Introduzione ai tipi e ai formati dei dati ambientali

Risultati di apprendimento

Al termine di questa unità, gli studenti saranno in grado di:

- Definire le principali tipologie di dati ambientali e comprenderne l'importanza negli studi ambientali.
- Distinguere tra dati ambientali strutturati, semistrutturati e non strutturati.
- Individuare i formati di dati più comuni nelle scienze ambientali (CSV, NetCDF, HDF5, GeoJSON, shapefile, ecc.).
- Valutare vantaggi e limiti dei diversi formati per l'archiviazione, la condivisione e l'analisi.
- Convertire e gestire file di dati ambientali con strumenti e librerie appropriati (ad esempio Python e QGIS).

Introduzione

I dati ambientali svolgono un ruolo fondamentale nella comprensione, nel monitoraggio e nella gestione dei sistemi naturali e di quelli influenzati dalle attività umane. Forniscono la base empirica per studiare i cambiamenti climatici, la qualità dell'aria e dell'acqua, la biodiversità, l'uso del suolo e molti altri processi ambientali. I progressi nelle tecnologie di rilevamento, nei sistemi di osservazione satellitare e negli strumenti computazionali hanno determinato una crescita senza precedenti del volume, della varietà e della complessità dei dati ambientali. Di conseguenza, professionisti e studenti delle scienze ambientali devono saper analizzare i dati e comprendere i diversi tipi e formati in cui vengono archiviati e scambiati.

I dati ambientali provengono da un'ampia gamma di fonti. Le misurazioni in situ sono raccolte mediante strumenti a terra, come stazioni meteorologiche, sensori della qualità dell'aria, idrometri e aree di monitoraggio ecologico. Le piattaforme di telerilevamento, tra cui satelliti, droni e aeromobili, forniscono osservazioni spaziali estese della superficie terrestre e dell'atmosfera. Inoltre, i dati ambientali possono essere generati da modelli numerici e simulazioni, come i modelli climatici o idrologici, che producono grandi insiemi di dati multidimensionali. Ciascuna fonte produce dati con strutture, scale e requisiti tecnici differenti, che ne influenzano la conservazione e l'utilizzo.

Comprendere i tipi e i formati dei dati ambientali è essenziale per gestirli e analizzarli efficacemente. I tipi di dati descrivono come sono organizzate le informazioni, ad esempio in dati

tabellari strutturati, record semistrutturati o file non strutturati come immagini e testi. I formati, invece, definiscono come i dati sono codificati in file quali CSV, Parquet, NetCDF, HDF5, GeoTIFF o GeoJSON. La scelta del formato influisce sull'accessibilità dei dati, sull'efficienza dell'archiviazione, sull'interoperabilità tra strumenti software e sulla possibilità di includere metadati che descrivono origine, unità di misura, riferimento spaziale e copertura temporale.

Nelle scienze ambientali, una gestione inadeguata dei formati può causare interpretazioni errate, perdita di dati o minore riproducibilità dei risultati. Ad esempio, ignorare i sistemi di riferimento delle coordinate nei dati spaziali può produrre sovrapposizioni cartografiche scorrette, mentre l'uso di formati eccessivamente semplici può comportare la perdita di metadati importanti. Al contrario, la scelta di formati adeguati consente una condivisione efficiente, la conservazione a lungo termine e l'integrazione di dataset eterogenei provenienti da più fonti. Ciò è particolarmente importante nella ricerca ambientale interdisciplinare, in cui occorre combinare e confrontare dati di diversi ambiti.

Questa lezione presenta i principali tipi di dati ambientali e i formati più utilizzati nelle scienze ambientali e nelle applicazioni geospaziali. Offre un quadro concettuale per comprendere come i dati ambientali siano strutturati e archiviati, con esempi pratici relativi al monitoraggio della qualità dell'aria, alle osservazioni climatiche e all'analisi dell'uso del suolo. La familiarità con questi concetti consentirà agli studenti di lavorare meglio con dataset ambientali reali, scegliere formati adeguati all'analisi e garantire processi di lavoro trasparenti, riproducibili e scientificamente fondati.

Tipi di dati ambientali

I dati ambientali possono essere classificati in diversi modi, a seconda della struttura, della dimensionalità, delle caratteristiche spaziali e temporali e del metodo di raccolta. Una chiara comprensione delle tipologie di dati è essenziale per scegliere formati di archiviazione, tecniche di analisi e metodi di visualizzazione adeguati. In questa sezione i dati ambientali sono suddivisi in tre grandi categorie in base alla loro struttura: strutturati, semistrutturati e non strutturati. Particolare attenzione è inoltre dedicata alle dimensioni spaziale e temporale, particolarmente importanti nelle applicazioni ambientali.

Dati ambientali strutturati

I dati strutturati sono fortemente organizzati e generalmente conservati in forma tabellare, con righe e colonne caratterizzate da attributi e tipi di dati chiaramente definiti. Ogni riga rappresenta di norma un'osservazione, mentre ogni colonna corrisponde a una variabile o a un attributo. Questa tipologia è comune nei programmi di monitoraggio ambientale e viene spesso raccolta mediante procedure di misurazione standardizzate.

Tra gli esempi di dati ambientali strutturati figurano serie temporali della temperatura dell'aria, misure delle precipitazioni, concentrazioni di inquinanti atmosferici, indicatori della qualità dell'acqua e proprietà del suolo misurate in luoghi fissi. Questi dataset sono spesso conservati in file CSV, fogli di calcolo o basi di dati relazionali. I dati strutturati sono relativamente semplici da interrogare, filtrare e analizzare mediante software statistici, database e linguaggi di programmazione come R e Python.

CSV è un semplice formato testuale organizzato per righe, facile da leggere e scambiare, ma poco efficiente per analisi su larga scala e archiviazione a lungo termine. Apache Parquet è invece un formato colonnare ottimizzato per carichi di lavoro analitici: consente interrogazioni più rapide su variabili selezionate, una compressione efficiente e la conservazione dei tipi di dati. Pertanto, Parquet è più adatto a grandi dataset ambientali e flussi di lavoro intensivi, mentre CSV rimane utile per piccoli dataset e per lo scambio di dati.

Uno dei principali vantaggi dei dati strutturati è la semplicità e la compatibilità con numerosi strumenti. Consentono aggregazioni efficienti, analisi statistiche e visualizzazioni, ad esempio il calcolo di medie, tendenze o correlazioni nel tempo. Presentano tuttavia anche dei limiti: sono meno adatti a rappresentare strutture spaziali complesse, variabili multidimensionali o grandi dataset raster. Inoltre, informazioni contestuali importanti possono andare perdute se i metadati non vengono documentati accuratamente e conservati separatamente.

Dati ambientali semistrutturati

I dati semistrutturati non seguono strutture tabellari rigide, ma contengono comunque elementi organizzativi come tag, chiavi o relazioni gerarchiche. Offrono maggiore flessibilità rispetto ai dati strutturati, mantenendo al tempo stesso un'organizzazione che ne consente l'elaborazione automatica.

Nelle scienze ambientali i dati semistrutturati sono comunemente usati per rappresentare metadati, risultati dei sensori ed elementi spaziali. Formati come XML e JSON sono ampiamente

utilizzati per lo scambio di dati tra sistemi, soprattutto nei servizi web e nelle interfacce di programmazione delle applicazioni (API). GeoJSON, un'estensione specializzata di JSON, viene spesso utilizzato per rappresentare elementi geografici insieme ai loro attributi e alla geometria.

I dati semistrutturati sono particolarmente utili con dataset eterogenei, la cui struttura varia o ai quali si aggiungono attributi nel tempo. Ad esempio, le reti di sensori ambientali producono spesso dati semistrutturati che includono misurazioni, marcature temporali, identificativi dei sensori e indicatori di qualità. Questa flessibilità facilita l'integrazione tra piattaforme, ma può richiedere operazioni di parsing e validazione più complesse rispetto ai dati tabellari strutturati.

Dati ambientali non strutturati

I dati non strutturati non hanno un modello predefinito e non sono facilmente rappresentabili mediante tabelle o strutture gerarchiche. Comprendono immagini, video, registrazioni audio, relazioni in testo libero e immagini grezze di telerilevamento. In ambito ambientale acquistano crescente importanza grazie ai progressi nelle tecnologie di osservazione della Terra e nei sistemi automatizzati di raccolta dei dati.

Ne sono esempi le immagini satellitari, le fotografie aeree, i filmati ripresi dai droni, gli appunti di campo digitalizzati e le valutazioni di impatto ambientale in forma testuale. Questi dati richiedono spesso tecniche specializzate, come l'analisi delle immagini, l'elaborazione del linguaggio naturale o l'apprendimento automatico. Pur contenendo informazioni ricche e dettagliate, risultano più difficili da analizzare e archiviare in modo efficiente.

L'analisi dei dati ambientali non strutturati richiede spesso ulteriori livelli di interpretazione o trasformazione. Ad esempio, le immagini satellitari possono essere elaborate per ricavare indici di vegetazione o classificazioni di copertura del suolo, trasformando dati raster non strutturati in dataset strutturati o semistrutturati adatti a successive analisi. I dati non strutturati costituiscono quindi spesso l'input iniziale per generare informazioni ambientali più strutturate.

Dati ambientali spaziali

I dati spaziali rappresentano informazioni esplicitamente associate a una posizione geografica. La posizione può essere definita tramite coordinate, confini amministrativi, griglie o sistemi di riferimento spaziale. Questi dati sono fondamentali nelle scienze ambientali, poiché la maggior parte dei fenomeni ambientali è intrinsecamente legata allo spazio.

I dati ambientali spaziali si distinguono generalmente in vettoriali e raster. I dati vettoriali rappresentano elementi discreti, come stazioni di monitoraggio, fiumi, strade o appezzamenti,

mediante punti, linee e poligoni. I dati raster rappresentano fenomeni continui, come campi di temperatura, modelli altimetrici o immagini satellitari, mediante una griglia di celle o pixel.

Comprendere i tipi di dati spaziali è fondamentale per l'analisi ambientale e la cartografia. Aspetti come i sistemi di riferimento delle coordinate, la risoluzione spaziale e la scala possono influire notevolmente sui risultati. Una gestione scorretta può produrre livelli disallineati, misurazioni imprecise o interpretazioni errate. I dati spaziali richiedono pertanto una gestione attenta e strumenti specializzati, come i Sistemi Informativi Geografici (GIS).

Dati ambientali temporali

Le caratteristiche temporali costituiscono un altro aspetto distintivo dei dati ambientali. Molti processi cambiano nel tempo, dalle fluttuazioni meteorologiche di breve periodo alle tendenze climatiche o ai cambiamenti degli ecosistemi su periodi lunghi. I dati temporali possono essere raccolti a intervalli regolari o irregolari oppure mediante misure continue.

Le serie temporali sono una forma comune di dati ambientali temporali. Ne sono esempi le misure orarie della qualità dell'aria, le portate giornaliere dei fiumi e gli indici climatici mensili. La risoluzione temporale, le convenzioni sui fusi orari e la presenza di lacune nei dati sono aspetti importanti. Un corretto allineamento temporale è indispensabile per combinare dataset di fonti differenti.

Spesso i dati ambientali presentano contemporaneamente una componente spaziale e una temporale, dando origine a dataset spazio-temporali. Possono essere complessi e avere un'elevata dimensionalità, come gli output dei modelli climatici globali o le serie temporali ricavate dai satelliti. La loro gestione richiede strutture e formati specializzati, capaci di conservare e rendere accessibili grandi quantità di informazioni in modo efficiente.

Dati multidimensionali e su griglia

I dati multidimensionali rappresentano variabili ambientali che cambiano lungo più dimensioni, come latitudine, longitudine, altitudine e tempo. Sono comuni nelle scienze atmosferiche e oceanografiche, in cui le misure riguardano uno spazio tridimensionale e intervalli temporali lunghi.

I dati su griglia sono un particolare tipo di dati multidimensionali, con valori definiti su una griglia regolare. Esempi comprendono i dataset di rianalisi climatica, gli output delle previsioni meteorologiche numeriche e i modelli di circolazione oceanica. Spesso contengono molte variabili e metadati estesi, risultando molto più complessi dei semplici dati tabellari.

La complessità dei dati ambientali multidimensionali rende necessari formati e strumenti software specializzati, in grado di gestire grandi array e conservare i metadati. Comprendere la natura di questi dati è indispensabile prima di scegliere formati e procedure di analisi appropriati.

Formati comuni dei dati ambientali

I formati dei dati ambientali definiscono le modalità con cui i dati vengono codificati, archiviati e scambiati tra sistemi. La scelta del formato incide notevolmente su accessibilità, efficienza di archiviazione, interoperabilità e utilizzabilità a lungo termine. Nelle scienze ambientali, i formati devono spesso supportare grandi volumi di dati, dimensioni spaziali e temporali e metadati dettagliati. Questa sezione presenta i formati più comuni nelle applicazioni ambientali e geospaziali, evidenziandone caratteristiche, vantaggi e limiti.

Formati di dati tabellari

I formati tabellari sono tra i più diffusi nelle scienze ambientali, grazie alla semplicità e all'ampia compatibilità. Memorizzano le informazioni in righe e colonne e sono comunemente usati per dataset ambientali strutturati, come serie temporali di misurazioni e statistiche riepilogative.

Il formato tabellare più comune è CSV (Comma-Separated Values). I file CSV sono file di testo semplice in cui ogni riga rappresenta un record e i valori sono separati da virgole o altri delimitatori. Sono facili da creare, leggere e scambiare tra diverse piattaforme software. Sono molto usati per condividere misure della qualità dell'aria, osservazioni meteorologiche e dati sulla qualità dell'acqua. Tuttavia, non offrono un supporto standardizzato per metadati, tipi di dati e riferimenti spaziali, con il rischio di ambiguità o perdita di informazioni in assenza di una documentazione adeguata.

Anche i formati dei fogli di calcolo, come i file Microsoft Excel, sono utilizzati di frequente, soprattutto nelle prime fasi di raccolta o con dataset piccoli. Pur offrendo interfacce intuitive e funzioni di base per la validazione, sono meno adatti a grandi dataset, elaborazioni automatizzate e processi riproducibili. Formattazioni nascoste, fogli multipli e modifiche manuali possono introdurre errori e ridurre la trasparenza.

Formati colonnari e analitici

Apache Parquet è un formato open source di archiviazione colonnare, ottimizzato per l'analisi e l'elaborazione dei dati su larga scala. A differenza dei tradizionali formati tabellari per righe,

come CSV, Parquet memorizza i dati colonna per colonna, migliorando notevolmente le prestazioni delle interrogazioni che accedono soltanto a un sottoinsieme di variabili. È quindi particolarmente adatto a grandi dataset ambientali con numerosi attributi e osservazioni.

Nelle scienze ambientali Parquet è sempre più utilizzato per conservare grandi quantità di dati tabellari, come misurazioni pluriennali della qualità dell'aria, output di modelli climatici convertiti in tabelle e dati aggregati dei sensori. Supporta compressione e codifica efficienti, riducendo lo spazio necessario e mantenendo letture rapide. Inoltre conserva i tipi di dati e le informazioni sullo schema, superando molti limiti dei formati testuali semplici.

Sebbene Parquet non sia un formato GIS nativo, svolge un ruolo importante nei processi di gestione dei dati ambientali elaborati su piattaforme cloud, data warehouse o ambienti di calcolo distribuito, prima della visualizzazione o dell'analisi spaziale in software GIS. I dati in formato Parquet possono essere convertiti in formati spaziali o collegati a strumenti GIS tramite elaborazioni intermedie con Python, motori basati su SQL o librerie per dati spaziali.

Formati scientifici e multidimensionali

I formati scientifici sono progettati per archiviare in modo efficiente grandi dataset multidimensionali e includere metadati ricchi, che descrivono struttura, unità di misura e sistemi di coordinate. Sono particolarmente importanti per climatologia, meteorologia, oceanografia e ricerca atmosferica.

NetCDF (Network Common Data Form) è uno dei formati scientifici più diffusi nelle scienze ambientali. Supporta array multidimensionali e metadati autodescrittivi ed è quindi adatto a dati spazio-temporali, come output di modelli climatici, dataset di rianalisi e variabili derivate da satellite. Un singolo file NetCDF può contenere più variabili e dimensioni, consentendo accesso e selezione efficienti dei dati.

HDF5 (Hierarchical Data Format, versione 5) è un altro formato potente per l'archiviazione di dataset ambientali complessi. Offre una struttura gerarchica che organizza dati in gruppi e dataset, simile a un file system all'interno di un file. È altamente scalabile ed efficiente, quindi adatto a dataset molto grandi. La sua complessità può tuttavia rappresentare una difficoltà per i principianti e spesso sono necessarie librerie specializzate per leggere e scrivere file HDF5.

Formati di dati vettoriali spaziali

I formati vettoriali vengono utilizzati per rappresentare elementi geografici discreti, come punti, linee e poligoni. Sono fondamentali per mappare stazioni di monitoraggio ambientale, confini amministrativi, fiumi, strade e caratteristiche dell'uso del suolo.

ESRI Shapefile è da molti anni un formato vettoriale standard nei GIS. Uno shapefile è composto da diversi file associati, che nel loro insieme conservano geometria, attributi e informazioni sul riferimento spaziale. Sebbene ampiamente supportato, presenta diversi limiti, tra cui restrizioni sulla dimensione dei file, sulla lunghezza dei nomi degli attributi e sui tipi di dati.

GeoJSON è un formato moderno, aperto e testuale per codificare elementi geografici mediante JSON. È particolarmente diffuso nella cartografia web e nello scambio di dati grazie alla semplicità e leggibilità. Consente di conservare attributi e geometrie in un unico file e si integra bene con le tecnologie web. Tuttavia, i file GeoJSON possono diventare voluminosi e poco efficienti per dataset complessi o estesi.

Formati di dati raster spaziali

I formati raster rappresentano fenomeni ambientali continui, memorizzando i valori in una griglia di celle o pixel. Esempi comuni sono immagini satellitari, modelli digitali di elevazione, mappe della temperatura superficiale e indici di vegetazione.

GeoTIFF è un formato raster molto utilizzato, che estende il TIFF standard includendo metadati geografici, come sistemi di riferimento delle coordinate e risoluzione spaziale. È comune nelle applicazioni di telerilevamento e GIS ed è supportato dalla maggior parte dei software geospaziali.

Anche altri formati raster, come NetCDF e HDF5, sono utilizzati frequentemente per dati ambientali su griglia, soprattutto quando sono presenti dimensioni temporali e verticali. La scelta dipende spesso dalla complessità dei dati, dalla dimensione dei file e dai requisiti dell'analisi.

Metadati e standard di descrizione dei dati

I metadati sono indispensabili per comprendere e utilizzare correttamente i dati ambientali. Descrivono origine, contenuto, struttura e qualità dei dati, oltre alla copertura spaziale e temporale. Senza metadati adeguati, i dataset ambientali possono essere difficili o impossibili da interpretare correttamente.

Nelle applicazioni ambientali e geospaziali si usano diversi standard per i metadati. ISO 19115 è uno standard internazionale per i metadati geografici che offre un quadro completo per descrivere i dataset spaziali. Dublin Core è uno standard più semplice e generale, utilizzato in diversi ambiti. Molti formati scientifici, come NetCDF, supportano metadati incorporati conformi a convenzioni consolidate.

Un corretto utilizzo degli standard dei metadati migliora interoperabilità, riproducibilità e conservazione dei dati nel lungo periodo. Facilita inoltre la ricerca dei dati nei portali e negli archivi, permettendo a ricercatori e decisori pubblici di riutilizzare efficacemente i dataset ambientali.

Scelta del formato di dati appropriato

La scelta di un formato adeguato è un passaggio cruciale nella gestione dei dati ambientali. Dipende da diversi fattori, tra cui dimensioni, struttura, componenti spaziali e temporali, obiettivi di analisi e destinatari. Semplici formati tabellari possono essere sufficienti per dataset piccoli e ben documentati, mentre per la modellazione e l'analisi ambientale su larga scala sono necessari formati multidimensionali più complessi.

In molti casi i processi di lavoro sui dati ambientali prevedono conversioni tra formati per rispondere a specifiche esigenze di analisi o visualizzazione. Conoscere punti di forza e limiti dei formati più comuni permette di prendere decisioni consapevoli e riduce il rischio di perdere o interpretare male i dati.

Caso applicativo: esplorazione dei formati di dati ambientali con QGIS

I Sistemi Informativi Geografici (GIS) offrono un ambiente potente per lavorare con dati ambientali spaziali provenienti da più fonti e in formati differenti. QGIS è una piattaforma GIS open source ampiamente utilizzata, che supporta numerosi formati di dati ambientali e procedure analitiche. Questo caso applicativo mostra come esplorare, visualizzare e convertire tipi e formati di dati ambientali con QGIS, evidenziando difficoltà pratiche e buone pratiche di gestione.

Panoramica del caso applicativo

Il caso applicativo riguarda l'integrazione e l'analisi, all'interno di un unico progetto GIS, di dataset ambientali conservati in diversi formati. Gli studenti lavorano con tre dataset rappresentativi: dati di monitoraggio della qualità dell'aria in CSV, dati sulla copertura del suolo in

formato vettoriale (shapefile o GeoJSON) e dati di temperatura derivati da satellite in formato raster (GeoTIFF o NetCDF). Le differenze di struttura, dimensionalità e metadati li rendono adatti a illustrare i concetti introdotti nelle sezioni precedenti.

L'obiettivo non è svolgere analisi spaziali avanzate, ma acquisire competenze pratiche nel caricamento, nell'esame e nella gestione di diversi formati ambientali. L'utilizzo diretto di QGIS permette agli studenti di comprendere come i formati influiscano su visualizzazione, allineamento spaziale e interpretabilità.

Caricamento di dati ambientali tabellari (CSV)

La prima fase consiste nel caricare misurazioni della qualità dell'aria conservate in un file CSV. Di norma il dataset include colonne come identificativo della stazione, latitudine, longitudine, data e concentrazione dell'inquinante. Durante l'importazione in QGIS gli studenti devono indicare le colonne che rappresentano le coordinate e scegliere il sistema di riferimento appropriato.

Questa fase evidenzia aspetti importanti dei dati ambientali tabellari. Coordinate definite in modo errato possono collocare i punti in posizioni sbagliate, mentre l'assenza di metadati può richiedere ipotesi su unità di misura o standard di misurazione. Dopo il caricamento, i dati CSV sono mostrati come punti e gli studenti possono esaminare visivamente la distribuzione spaziale delle stazioni.

Gli studenti sono invitati a esplorare le tabelle degli attributi, applicare filtri di base e analizzare gli andamenti temporali mediante interrogazioni sugli attributi. Si dimostra così come i dati strutturati possano acquisire una componente spaziale ed essere integrati nei flussi di lavoro GIS.

Lavorare con dati ambientali vettoriali

Il secondo dataset contiene informazioni sulla copertura del suolo in un formato vettoriale, come shapefile o GeoJSON. Rappresenta elementi spaziali discreti e include attributi che descrivono le categorie di uso del suolo. Dopo averlo caricato in QGIS, gli studenti ne esaminano i tipi di geometria, i campi degli attributi e il sistema di riferimento delle coordinate.

Questa parte dell'esercizio sottolinea l'importanza dei sistemi di riferimento spaziale e della coerenza degli attributi. Gli studenti confrontano il sistema di coordinate dei dati sulla copertura del suolo con quello dei punti di qualità dell'aria e, se necessario, riproiettano i livelli. Esplorano inoltre opzioni di simbologia di base per visualizzare le classi di copertura del suolo e comprendere il collegamento tra attributi vettoriali ed elementi spaziali.

Questo processo permette agli studenti di comprendere punti di forza e limiti dei formati vettoriali, tra cui struttura dei file, archiviazione degli attributi e interoperabilità.

Esplorazione di dati ambientali raster

Il terzo dataset è un raster che rappresenta la temperatura superficiale derivata da satellite o un'altra variabile ambientale continua. Può essere fornito in GeoTIFF o NetCDF. I dati raster vengono caricati in QGIS come livelli a griglia, nei quali ogni cella contiene un valore numerico della variabile misurata o modellata.

Gli studenti esaminano proprietà del raster quali risoluzione spaziale, estensione e tipo di dati. Per interpretare i pattern spaziali utilizzano tecniche di visualizzazione come scale cromatiche e classificazione dei valori. Con un dataset NetCDF esplorano come accedere a più variabili o istanti temporali contenuti nello stesso file.

Questa fase mostra perché i formati raster sono adatti a rappresentare fenomeni ambientali continui e come i metadati incorporati sostengano un'interpretazione spaziale accurata.

Confronto e integrazione di più formati di dati

Una volta caricati tutti i dataset, gli studenti sovrappongono i livelli vettoriali e raster nello stesso progetto QGIS. L'integrazione consente di valutare visivamente le relazioni tra misure della qualità dell'aria, tipi di copertura del suolo e andamenti della temperatura. Si possono eseguire semplici interrogazioni spaziali, come individuare le stazioni di monitoraggio situate all'interno di determinate classi di copertura del suolo, per dimostrare l'integrazione pratica di diversi tipi di dati.

Il confronto evidenzia la complementarità dei formati ambientali e il ruolo dei GIS come piattaforme unificanti per analisi multiformato. Gli studenti osservano inoltre come differenze nella qualità e nella risoluzione dei dati influenzino l'interpretazione.

Conversione ed esportazione dei dati

Una componente importante del caso applicativo è la conversione dei dati. Gli studenti si esercitano a esportare livelli vettoriali da shapefile a GeoJSON e livelli raster da GeoTIFF a un altro formato supportato. L'esercizio ribadisce che i formati vengono spesso scelti in funzione di compiti specifici, requisiti software o esigenze di condivisione.

Particolare attenzione è dedicata alla conservazione dei metadati durante la conversione e alla comprensione di possibili limitazioni o perdite di dati. Gli studenti sono invitati a documentare i passaggi della conversione e a verificare il risultato.

Esercitazioni pratiche — Attività guidate passo dopo passo con QGIS

Queste esercitazioni pratiche sono progettate per offrire un'esperienza diretta con i tipi e i formati di dati ambientali utilizzando QGIS, un Sistema Informativo Geografico gratuito e open source. Gli studenti lavoreranno con dataset ambientali reali in formati differenti, imparando a caricarli, visualizzarli, integrarli e convertirli.

Fonti dei dati (gratuite e ad accesso aperto)

1. Dati di monitoraggio della qualità dell'aria (CSV)

Dataset gratuiti sulla qualità dell'aria e sulle stazioni di monitoraggio sono disponibili presso l'Agenzia europea dell'ambiente (EEA) e altri portali pubblici di dati aperti.

• Agenzia europea dell'ambiente - Dati e mappe ambientali

https://www.eea.europa.eu/Data and Maps

Gli studenti possono scaricare le misurazioni della qualità dell'aria ed esportarle in file CSV contenenti coordinate delle stazioni, marcature temporali e concentrazioni di inquinanti.

Opzione alternativa:

Il plugin NASA POWER per QGIS offre accesso diretto a variabili climatiche e atmosferiche (temperatura, precipitazioni, radiazione) senza dover scaricare manualmente i dati.

---

2. Dati sulla copertura del suolo (raster / GeoTIFF o servizi web)

ESA WorldCover fornisce un dataset globale sulla copertura del suolo derivato dalle immagini dei satelliti Sentinel.

• ESA WorldCover (download o servizi web)

http://ESA WorldCover.org

Servizi WMTS/WMS (utilizzabili direttamente in QGIS):
• WMTS: https://services.terrascope.be/wmts/v2
• WMS: https://services.terrascope.be/wms/v2

---

3. Immagini satellitari (raster / GeoTIFF)

Immagini satellitari gratuite adatte all'analisi ambientale possono essere ottenute da:

• Copernicus Open Access Hub (missioni Sentinel)
https://dataspace.copernicus.euhttps://dataspace.copernicus.eu

• USGS EarthExplorer (archivio Landsat)
https://earthexplorer.usgs.govhttps://earthexplorer.usgs.gov

A causa delle dimensioni dei file, si consiglia agli studenti di scaricare piccole porzioni o tasselli delle immagini.

---

4. Dati vettoriali (OpenStreetMap)

OpenStreetMap offre dati vettoriali gratuiti relativi a strade, uso del suolo, confini e infrastrutture.

• OpenStreetMap

https://www.openstreetmap.org

I dati possono essere importati in QGIS tramite il plugin QuickOSM.

Esercizio 1 — Importare e visualizzare dati CSV sulla qualità dell'aria

Obiettivo: caricare dati tabellari sulla qualità dell'aria e visualizzare le stazioni di monitoraggio come elementi puntuali.

Passaggi:

1. Aprire QGIS e selezionare Layer → Aggiungi layer → Aggiungi layer testo delimitato.
2. Selezionare il file CSV scaricato (ad esempio air_quality.csv).
3. Impostare il campo X sulla longitudine e il campo Y sulla latitudine.
4. Selezionare il sistema di riferimento delle coordinate (EPSG:4326).

5. Fare clic su Aggiungi per visualizzare i punti sulla mappa.
6. Aprire la tabella degli attributi (F6) per esaminare le variabili disponibili.

Discussione:
Quali attributi sono inclusi nel dataset? Le unità di misura e le marcature temporali sono definite chiaramente?

Esercizio 2 — Caricare e rappresentare graficamente i dati

Obiettivo: visualizzare dati sulla copertura del suolo mediante livelli raster e applicare una simbologia tematica.

Passaggi:

1. In QGIS scegliere Layer → Aggiungi layer → Aggiungi layer WMS/WMTS.
2. Creare una nuova connessione utilizzando l'URL WMTS di ESA WorldCover.
3. Selezionare un livello di copertura del suolo e aggiungerlo al progetto.
4. Aprire lo stile del layer (F7) e applicare una combinazione di colori categoriale alle classi di copertura del suolo.
5. Aggiungere una legenda per interpretare le categorie di copertura del suolo.

Discussione:
Come varia la distribuzione della copertura del suolo nell'area di studio?

Esercizio 3 — Caricare ed esplorare dati raster satellitari

Obiettivo: caricare dati raster satellitari ed esplorare gli andamenti spaziali.

Passaggi:

1. Scaricare un'immagine GeoTIFF Sentinel-2 o Landsat dai portali Copernicus o USGS.
2. In QGIS, selezionare Layer → Aggiungi layer → Aggiungi layer raster e caricare il file.
3. Esaminare le proprietà del raster, come risoluzione, estensione e sistema di coordinate.
4. Applicare scale cromatiche o creare una composizione a falsi colori.
5. (Facoltativo) Utilizzare il Calcolatore raster per calcolare l'NDVI.

Discussione:
Quali elementi ambientali sono visibili nei dati raster?

Esercizio 4 — Integrare livelli vettoriali e raster

Obiettivo: combinare più formati e svolgere una semplice interpretazione spaziale.

Passaggi:

1. Sovrapporre i punti di monitoraggio della qualità dell'aria ai livelli di copertura del suolo e ai raster satellitari.
2. Utilizzare lo strumento Identifica elementi per esaminare i valori raster in corrispondenza dei punti.
3. (Facoltativo) Utilizzare Raster → Statistiche zonali per riassumere i valori raster per classe di copertura del suolo.

Discussione:
Si osservano relazioni tra le misure della qualità dell'aria e i tipi di copertura del suolo?

Esercizio 5 — Convertire ed esportare formati di dati

Obiettivo: convertire dati ambientali tra formati differenti.

Passaggi:

1. Fare clic con il pulsante destro su un layer vettoriale e scegliere Esporta → Salva elementi con nome….
2. Esportare il layer in formato GeoJSON.
3. Per i dati raster, scegliere Raster → Conversione → Traduci (converti formato).
4. Salvare il raster in un formato alternativo (ad esempio GeoTIFF o GeoPackage).
5. Verificare che il riferimento spaziale e gli attributi siano stati conservati.

Discussione:
Quali formati sono più adatti alla condivisione dei dati e alla loro conservazione a lungo termine?

Valutazione

Domanda 1. Quale tipo di dati ambientali è rappresentato meglio da una struttura tabellare con righe e colonne?

A) Dati non strutturati

B) Dati semistrutturati

C) Dati strutturati

D) Dati raster

Risposta corretta: C) Dati strutturati

Domanda 2. Quale formato è più adatto a conservare dati climatici o atmosferici multidimensionali con metadati incorporati?

A) CSV
B) GeoJSON
C) NetCDF
D) Shapefile

Risposta corretta: C) NetCDF

Domanda 3. Quale tipo di dati utilizzato in QGIS rappresenta principalmente fenomeni ambientali continui, come temperatura o altitudine?

A) Vettoriale (punti, linee, poligoni)
B) Raster

C) Tabellare

D) Testuale

Risposta corretta: B) Raster

Domanda 4. Qual è un vantaggio fondamentale di GeoJSON rispetto a ESRI Shapefile?

A) Supporta file più grandi di qualsiasi altro formato

B) Conserva geometria e attributi in un unico file leggibile dalle persone

C) È più veloce per l'analisi raster

D) È obbligatorio per tutte le applicazioni GIS

Risposta corretta: B) Conserva geometria e attributi in un unico file leggibile dalle persone

---

Domanda 5. Perché i metadati sono particolarmente importanti nei dataset ambientali?

A) Riducono la dimensione dei file

B) Migliorano la velocità di visualizzazione

C) Descrivono origine dei dati, unità di misura e riferimenti spaziali e temporali

D) Sostituiscono la necessità di analizzare i dati

Risposta corretta: C) Descrivono origine dei dati, unità di misura e riferimenti spaziali e temporali

---

Attività pratiche (4–5 esercizi)

Gli studenti vengono valutati sulla base del completamento delle seguenti esercitazioni pratiche con QGIS e dataset ambientali aperti.

1. **Importazione ed esame dei dati. Caricare in QGIS un dataset CSV contenente misurazioni ambientali, definire il sistema di riferimento corretto ed esaminare i campi degli attributi e i metadati.**
2. **Gestione di dati vettoriali e raster. Caricare dati vettoriali sulla copertura del suolo e immagini satellitari raster, verificare l'allineamento spaziale e applicare una simbologia appropriata.**
3. **Integrazione dei dati. Sovrapporre dataset puntuali, vettoriali e raster e interpretare le relazioni spaziali tra variabili ambientali.**
4. **Conversione dei formati. Convertire dati vettoriali da Shapefile a GeoJSON ed esportare dati raster in un altro formato supportato, conservando le informazioni sul riferimento spaziale.**
5. **Visualizzazione e interpretazione (facoltativo / bonus). Creare una semplice composizione cartografica con legenda e titolo e interpretare brevemente gli andamenti ambientali osservati.**

Riferimenti bibliografici / Letture di approfondimento

- Davis, Jerry. *Introduction to Environmental Data Science.* Bookdown.org, 2022. Open Access. https://bookdown.org/igisc/EnvDataSci.
- Earth-Env-Data-Science Project. *Earth & Environmental Data Science – An Introduction.* 2023. Open Educational Resource. https://earth-env-data-science.github.io/intro.html.

- Lovelace, Robin, Jakub Nowosad, and Jannes Muenchow. *Geocomputation with R.* CRC Press, 2019. Open Access edition. https://r.geocompx.org/.
- Rey, Sergio J. and Levi J. Wolf. Geo-data science with Python. Geographicdata.science, 2022. Open Textbook. https://geographicdata.science/book/intro.html.
- Bolstad, Paul, and Michael Shin. *Essentials of Geographic Information Systems.* 2nd ed., University of Minnesota Libraries Publishing, 2018. Open Textbook Library, https://open.umn.edu/opentextbooks/textbooks/67.