Uvod v vrste in formate okoljskih podatkov

Učni izidi

Po zaključku te enote bodo študenti sposobni:

- Opredeliti ključne vrste okoljskih podatkov in razumeti njihov pomen pri okoljskih raziskavah.
- Razlikovati med strukturiranimi, polstrukturiranimi in nestrukturiranimi okoljskimi podatki.
- Prepoznati običajne formate podatkov, ki se uporabljajo v okoljski znanosti (CSV, NetCDF, HDF5, GeoJSON, Shapefile itd.).
- Oceniti prednosti in omejitve različnih formatov podatkov pri shranjevanju, izmenjavi in analizi.
- Pretvarjati in obdelovati datoteke z okoljskimi podatki z ustreznimi orodji in knjižnicami (npr. Python, QGIS).

Uvod

Okoljski podatki imajo temeljno vlogo pri razumevanju, spremljanju in upravljanju naravnih sistemov ter sistemov pod vplivom človeka. Zagotavljajo empirično osnovo za preučevanje podnebnih sprememb, kakovosti zraka in vode, biotske raznovrstnosti, rabe tal ter številnih drugih okoljskih procesov. Napredek na področju senzorskih tehnologij, satelitskih opazovalnih sistemov in računalniških orodij je povzročil rast obsega, raznovrstnosti in kompleksnosti okoljskih podatkov brez primere. Zato morajo strokovnjaki in študenti okoljskih ved znati analizirati podatke ter razumeti različne vrste in formate, v katerih se okoljski podatki shranjujejo in izmenjujejo.

Okoljski podatki izvirajo iz različnih virov. Meritve na lokaciji se zbirajo s talnimi instrumenti, kot so vremenske postaje, senzorji kakovosti zraka, rečne merilne postaje in ploskve za ekološko spremljanje. Platforme za daljinsko zaznavanje, med katere spadajo sateliti, brezpilotni letalniki in letala, zagotavljajo obsežna prostorska opazovanja Zemljinega površja in ozračja. Poleg tega lahko okoljski podatki nastajajo z numeričnimi modeli in simulacijami kot so podnebni ali hidrološki modeli, ki ustvarjajo velike večrazsežne podatkovne nize. Vsak od teh virov ustvarja podatke z različnimi strukturami, merili in tehničnimi zahtevami, ki vplivajo na način njihovega shranjevanja in uporabe.

Razumevanje vrst in formatov okoljskih podatkov je bistveno za učinkovito ravnanje s podatki in njihovo analizo. Vrste podatkov opisujejo, kako so informacije organizirane, na primer kot

strukturirani tabelarični podatki, polstrukturirani zapisi ali nestrukturirane datoteke, kot so slike ali besedilo. Formati podatkov pa določajo, kako so podatki kodirani v datotekah kot so CSV, Parquet, NetCDF, HDF5, GeoTIFF ali GeoJSON. Izbira formata podatkov vpliva na dostopnost podatkov, učinkovitost shranjevanja, interoperabilnost programskih orodij in možnost vključitve metapodatkov, ki opisujejo izvor podatkov, enote, prostorsko referenco in časovni obseg.

Neustrezno ravnanje s formati podatkov v okoljski znanosti lahko povzroči napačno razlago, izgubo podatkov ali slabšo ponovljivost rezultatov. Če na primer pri prostorskih podatkih ne upoštevamo koordinatnih referenčnih sistemov (CRS), se lahko kartografski sloji nepravilno prekrivajo, uporaba preveč preprostih formatov pa lahko povzroči izgubo pomembnih metapodatkov. Nasprotno pa izbira ustreznih formatov podatkov omogoča učinkovito izmenjavo podatkov, njihovo dolgoročno ohranjanje ter povezovanje raznovrstnih podatkovnih nizov iz več virov. To je zlasti pomembno pri interdisciplinarnih okoljskih raziskavah, pri katerih je treba združevati in primerjati podatkovne nize z različnih področij.

Pričujoče predavanje predstavlja glavne vrste okoljskih podatkov ter najpogosteje uporabljene formate podatkov v okoljski znanosti in geoprostorskih aplikacijah. Podaja konceptualni okvir za razumevanje strukturiranja in shranjevanja okoljskih podatkov, podprt s praktičnimi primeri spremljanja kakovosti zraka, podnebnih opazovanj in analize rabe tal. S poznavanjem teh konceptov bodo študenti bolje pripravljeni na delo z dejanskimi okoljskimi podatkovnimi nizi, izbiro ustreznih formatov za analizo ter zagotavljanje preglednih, ponovljivih in znanstveno utemeljenih delovnih postopkov pri delu s podatki.

Vrste okoljskih podatkov

Okoljske podatke je mogoče razvrstiti na več načinov glede na njihovo strukturo, razsežnost, prostorske in časovne značilnosti ter način zbiranja. Jasno razumevanje vrst okoljskih podatkov je bistveno za izbiro ustreznih formatov shranjevanja, analitičnih tehnik in načinov vizualizacije. V tem razdelku so okoljski podatki glede na strukturne značilnosti razvrščeni v tri široke skupine: strukturirane, polstrukturirane in nestrukturirane podatke. Posebna pozornost je namenjena tudi prostorski in časovni razsežnosti, ki sta pri okoljskih aplikacijah še posebej pomembni.

Strukturirani okoljski podatki

Strukturirani podatki so zelo urejeni in so običajno shranjeni v tabelarični obliki, v kateri so podatki razporejeni v vrstice in stolpce z jasno določenimi atributi in tipi podatkov. Vsaka vrstica

praviloma predstavlja opazovanje, vsak stolpec pa spremenljivko ali atribut. Ta vrsta podatkov je pogosta v programih okoljskega spremljanja in se pogosto zbira s standardiziranimi merilnimi postopki.

Primeri strukturiranih okoljskih podatkov vključujejo časovne vrste temperature zraka, meritve padavin, koncentracije onesnaževal zraka, kazalnike kakovosti vode ali lastnosti tal, izmerjene na stalnih lokacijah. Ti podatkovni nizi se pogosto shranjujejo v formatih, kot so datoteke CSV, preglednice ali relacijske podatkovne baze. Po strukturiranih podatkih je razmeroma enostavno poizvedovati ter jih filtrirati in analizirati s statistično programsko opremo, podatkovnimi bazami in programskimi jeziki, kot sta R ali Python.

CSV je preprost vrstično zasnovan besedilni format, ki ga je enostavno brati in izmenjevati, vendar je neučinkovit pri obsežni analitični obdelavi in dolgoročnem shranjevanju. Apache Parquet pa je stolpčno zasnovan format, optimiziran za analitična opravila, ki omogoča hitrejše poizvedbe po izbranih spremenljivkah, učinkovito stiskanje in ohranjanje tipov podatkov. Zato je Parquet primernejši za velike okoljske podatkovne nize in podatkovno intenzivne delovne postopke, CSV pa ostaja uporaben za majhne podatkovne nize in izmenjavo podatkov.

Ena ključnih prednosti strukturiranih podatkov je njihova preprostost in združljivost s številnimi orodji. Omogočajo učinkovito združevanje, statistično analizo in vizualizacijo, kot so izračunavanje povprečij, trendov ali korelacij skozi čas. Vendar imajo strukturirani podatki tudi omejitve. Manj primerni so za prikaz kompleksnih prostorskih struktur, večrazsežnih spremenljivk ali velikih rastrskih podatkovnih nizov. Poleg tega se lahko pomembne kontekstualne informacije izgubijo, če metapodatki niso skrbno dokumentirani in shranjeni ločeno.

Polstrukturirani okoljski podatki

Polstrukturirani podatki niso skladni s togimi tabelaričnimi strukturami, vendar vseeno vsebujejo organizacijske elemente, kot so oznake, ključi ali hierarhična razmerja. Ta vrsta podatkov je prilagodljivejša od strukturiranih podatkov, hkrati pa ohranja določeno raven organiziranosti, ki omogoča avtomatizirano obdelavo.

V okoljski znanosti se polstrukturirani podatki pogosto uporabljajo za predstavitev metapodatkov, izhodnih podatkov senzorjev in prostorskih objektov. Formata, kot sta XML in JSON, se pogosto uporabljata za izmenjavo podatkov med sistemi, zlasti v spletnih storitvah in programskih vmesnikih (APIs). GeoJSON, specializirana razširitev formata JSON, se pogosto uporablja za predstavitev geografskih objektov skupaj z njihovimi atributi in geometrijo.

Polstrukturirani podatki so še posebej uporabni pri delu z raznovrstnimi podatkovnimi nizi, ki se razlikujejo po strukturi, ali kadar se lahko sčasoma dodajajo novi atributi. Omrežja okoljskih senzorjev na primer pogosto ustvarjajo polstrukturirane izhodne podatke, ki vključujejo meritve, časovne žige, identifikatorje senzorjev in oznake kakovosti. Prilagodljivost polstrukturiranih podatkov omogoča lažje povezovanje med platformami, vendar lahko v primerjavi s strukturiranimi tabelaričnimi podatki zahteva zahtevnejše razčlenjevanje in preverjanje veljavnosti.

Nestrukturirani okoljski podatki

Nestrukturirani podatki nimajo vnaprej določenega podatkovnega modela in jih ni mogoče preprosto predstaviti v tabelah ali hierarhičnih strukturah. V to skupino spadajo slike, videoposnetki, zvočni posnetki, poročila v prostem besedilu in neobdelani posnetki daljinskega zaznavanja. V okoljskem kontekstu so nestrukturirani podatki zaradi napredka tehnologij opazovanja Zemlje in avtomatiziranih sistemov za zbiranje podatkov vse pomembnejši.

Primeri nestrukturiranih okoljskih podatkov vključujejo satelitske posnetke, aerofotografije, posnetke z brezpilotnimi letalniki, skenirane terenske zapiske in besedilne presoje vplivov na okolje. Te vrste podatkov pogosto zahtevajo specializirane tehnike obdelave, kot so analiza slik, obdelava naravnega jezika ali metode strojnega učenja. Čeprav lahko nestrukturirani podatki vsebujejo bogate in podrobne informacije, sta njihova analiza in učinkovito shranjevanje zahtevnejša.

Analiza nestrukturiranih okoljskih podatkov je pogosto odvisna od dodatnih ravni interpretacije ali pretvorbe. Satelitske posnetke je mogoče na primer obdelati za izpeljavo vegetacijskih indeksov ali klasifikacij pokrovnosti tal, s čimer se nestrukturirani rastrski podatki pretvorijo v strukturirane ali polstrukturirane podatkovne nize, primerne za nadaljnjo analizo. Nestrukturirani podatki zato pogosto služijo kot surovi vhodni podatki za ustvarjanje bolj strukturiranih okoljskih informacij.

Prostorski okoljski podatki

Prostorski podatki predstavljajo informacije, ki so izrecno povezane z geografsko lokacijo. Lokacija je lahko določena s koordinatami, upravnimi mejami, mrežami ali prostorskimi referenčnimi sistemi. Prostorski podatki so temeljnega pomena za okoljsko znanost, saj je večina okoljskih pojavov neločljivo povezana s prostorom.

Prostorski okoljski podatki se običajno delijo na vektorske in rastrske podatke. Vektorski podatki predstavljajo diskretne objekte, kot so merilne postaje, reke, ceste ali zemljiške parcele, s točkami, črtami in poligoni. Rastrski podatki predstavljajo zvezne pojave, kot so temperaturna polja, modeli nadmorskih višin ali satelitski posnetki, z mrežo celic oziroma slikovnih pik.

Razumevanje vrst prostorskih podatkov je ključno za okoljske analize in kartiranje. Dejavniki, kot so koordinatni referenčni sistemi (CRS), prostorska ločljivost in merilo, lahko pomembno vplivajo na rezultate analize. Neustrezno ravnanje s prostorskimi podatki lahko povzroči neusklajene sloje, netočne meritve ali napačne razlage. Zato je treba prostorske podatke skrbno upravljati s specializiranimi orodji, kot so geografski informacijski sistemi (GIS).

Časovni okoljski podatki

Časovne značilnosti so še en opredeljujoč vidik okoljskih podatkov. Številni okoljski procesi se skozi čas spreminjajo, od kratkoročnih nihanj vremenskih razmer do dolgoročnih trendov podnebja ali sprememb ekosistemov. Časovni podatki se lahko zbirajo v rednih ali nerednih časovnih presledkih ali pa kot neprekinjene meritve.

Podatki časovnih vrst so pogosta oblika časovnih okoljskih podatkov. Mednje spadajo urne meritve kakovosti zraka, dnevne vrednosti pretoka rek ali mesečni podnebni indeksi. Pri delu s časovno odvisnimi okoljskimi podatki je pomembno upoštevati časovno ločljivost, dogovore glede časovnih pasov in vrzeli v podatkih. Pri združevanju podatkovnih nizov iz različnih virov je bistvena ustrezna časovna uskladitev.

Okoljski podatki so pogosto hkrati prostorski in časovni, zato nastanejo prostorsko-časovni podatkovni nizi. Ti podatkovni nizi so lahko kompleksni in večrazsežni, kot so izhodni podatki globalnih podnebnih modelov ali časovne vrste, izpeljane iz satelitskih podatkov. Ravnanje s prostorsko-časovnimi podatki zahteva specializirane podatkovne strukture in formate, ki omogočajo učinkovito shranjevanje velikih količin informacij ter dostop do njih.

Večrazsežni in mrežni podatki

Večrazsežni podatki predstavljajo okoljske spremenljivke, ki se spreminjajo v več razsežnostih, kot so geografska širina, geografska dolžina, nadmorska višina in čas. Ti podatki so pogosti v atmosferskih in oceanografskih znanostih, kjer se meritve izvajajo v tridimenzionalnem prostoru in v dolgih časovnih obdobjih.

Mrežni podatki so posebna vrsta večrazsežnih podatkov, pri katerih so vrednosti določene na pravilni mreži. Mednje spadajo podatkovni nizi podnebnih reanaliz, izhodni podatki numeričnega

napovedovanja vremena in modeli oceanskega kroženja. Ti podatkovni nizi pogosto vsebujejo več spremenljivk in obsežne metapodatke, zaradi česar so bistveno kompleksnejši od preprostih tabelaričnih podatkov.

Zaradi kompleksnosti večrazsežnih okoljskih podatkov je treba uporabljati specializirane formate in programska orodja, ki lahko učinkovito obdelujejo velika podatkovna polja in ohranjajo metapodatke. Pred izbiro ustreznih formatov podatkov in analitičnih delovnih postopkov je bistveno razumeti naravo teh vrst podatkov.

Običajni formati okoljskih podatkov

Formati okoljskih podatkov določajo, kako se podatki kodirajo, shranjujejo in izmenjujejo med sistemi. Izbira formata podatkov pomembno vpliva na njihovo dostopnost, učinkovitost shranjevanja, interoperabilnost in dolgoročno uporabnost. V okoljski znanosti morajo formati podatkov pogosto podpirati velike količine podatkov, prostorsko in časovno razsežnost ter podrobne metapodatke. To podpoglavje predstavlja najpogosteje uporabljene formate podatkov v okoljskih in geoprostorskih aplikacijah ter izpostavlja njihove značilnosti, prednosti in omejitve.

Tabelarični formati podatkov

Tabelarični formati so zaradi svoje preprostosti in široke združljivosti med najpogosteje uporabljenimi formati podatkov v okoljski znanosti. Podatke shranjujejo v vrsticah in stolpcih ter se pogosto uporabljajo za strukturirane okoljske podatkovne nize, kot so meritve časovnih vrst in zbirne statistike.

Najpogostejši tabelarični format je Comma-Separated Values (CSV). Datoteke CSV so navadne besedilne datoteke, v katerih vsaka vrstica predstavlja zapis, vrednosti pa so ločene z vejicami ali drugimi ločili. CSV je preprosto ustvariti, brati in izmenjevati med različnimi programskimi platformami. Pogosto se uporablja za izmenjavo meritev kakovosti zraka, meteoroloških opazovanj in podatkov o kakovosti vode. Vendar datoteke CSV nimajo standardizirane podpore za metapodatke, tipe podatkov in informacije o prostorski referenci, kar lahko ob pomanjkljivi dokumentaciji povzroči dvoumnost ali izgubo podatkov.

Formati preglednic, kot so datoteke Microsoft Excel, se prav tako pogosto uporabljajo, zlasti v začetnih fazah zbiranja podatkov ali pri majhnih podatkovnih nizih. Čeprav preglednice ponujajo uporabniku prijazne vmesnike in osnovne funkcije preverjanja veljavnosti podatkov, so manj primerne za velike podatkovne nize, avtomatizirano obdelavo ali ponovljive delovne postopke.

Skrito oblikovanje, več delovnih listov in ročne spremembe lahko povzročijo napake ter zmanjšajo preglednost.

Stolpčni in analitični formati podatkov

Apache Parquet je odprtokodni stolpčni format za shranjevanje podatkov, optimiziran za analitična opravila in obsežno obdelavo podatkov. V nasprotju s tradicionalnimi vrstično zasnovanimi tabelaričnimi formati, kot je CSV, Parquet shranjuje podatke po stolpcih, kar znatno izboljša zmogljivost poizvedb, ki dostopajo le do podmnožice spremenljivk. Zato je Parquet še posebej primeren za velike okoljske podatkovne nize z veliko atributi in opazovanji.

V okoljski znanosti se Parquet vse pogosteje uporablja za shranjevanje tabelaričnih podatkov velikega obsega, kot so dolgoročne meritve kakovosti zraka, izhodni podatki podnebnih modelov, pretvorjeni v tabelarično obliko, in združeni senzorski podatki. Format podpira učinkovito stiskanje in kodiranje, zato zmanjšuje potrebe po prostoru za shranjevanje, hkrati pa ohranja hitro branje. Parquet ohranja tudi tipe podatkov in informacije o shemi, s čimer odpravlja številne omejitve formatov navadnega besedila.

Čeprav Parquet ni izvorni format GIS, ima pomembno vlogo v postopkih obdelave okoljskih podatkov, pri katerih se podatki obdelajo na oblačnih platformah, v podatkovnih skladiščih ali porazdeljenih računalniških okoljih, preden se vizualizirajo ali prostorsko analizirajo v programski opremi GIS. Podatke, shranjene v formatu Parquet, je mogoče pretvoriti v prostorske formate ali jih z orodji GIS povezati prek vmesnih korakov obdelave z uporabo jezika Python, mehanizmov na osnovi SQL ali knjižnic za prostorske podatke.

Znanstveni in večrazsežni formati podatkov

Znanstveni formati podatkov so namenjeni učinkovitemu shranjevanju velikih večrazsežnih podatkovnih nizov in vključevanju bogatih metapodatkov, ki opisujejo strukturo podatkov, enote in koordinatne sisteme. Ti formati so še posebej pomembni v podnebni znanosti, meteorologiji, oceanografiji in raziskavah ozračja.

NetCDF (Network Common Data Form) je eden najpogosteje uporabljenih znanstvenih formatov podatkov v okoljski znanosti. Podpira večrazsežna podatkovna polja in samoopisne metapodatke, zato je primeren za shranjevanje prostorsko-časovnih podatkov, kot so izhodni podatki podnebnih modelov, podatkovni nizi reanaliz in spremenljivke, izpeljane iz satelitskih podatkov. Datoteke NetCDF lahko vsebujejo več spremenljivk in razsežnosti v eni sami datoteki, kar omogoča učinkovit dostop do podatkov in izbiranje podmnožic.

HDF5 (Hierarchical Data Format version 5) je še en zmogljiv format za shranjevanje kompleksnih okoljskih podatkovnih nizov. Omogoča hierarhično strukturo, v kateri so podatki lahko organizirani v skupine in podatkovne nize, podobno kot datotečni sistem znotraj datoteke. HDF5 je zelo razširljiv in učinkovit, zato je primeren za zelo velike podatkovne nize. Vendar je lahko njegova kompleksnost za začetnike zahtevna, za branje in zapisovanje datotek HDF5 pa so pogosto potrebne specializirane knjižnice.

Formati prostorskih vektorskih podatkov

Formati vektorskih podatkov se uporabljajo za predstavitev diskretnih geografskih objektov, kot so točke, črte in poligoni. Ti formati so bistveni za kartiranje lokacij okoljskega spremljanja, upravnih meja, rek, cest in objektov rabe tal.

ESRI Shapefile je že vrsto let standardni vektorski format v sistemih GIS. Shapefile sestavlja več povezanih datotek, ki skupaj shranjujejo geometrijo, atribute in informacije o prostorski referenci. Čeprav je format Shapefile široko podprt, ima več omejitev, med njimi omejitve velikosti datotek, dolžine imen atributov in tipov podatkov.

GeoJSON je sodoben, odprt besedilni format za kodiranje geografskih objektov z uporabo JSON. Zaradi svoje preprostosti in berljivosti je še posebej priljubljen pri spletnem kartiranju in izmenjavi podatkov. GeoJSON podpira atribute in geometrijo v eni sami datoteki ter se dobro povezuje s spletnimi tehnologijami. Vendar lahko datoteke GeoJSON pri kompleksnih ali obsežnih podatkovnih nizih postanejo velike in neučinkovite.

Formati prostorskih rastrskih podatkov

Formati rastrskih podatkov se uporabljajo za predstavitev zveznih okoljskih pojavov, pri katerih so vrednosti shranjene v mreži celic oziroma slikovnih pik. Pogosti primeri so satelitski posnetki, digitalni modeli višin, karte temperature površja in vegetacijski indeksi.

GeoTIFF je pogosto uporabljen rastrski format, ki standardni format TIFF razširja z vdelanimi geografskimi metapodatki, kot so koordinatni referenčni sistemi (CRS) in prostorska ločljivost. Datoteke GeoTIFF se pogosto uporabljajo pri daljinskem zaznavanju in v aplikacijah GIS ter jih podpira večina geoprostorske programske opreme.

Drugi rastrski formati, kot sta NetCDF in HDF5, se prav tako pogosto uporabljajo za mrežne okoljske podatke, zlasti kadar so vključene časovne in navpične razsežnosti. Izbira rastrskega formata je pogosto odvisna od kompleksnosti podatkov, velikosti datoteke in zahtev analize.

Standardi za metapodatke in opis podatkov

Metapodatki so ključni za pravilno razumevanje in uporabo okoljskih podatkov. Opisujejo izvor, vsebino, strukturo in kakovost podatkov ter informacije o prostorskem in časovnem obsegu. Brez ustreznih metapodatkov je lahko okoljske podatkovne nize težko ali celo nemogoče pravilno razložiti.

V okoljskih in geoprostorskih aplikacijah se pogosto uporablja več standardov za metapodatke. ISO 19115 je mednarodni standard za geografske metapodatke, ki zagotavlja celovit okvir za opisovanje prostorskih podatkovnih nizov. Dublin Core je preprostejši in splošnejši standard za metapodatke, ki se uporablja na različnih področjih. Številni znanstveni formati podatkov, kot je NetCDF, podpirajo vdelane metapodatke, ki upoštevajo uveljavljene dogovore.

Pravilna uporaba standardov za metapodatke izboljšuje interoperabilnost, ponovljivost in dolgoročno ohranjanje podatkov. Omogoča tudi lažje odkrivanje podatkov prek podatkovnih portalov in repozitorijev, s čimer raziskovalcem in oblikovalcem politik omogoča učinkovito ponovno uporabo okoljskih podatkovnih nizov.

Izbira ustreznega formata podatkov

Izbira ustreznega formata podatkov je ključni korak pri upravljanju okoljskih podatkov. Odvisna je od več dejavnikov, med drugim od velikosti in strukture podatkov, njihovih prostorskih in časovnih razsežnosti, ciljev analize ter ciljne skupine. Preprosti tabelarični formati lahko zadostujejo za majhne in dobro dokumentirane podatkovne nize, za obsežno okoljsko modeliranje in analizo pa so potrebni kompleksni večrazsežni formati.

Delovni postopki pri delu z okoljskimi podatki pogosto vključujejo pretvarjanje podatkov med formati, da se izpolnijo posebne zahteve analize ali vizualizacije. Razumevanje prednosti in omejitev običajnih formatov okoljskih podatkov omogoča premišljeno odločanje ter zmanjšuje tveganje izgube ali napačne razlage podatkov.

Praktični primer uporabe: Raziskovanje formatov okoljskih podatkov s programom QGIS

Geografski informacijski sistemi (GIS) zagotavljajo zmogljivo okolje za delo s prostorskimi okoljskimi podatki iz več virov in v različnih formatih. QGIS je široko uporabljana odprtokodna platforma GIS, ki podpira širok nabor formatov okoljskih podatkov in analitičnih delovnih

postopkov. Ta praktični primer uporabe prikazuje, kako je mogoče s programom QGIS raziskovati, vizualizirati in pretvarjati vrste in formate okoljskih podatkov, pri tem pa poudarja praktične izzive in dobre prakse pri ravnanju z okoljskimi podatki.

Pregled primera uporabe

Primer uporabe se osredotoča na združevanje in analizo okoljskih podatkovnih nizov, shranjenih v različnih formatih, znotraj enega projekta GIS. Študenti delajo s tremi reprezentativnimi podatkovnimi nizi: podatki spremljanja kakovosti zraka, shranjeni v formatu CSV, podatki o pokrovnosti tal v obliki vektorskega podatkovnega niza (Shapefile ali GeoJSON) ter satelitsko pridobljeni podatki o temperaturi, shranjeni v rastrskem formatu (GeoTIFF ali NetCDF). Ti podatkovni nizi se razlikujejo po strukturi, razsežnosti in vsebini metapodatkov, zato so primerni za ponazoritev ključnih konceptov iz prejšnjih poglavij.

Cilj tega primera uporabe ni napredna prostorska analiza, temveč razvoj praktičnih veščin nalaganja, pregledovanja in upravljanja različnih formatov okoljskih podatkov. S praktičnim delom v programu QGIS študenti spoznajo, kako formati podatkov vplivajo na vizualizacijo, prostorsko usklajenost in možnost interpretacije podatkov.

Nalaganje tabelaričnih okoljskih podatkov (CSV)

Prvi korak vključuje nalaganje meritev kakovosti zraka, shranjenih v datoteki CSV. Podatkovni niz običajno vsebuje stolpce, kot so identifikator postaje, zemljepisna širina, zemljepisna dolžina, datum in koncentracija onesnaževala. Pri uvozu podatkov CSV v QGIS morajo študenti določiti, kateri stolpci predstavljajo prostorske koordinate, in izbrati ustrezen koordinatni referenčni sistem (CRS).

Ta korak izpostavlja pomembne vidike dela s tabelaričnimi okoljskimi podatki. Zaradi napačno opredeljenih koordinat se lahko točke prikažejo na napačnih lokacijah, manjkajoči metapodatki pa lahko zahtevajo predpostavke o merskih enotah ali merilnih standardih. Po nalaganju se podatki CSV prikažejo kot točkovni prostorski objekti, kar študentom omogoča vizualni pregled prostorske razporeditve merilnih postaj.

Študenti naj raziščejo atributne tabele, uporabijo osnovne filtre in z atributnimi poizvedbami preučijo časovne vzorce. To prikazuje, kako je mogoče strukturirane podatke opremiti s prostorsko komponento in jih vključiti v delovne postopke GIS.

Delo z vektorskimi okoljskimi podatki

Drugi podatkovni niz vsebuje informacije o pokrovnosti tal v vektorskem formatu, kot sta Shapefile ali GeoJSON. Ta podatkovni niz predstavlja diskretne prostorske objekte in vključuje atributne podatke, ki opisujejo kategorije rabe tal. Po nalaganju vektorskega podatkovnega niza v QGIS študenti pregledajo vrste geometrije, atributna polja in koordinatni referenčni sistem (CRS).

Ta del vaje poudarja pomen prostorskih referenčnih sistemov in doslednosti atributov. Študenti primerjajo koordinatni referenčni sistem (CRS) podatkovnega niza o pokrovnosti tal s koordinatnim referenčnim sistemom (CRS) točk kakovosti zraka ter po potrebi ponovno projicirajo sloje. Raziščejo tudi osnovne možnosti simbologije za vizualizacijo razredov pokrovnosti tal in razumevanje povezave med vektorskimi atributi ter prostorskimi objekti.

V tem postopku študenti pridobijo vpogled v prednosti in omejitve formatov vektorskih podatkov, vključno s strukturo datotek, shranjevanjem atributov in interoperabilnostjo.

Raziskovanje rastrskih okoljskih podatkov

Tretji podatkovni niz je rastrski podatkovni niz, ki predstavlja satelitsko pridobljeno površinsko temperaturo ali drugo zvezno okoljsko spremenljivko. Na voljo je lahko v formatu GeoTIFF ali NetCDF. Rastrski podatki se v QGIS naložijo kot sloji, temelječi na mreži, v katerih vsaka celica vsebuje številsko vrednost, ki predstavlja izmerjeno ali modelirano spremenljivko.

Študenti pregledajo lastnosti rastra, kot so prostorska ločljivost, prostorski obseg in podatkovni tip. Za razlago prostorskih vzorcev uporabijo tehnike vizualizacije, vključno z barvnimi lestvicami in razvrščanjem vrednosti. Če uporabijo podatkovni niz NetCDF, raziščejo, kako je mogoče znotraj ene datoteke dostopati do več spremenljivk ali časovnih korakov.

Ta korak ponazarja, kako primerni so rastrski formati za predstavljanje zveznih okoljskih pojavov in kako metapodatki, vdelani v rastrske datoteke, podpirajo natančno prostorsko razlago.

Primerjava in povezovanje več formatov podatkov

Ko so vsi podatkovni nizi naloženi, študenti v istem projektu QGIS prekrijejo vektorske in rastrske sloje. Tako lahko vizualno ocenijo povezave med meritvami kakovosti zraka, vrstami pokrovnosti tal in temperaturnimi vzorci. Za prikaz praktičnega povezovanja različnih vrst podatkov lahko izvedejo osnovne prostorske poizvedbe, na primer poiščejo merilne postaje znotraj določenih razredov pokrovnosti tal.

Ta primerjava poudarja, kako se formati okoljskih podatkov medsebojno dopolnjujejo in kako GIS deluje kot enotna platforma za analizo podatkov v več formatih. Študenti opazujejo tudi, kako razlike v kakovosti in ločljivosti podatkov vplivajo na razlago.

Pretvorba in izvoz podatkov

Pomemben del primera uporabe je pretvorba podatkov. Študenti vadijo izvoz vektorskih slojev iz formata Shapefile v format GeoJSON in rastrskih slojev iz formata GeoTIFF v drug podprt format. Vaja utrjuje spoznanje, da se formati podatkov pogosto izberejo glede na posamezna opravila, zahteve programske opreme ali potrebe po izmenjavi podatkov.

Posebna pozornost je namenjena ohranjanju metapodatkov med pretvorbo ter razumevanju morebitnih omejitev ali izgube podatkov. Študenti naj dokumentirajo korake pretvorbe in preverijo rezultat.

Praktične vaje — naloge v programu QGIS po korakih

Te praktične vaje so zasnovane tako, da študentom omogočijo pridobivanje praktičnih izkušenj z vrstami in formati okoljskih podatkov v brezplačnem in odprtokodnem geografskem informacijskem sistemu QGIS. Delali bodo z dejanskimi okoljskimi podatkovnimi nizi v različnih formatih ter se jih naučili nalagati, vizualizirati, povezovati in pretvarjati.

Viri podatkov (brezplačen in odprt dostop)

1. Podatki spremljanja kakovosti zraka (CSV)

Brezplačni podatkovni nizi o kakovosti zraka in merilnih postajah so na voljo pri Evropski agenciji za okolje (EEA) in na drugih portalih odprtih podatkov državnih organov.

• Evropska agencija za okolje – okoljski podatki in zemljevidi

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

Študenti lahko prenesejo meritve kakovosti zraka in jih izvozijo kot datoteke CSV, ki vsebujejo koordinate postaj, časovne žige in koncentracije onesnaževal.

Alternativna možnost:

Vtičnik NASA POWER za QGIS omogoča neposreden dostop do podnebnih in atmosferskih spremenljivk (temperatura, padavine, sevanje) brez ročnega prenosa podatkov.

2. Podatki o pokrovnosti tal (rastrski / GeoTIFF ali spletne storitve)

ESA WorldCover zagotavlja globalni podatkovni niz o pokrovnosti tal, izdelan iz satelitskih posnetkov Sentinel.

• ESA WorldCover (prenos ali spletne storitve)

http://ESA WorldCover.org

Storitve WMTS/WMS (uporabiti jih je mogoče neposredno v programu QGIS):
• WMTS: https://services.terrascope.be/wmts/v2
• WMS: https://services.terrascope.be/wms/v2

3. Satelitski posnetki (rastrski / GeoTIFF)

Brezplačne satelitske posnetke, primerne za okoljske analize, je mogoče pridobiti pri:

• **Copernicus Open Access Hub** (misije Sentinel)
https://dataspace.copernicus.eu

• **USGS EarthExplorer** (arhiv Landsat)
https://earthexplorer.usgs.gov

Študentom svetujemo, naj zaradi velikosti datotek prenesejo manjše izseke ali ploščice.

4. Vektorski podatki (OpenStreetMap)

OpenStreetMap zagotavlja brezplačne vektorske podatke o cestah, rabi tal, mejah in infrastrukturi.

• OpenStreetMap

https://www.openstreetmap.org

Podatke je mogoče uvoziti v QGIS z vtičnikom **QuickOSM**.

Vaja 1 — uvoz in vizualizacija podatkov o kakovosti zraka v formatu CSV

Cilj: Naložiti tabelarične podatke o kakovosti zraka in merilne postaje vizualizirati kot točkovne prostorske objekte.

Koraki:

1. Odprite QGIS in izberite **Layer → Add Layer → Add Delimited Text Layer**.
2. Poiščite preneseno datoteko CSV (npr. air_quality.csv).
3. Polje X nastavite na longitude, polje Y pa na latitude.
4. Izberite koordinatni referenčni sistem (CRS) EPSG:4326.
5. Kliknite **Add**, da se točke prikažejo na zemljevidu.
6. Odprite Attribute Table (F6), da pregledate razpoložljive spremenljivke.

Razprava:
Kateri atributi so vključeni v podatkovni niz? Ali so merske enote in časovni žigi jasno opredeljeni?

Vaja 2 — nalaganje in slogovno oblikovanje podatkov

Cilj: Vizualizirati podatke o pokrovnosti tal z rastrskimi sloji in uporabiti tematsko oblikovanje.

Koraki:

1. V programu QGIS pojdite na **Layer → Add Layer → Add WMS/WMTS Layer**.
2. Ustvarite novo povezavo z URL-jem storitve WMTS ESA WorldCover.
3. Izberite sloj pokrovnosti tal in ga dodajte projektu.
4. Odprite Layer Styling (F7) in uporabite barvno shemo po kategorijah za razrede pokrovnosti tal.
5. Dodajte legendo za razlago kategorij pokrovnosti tal.

Razprava:
Kako se razporeditev pokrovnosti tal spreminja na preučevanem območju?

Vaja 3 — nalaganje in raziskovanje satelitskih rastrskih podatkov

Cilj: Naložiti rastrske satelitske podatke in raziskati prostorske vzorce.

Koraki:

1. Prenesite posnetek Sentinel-2 ali Landsat v formatu GeoTIFF s portalov Copernicus ali USGS.
2. V QGIS-u izberite **Layer → Add Layer → Add Raster Layer** in naložite datoteko.
3. Preverite lastnosti rastra, kot so ločljivost, prostorski obseg in koordinatni sistem.
4. Uporabite barvne lestvice ali ustvarite kompozit v lažnih barvah.
5. (Izbirno) Z orodjem **Raster Calculator** izračunajte NDVI.

Razprava:
Katere okoljske značilnosti so vidne v rastrskih podatkih?

Vaja 4 — Združevanje vektorskih in rastrskih slojev

Cilj: Združiti več podatkovnih formatov in opraviti osnovno prostorsko interpretacijo.

Koraki:

1. Točke za spremljanje kakovosti zraka prikažite čez sloje pokrovnosti tal in satelitske rastrske sloje.
2. Z orodjem **Identify Features** preglejte vrednosti rastra na lokacijah točk.
3. (Izbirno) Z ukazom Raster → Zonal Statistics povzemite vrednosti rastra po razredih pokrovnosti tal.

Razprava:
Ali je mogoče opaziti povezave med meritvami kakovosti zraka in vrstami pokrovnosti tal?

Vaja 5 — Pretvorba in izvoz podatkovnih formatov

Cilj: Pretvoriti okoljske podatke iz enega formata v drugega.

Koraki:

1. Z desno tipko miške kliknite vektorski sloj in izberite Export → Save Features As….
2. Izvozite sloj v formatu **GeoJSON**.
3. Za rastrske podatke izberite **Raster → Conversion → Translate (Convert Format)**.
4. Shranite raster v drugem formatu (npr. GeoTIFF ali GeoPackage).
5. Preverite, ali so prostorska referenca in atributi ohranjeni.

Razprava:
Kateri formati so najprimernejši za izmenjavo podatkov in dolgoročno shranjevanje?

Ocenjevanje

Vprašanje 1: Katera vrsta okoljskih podatkov je najprimerneje predstavljena v tabelarni strukturi z vrsticami in stolpci?

A) Nestrukturirani podatki

B) Polstrukturirani podatki

C) Strukturirani podatki

D) Rastrski podatki

Pravilni odgovor: C) Strukturirani podatki

Vprašanje 2: Kateri podatkovni format je najprimernejši za shranjevanje večrazsežnih podnebnih ali atmosferskih podatkov z vgrajenimi metapodatki?

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

Pravilni odgovor: C) NetCDF

Vprašanje 3: Kateri podatkovni format v QGIS-u se uporablja predvsem za prikaz zveznih okoljskih pojavov, kot sta temperatura ali nadmorska višina?

A) Vektorski (točke, črte, poligoni)
B) Rastrski

C) Tabelarni

D) Besedilni

Pravilni odgovor: (B) Rastrski podatki

Vprašanje 4: Katera je ključna prednost uporabe formata GeoJSON v primerjavi s formatom ESRI Shapefile?

A) Podpira večje datoteke kot kateri koli drug format

B) Geometrijo in atribute hrani v eni sami človeku berljivi datoteki

C) Hitrejši je pri rastrski analizi

D) Obvezen je za vse GIS-aplikacije

Pravilni odgovor: B) Geometrijo in atribute hrani v eni sami človeku berljivi datoteki

---

Vprašanje 5: Zakaj so metapodatki pri okoljskih podatkovnih nizih še posebej pomembni?

A) Zmanjšajo velikost datoteke

B) Pospešijo vizualizacijo

C) Opisujejo izvor podatkov, merske enote ter prostorsko in časovno referenco

D) Odpravijo potrebo po analizi podatkov

Pravilni odgovor: C) Opisujejo izvor podatkov, merske enote ter prostorsko in časovno referenco

---

Praktične naloge (4–5 vaj)

Ocenjevanje študentov temelji na uspešni izvedbi naslednjih praktičnih vaj z uporabo QGIS-a in odprtih okoljskih podatkovnih nizov.

1. **Uvoz in pregled podatkov: V QGIS naložite podatkovni niz CSV, ki vsebuje okoljske meritve, določite pravilen koordinatni referenčni sistem (CRS) ter preglejte atributna polja in metapodatke.**
2. **Obravnava vektorskih in rastrskih podatkov: Naložite vektorske podatke o pokrovnosti tal in rastrske satelitske posnetke, preverite prostorsko poravnavo ter uporabite ustrezno simbologijo.**
3. **Integracija podatkov: Prekrijte točkovne, vektorske in rastrske podatkovne nize ter interpretirajte prostorske odnose med okoljskimi spremenljivkami.**
4. **Pretvorba podatkovnih formatov**
   Vektorske podatke iz formata Shapefile pretvorite v GeoJSON in rastrske podatke izvozite v drug podprt format, pri čemer ohranite informacije o prostorski referenci.
5. **Vizualizacija in interpretacija (izbirno / dodatne točke)**
   Ustvarite preprosto postavitev karte z legendo in naslovom ter na kratko interpretirajte opažene okoljske vzorce.

Viri / dodatno branje

- Davis, Jerry. *Introduction to Environmental Data Science.* Bookdown.org, 2022. Odprti dostop. https://bookdown.org/igisc/EnvDataSci.
- Earth-Env-Data-Science Project. *Earth & Environmental Data Science – An Introduction.* 2023. Odprto izobraževalno gradivo. https://earth-env-data-science.github.io/intro.html.
- Lovelace, Robin, Jakub Nowosad in Jannes Muenchow. *Geocomputation with R.* CRC Press, 2019. Izdaja v odprtem dostopu. https://r.geocompx.org/.
- Rey, Sergio J. in Levi J. Wolf. Geo-data science with Python. Geographicdata.science, 2022. Odprti učbenik. https://geographicdata.science/book/intro.html.
- Bolstad, Paul, in Michael Shin. *Essentials of Geographic Information Systems.* 2. izd., University of Minnesota Libraries Publishing, 2018. Open Textbook Library, https://open.umn.edu/opentextbooks/textbooks/67.