutti i campioni catalogati con il loro posizionamento puntuale e le informazioni principali;
- *Areal\_section* (layer poligonale), contenente l'insieme degli areali che raggruppano collezioni di campioni;
- *Stratigraphic\_section* (layer lineare), contenente le sezioni stratigrafiche nelle quali sono stati prelevati campioni.

**Tabella 1.** Domini implementati per facilitare la compilazione del Database Stratigrafico.

<b>Dominio</b>	<b>Campo</b>	<b>Tabella</b>
<i>D_Country</i>	<i>Country</i>	<i>Sample_Registry</i>
<i>D_Region</i>	<i>Region</i>	<i>Sample_Registry</i>
<i>D_Province</i>	<i>Province</i>	<i>Sample_Registry</i>
<i>D_Municipality</i>	<i>Municipality</i>	<i>Sample_Registry</i>
<i>D_Aim_of_sampling</i>	<i>Aim_of_sampling</i>	<i>Sample_Registry</i>
<i>D_Type_of_prepared</i>	<i>Type_of_prepared</i>	<i>Associated_Registry</i>
<i>D_Preservation_of_the_prepared</i>	<i>Preservation</i>	<i>Associated_Registry</i>
<i>D_Era</i>	<i>Era_from, Era_to</i>	<i>Dating_Sample,</i> <i>Dating_Associated</i>
<i>D_Period</i>	<i>Period_from, Period_to</i>	<i>Dating_Sample,</i> <i>Dating_Associated</i>
<i>D_Epoch</i>	<i>Epoch_from, Epoch_to</i>	<i>Dating_Sample,</i> <i>Dating_Associated</i>
<i>D_Stage</i>	<i>Stage_from, Stage_to</i>	<i>Dating_Sample,</i> <i>Dating_Associated</i>
<i>D_Pro_partis</i>	<i>Pro_partis_from, Pro_partis_to</i>	<i>Dating_Sample,</i> <i>Dating_Associated</i>
<i>D_Lithology</i>	<i>Lithology_1, Lithology_2, Lithology_3</i>	<i>Lithostratigraphy_Sample,</i> <i>Lithology_Associated</i>
<i>D_Depositional_setting</i>	<i>Depositional_setting</i>	<i>Lithostratigraphy_Sample</i> <i>Lithology_Associated</i>
<i>D_Texture</i>	<i>Texture</i>	<i>Sedimentology_Sample,</i> <i>Sedimentology_Associated</i>
<i>D_Granulometry</i>	<i>Granulometry</i>	<i>Sedimentology_Sample,</i> <i>Sedimentology_Associated</i>
<i>D_Secondary_Granulometry</i>	<i>Secondary_Granulometry</i>	<i>Sedimentology_Sample,</i> <i>Sedimentology_Associated</i>
<i>D_Sedimentary_structure</i>	<i>Sedimentary_structure</i>	<i>Sedimentology_Sample,</i> <i>Sedimentology_Associated</i>
<i>D_Component</i>	<i>Components_1, Components_2,</i> <i>Components_3, Components_4</i>	<i>Sedimentology_Sample,</i> <i>Sedimentology_Associated</i>

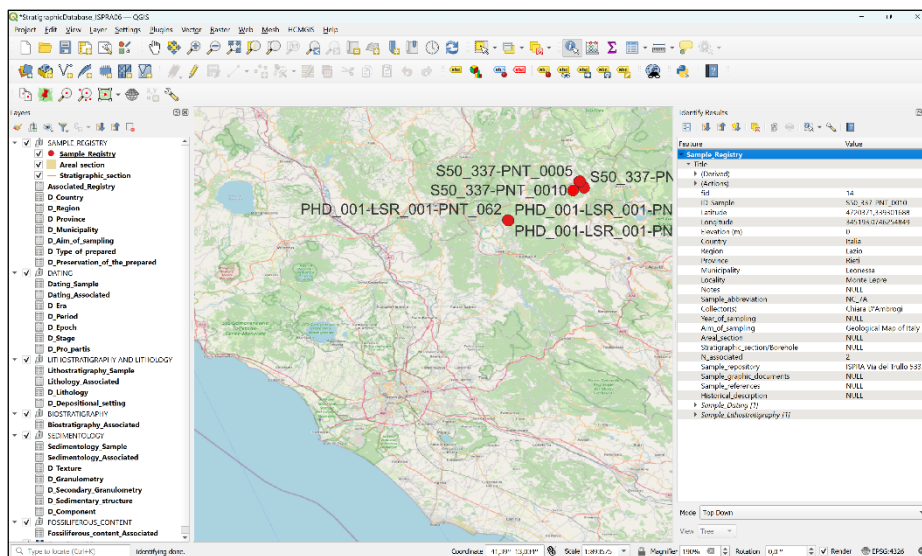
Accanto alla *Sample\_Registry*, l'altra tabella degli attributi principale è la *Associated\_Registry*, contenente le informazioni principali di tutti i preparati associati ad ogni campione.

La struttura del Database Stratigrafico è completata da alcune tabelle degli attributi collegate da relazioni, tramite l'*id-sample* e l'*id-associated*, alle due tabelle principali *Sample\_Registry* e *Associated\_Registry*. Le tabelle sono le seguenti:

- *Dating* (presente sia per i campioni che per gli associati), contenente tutte le informazioni sulla datazione;
- *Lithostratigraphy* (presente sia per i campioni che per gli associati), contenente le informazioni litostratigrafiche e litologiche;
- *Biostratigraphy* (presente solo per gli associati), contenente le informazioni biostratigrafiche;
- *Sedimentology* (presente sia per i campioni che per gli associati), raccoglie i risultati delle analisi sedimentologiche;
- *Fossiliferous\_Content* (presente solo per gli associati), contenente informazioni sul contenuto fossilifero.

Per facilitare la compilazione delle diverse tabelle degli attributi che costituiscono il database, sono stati creati dei domini, consistenti in delle liste precompilate di valori, tra i quali il compilatore potrà scegliere il valore del campo in questione. I domini implementati sono elencati nella Tabella 1.

Nella Figura 1 è riportato uno *screenshot* del *GeoPackage* implementato per il Database Stratigrafico.



**Figura 1:** *GeoPackage* del Database Stratigrafico ISPRA. Nella *Table Of Contents* (TOC) sono visibili tutti i layer che costituiscono il database.

### 3.2 Catalogo ITHACA

Il Catalogo ITHACA (*ITaly HAZard from Capable faults*) iniziò ad essere implementato negli anni '90. Esso rappresenta un importante strumento per analisi di pericolosità ambientale e sismica, pianificazione territoriale, comprensione dell'evoluzione recente del paesaggio, gestione delle emergenze di Protezione Civile e ricerca scientifica nel campo della geodinamica [8, 9].

Il Catalogo ITHACA è costituito da due layer geometrici:

- *Faults* (layer lineare), contenente le tracce delle faglie capaci e le informazioni disponibili;
- *Paleosites* (layer puntuale), contenente il posizionamento dei siti paleosismologici e le loro informazioni derivanti dall'analisi paleosismologica.

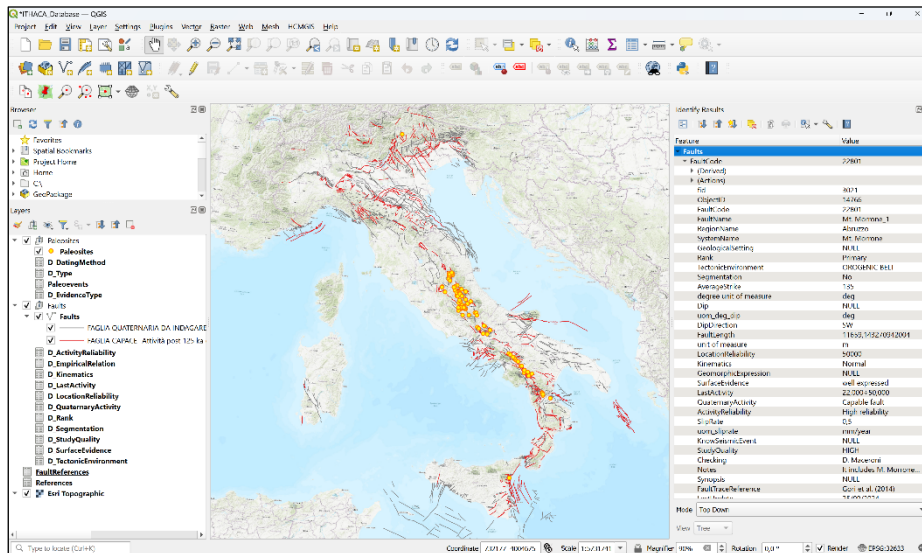
La struttura di questo database è arricchita da un'altra tabella degli attributi, la tabella *Paleoevents*, che fornisce una descrizione delle evidenze paleosismologiche.

Anche in questo caso d'uso, allo scopo di facilitare la compilazione di alcuni campi delle diverse tabelle degli attributi che costituiscono il database, sono stati creati dei domini. I domini implementati sono elencati nella Tabella 2.

**Tabella 2.** Domini implementati per la compilazione del Catalogo ITHACA.

<b>Dominio</b>	<b>Campo</b>	<b>Tabella</b>
<i>D Rank</i>	<i>Rank</i>	<i>Faults</i>
<i>D TectonicEnvironment</i>	<i>TectonicEnvironment</i>	<i>Faults</i>
<i>D Segmentation</i>	<i>Segmentation</i>	<i>Faults</i>
<i>D LocationReliability</i>	<i>LocationReliability</i>	<i>Faults</i>
<i>D Kinematics</i>	<i>Kinematics</i>	<i>Faults</i>
<i>D SurfaceEvidence</i>	<i>SurfaceEvidence</i>	<i>Faults</i>
<i>D LastActivity</i>	<i>LastActivity</i>	<i>Faults</i>
<i>D QuaternaryActivity</i>	<i>QuaternaryActivity</i>	<i>Faults</i>
<i>D ActivityReliability</i>	<i>ActivityReliability</i>	<i>Faults</i>
<i>D StudyQuality</i>	<i>StudyQuality</i>	<i>Faults</i>
<i>D Type</i>	<i>type</i>	<i>Paleosites</i>
<i>D EvidenceType</i>	<i>EvidenceType</i>	<i>Paleoevents</i>
<i>D DatingMethod</i>	<i>DatingMethod</i>	<i>Paleoevents</i>
<i>D EmpiricalRelation</i>	<i>EmpiricalRelation</i>	<i>Paleoevents</i>

In Figura 2 è riportato uno *screenshot* del *GeoPackage* implementato per il Catalogo ITHACA. Nel layer *Faults* sono state distinte le faglie capaci (in rosso), le quali sono state attive negli ultimi 125 ka, dalle faglie quaternarie (in grigio), la cui ultima attività accertata è risalente a più di 125 ka.



**Figura 2:** *GeoPackage* del Catalogo ITHACA. Nella *Table Of Contents* (TOC) sono visibili tutti i layer che costituiscono il database.

#### 4. Conclusioni

Nell'ambito dei progetti PNRR GeoSciences-IR e MEET si è scelto di utilizzare il formato *GeoPackage* per gestire tutti i dati geospaziali che confluiranno nelle rispettive Infrastrutture di Ricerca.

Il *GeoPackage* è un formato aperto e basato su standard, sviluppato per la condivisione ed il trasferimento di informazioni geospaziali. Questo formato è caratterizzato da un supporto versatile di dati, che consente di archiviare in maniera efficace differenti tipologie di dati, e da una grande capacità di archiviazione, che consente di gestire in maniera ottimale dataset molto estesi. Il *GeoPackage* supporta inoltre l'uso diretto, consentendo dunque l'accesso e l'aggiornamento dei dati in un formato di archiviazione "nativo", senza la necessità di ricorrere a conversioni intermedie di formato. Altre caratteristiche vantaggiose di questo formato sono il fatto che esso può essere utilizzato in tutte le piattaforme, l'interoperabilità, la portabilità, ed il fatto che esso supera la limitazione di 10 caratteri per i nomi dei campi delle tabelle e di 255 caratteri per gli attributi testuali.

I due casi d'uso descritti in questo lavoro riguardano il Database Stratigrafico ed il Catalogo ITHACA, entrambi di proprietà del Dipartimento per il Servizio Geologico d'Italia (ISPRA).

Questi due casi d'uso mettono in evidenza come il formato *GeoPackage*, a causa delle sue caratteristiche vantaggiose, sia fortemente adatto ad essere utilizzato per la gestione e la condivisione di dati geospaziali.

## 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 e Progetto IR0000025 – MEET.

## Riferimenti bibliografici

1. Sito web Progetto GeoSciencesIR, <https://geosciences-ir.it>, ultimo accesso 2024/09/18.
2. Sito web Progetto MEET, <https://meet.ingv.it>, ultimo accesso 2024/09/18.
3. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.
4. Jones, S., Grootveld, M.: How FAIR are your data? EUDAT (2017). [10.5281/zenodo.1065990](https://zenodo.org/record/1065990).
5. Direttiva 2007/2/CE del Parlamento Europeo e del Consiglio del 14 marzo 2007, Gazzetta ufficiale dell'Unione europea (2007).
6. Congi, M. P., Campo, V., Cipolloni, C., Funaro, M., Manganello, P. L.: The Research Infrastructure of GeoSciences: from datasets analysis to harmonization process. XVIII Convegno Nazionale della Sezione “GIT – Geosciences and Information Technologies” della Società Geologica Italiana, Malnisio, Montereale Valcellina (PN), 17-19 giugno 2024 (2024).
7. Sito web GeoPackage, <https://www.geopackage.org>, ultimo accesso 2024/10/11.
8. Pagina web Progetto ITHACA, <https://www.isprambiente.gov.it/it/progetti/cartella-progetti-in-corso/suolo-e-territorio-1/ithaca-catalogo-delle-faglie-capaci/default>, ultimo accesso 2024/10/15.
9. Maceroni, D., Blumetti, A. M., Bonadeo, L., Congi, M. P., Comerci, V., Di Manna, P., Guerrieri, L.: The implementation of the paleoseismological sites database into ITHACA Catalogue in the framework of the GeoSciencesIR project. GNGTS 2024, Sessione 1.1, Ferrara, 13-16 febbraio 2024 (2024).