



Univerza
v Ljubljani
Fakulteta
*za gradbeništvo
in geodezijo*

**V2-2295 - Razvoj usmeritev za izboljšanje semantične medopravilnosti na področju
upravljanja prostorskih zbirk podatkov in geoinformatike v Sloveniji**

Marjan Čeh, Gregor Završnik, Jernej Tekavec, Anka Lisec, Robert Klinc, Barbara Trobec, Bujar Fetai



Ljubljana, januar 2025

V2-2295 - Razvoj usmeritev za izboljšanje semantične medopravilnosti na področju upravljanja prostorskih zbirk podatkov in geoinformatike v Sloveniji

Končno poročilo, januar 2025

Naročnik: Ministrstvo za okolje in prostor

Po pogodbi št. C2552-22-000093

Predstavnik naročnika: g. Tomaž Petek

Skrbnik pogodbe: mag. Matej Sotlar

Izvajalec: Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo (UL FGG)

Člani projektne skupine:

dr. Marjan Čeh, UL FGG

Gregor Završnik, UL FGG

dr. Anka Lisec, UL FGG

dr. Jernej Tekavec, UL FGG

dr. Robert Klinc, UL FGG

dr. Bujar Fetai, UL FGG

Barbara Trobec, UL FGG

Ljubljana, 27. 1. 2025

Pregled nad izvedenimi dejavnostmi in načrtovane naloge

Št.	Delovni svežnji in naloge	Načrtovano in opravljeno	Nosilec	Rezultati
DS1	Delovni sveženj 1 (DS1): Pregled semantičnih spletnih tehnologij za geoprostorske podatke			
	Naloga 1.1: Analiza mogočih kombinacij semantičnih tehnologij za učinkovito večkratno uporabo	28. 2. 2023	UL FGG, KGKN	R1.1
	Naloga 1.2: Analiza aktualnih odprtih kodnih in drugih standardov na področju geoprostorskih podatkov in programskih rešitev	28. 2. 2023	UL FGG, KGKN	R1.2
DS2	Delovni sveženj 2 (DS2): Pregled predlogov semantičnih referenčnih sistemov za geografski prostor			
	Naloga 2.1: Izdelava konceptualne zasnove tematskega semantičnega referenčnega sistema	30. 8. 2023	UL FGG, KGKN	R2.1
	Naloga 2.2: Izdelava konceptualne zasnove za ocenjevanje stopnje podobnosti podatkovnih prostorov	30. 8. 2023	UL FGG, KGKN	R2.2
	Naloga 2.3: Izdelava idejne zasnove platforme za podporo procesu pomenskega usklajevanja	30. 8. 2023	UL FGG, KGKN	R2.3
DS3	Delovni sveženj 3 (DS3): Testno pregledovanje izbranih podatkovnih shem in šifrantov			
	Naloga 3.1: Testna analiza možnosti semantične povezljivosti topografskih zbirk podatkov (občinska in državna raven)	8. 11. 2024	UL FGG, KGKN	R3.1
	Naloga 3.2: Testna analiza možnosti <i>interne</i> semantične povezljivosti nepremičninskih zbirk podatkov GURS (državna raven)	8. 11. 2023	UL FGG, KGKN	R3.2
	Naloga 3.3: Testna analiza možnosti semantične povezljivosti GJI zbirk podatkov (državne in občinske)	8. 11. 2024	UL FGG, KGKN	R3.3
	Naloga 3.4: Testna analiza možnosti semantične povezljivosti kmetijskih, gozdarskih zbirk podatkov (državne)	8. 11. 2024	UL FGG, KGKN	R3.3
DS4	Delovni sveženj 4 (DS4): Izdelava usmeritev, vodil za upravljanje procesa semantičnega povezovanja za GURS			
	Naloga 2.1: Sinteza: povzetek ugotovitev na osnovi rezultatov posameznih delovnih sklopov	20. 1. 2025	UL FGG, KGKN	R4.1
	Naloga 2.2: Usmeritve za nadaljnje korake v razvoju semantičnega povezovanja zbirk geoprostorskih podatkov v pristojnosti GURS	20. 1. 2025	UL FGG, KGKN	R4.2

KAZALO

Pregled opravljenih nalog projekta.....	15
Delovni sveženj 1 (DS1): Pregled semantičnih spletnih tehnologij za geoprostorske podatke.....	15
Delovni sveženj 2 (DS2): Pregled predlogov semantičnih referenčnih sistemov za geografski prostor	16
Delovni sveženj (DS3): Testno pregledovanje izbranih podatkovnih shem in šifrantov GURS	17
Delovni sveženj (DS4): Izdelava usmeritev, vodil za upravljanje procesa semantičnega povezovanja za GURS.....	17
1 Delovni sveženj 1 (DS1): Pregled semantičnih spletnih tehnologij za geoprostorske podatke.....	19
1.1.1 Uvod v analizo.....	19
1.1.1.1 Motivacija za analizo	20
1.1.1.2 Obravnava raziskovalnih vprašanj	20
1.1.2 Metodologija raziskave.....	21
1.1.2.1 Načrtovanje pregleda in protokola.....	21
1.1.2.2 Protokol za pregled literature.....	22
1.1.3 Rezultati uvodne sistematične raziskave	23
1.1.3.1 Rezultat iskanja s postopkom »snežna kepa« iskanja	25
1.1.4 Analize in sinteze objavljanih povezanih podatkov pri geodetskih upravah	27
1.1.5 Razprava o raziskovalnih vprašanjih	30
1.1.6 Sklepi	34
1.2 Naloga 1.1: Analiza mogočih kombinacij semantičnih tehnologij za učinkovito večkratno uporabo (R1.1).....	36
1.2.1 Uvod	36
1.2.2 Tehnološki kontekst ustvarjanja povezanih geoprostorskih podatkov	36
1.2.3 Načrtovanje in razvoj podpornih arhitektur.....	40
1.2.4 Vizija za integracijo geoprostorskih podatkov	43
1.2.5 Vrednotenje in pridobljena spoznanj	45
1.3 Naloga 1.2: Analiza aktualnih odprtih in drugih standardov (R1.2).....	48
1.3.1 »Plačljivi« ISO standardi za področje obravnave geoprostorskih metapodatkov.....	48
1.3.2 Standardi konzorcija W3C Semantic web	50
1.3.3 Standardi OGC	50
2 Delovni sveženj 2 (DS2): Pregled predlogov semantičnih referenčnih sistemov za geografski prostor.....	65
2.1 Rezultati naloge 2.1: Izdelava konceptualne zasnove tematskega semantičnega referenčnega sistema (SRS)	65
2.1.1 Prostorski objekti geografskega obsega.....	66
2.1.1.1 Koncept lokacije v ontologiji prostora.....	67

2.1.2	Pravila za modeliranje ontologije prostora geografskih razsežnosti.....	67
2.1.2.1	Opredelitev pravil semantične podobnosti.....	67
2.1.3	Opredelitev ontologije prostora geografskih razsežnosti.....	68
2.1.3.1	Določitev izhodišča ontologije prostora	69
2.1.3.2	Izdelava inducirane ravni ontologije prostora.....	69
2.1.3.3	Predpostavke modela	70
2.1.3.4	Izdelava stvarnih ravni ontologije prostora.....	71
2.2	Rezultati naloge 2.2: Izdelava konceptualne zasnove za ocenjevanje stopnje podobnosti podatkovnih prostorov.....	75
2.2.1.1	Analiza semantične ustreznosti OPGR za povezovanje.....	78
2.2.2	Semantična globina.....	78
2.2.2.1	Določitev absolutne semantične globine.....	78
2.2.2.2	Absolutna semantična globina posameznega prostorskega koncepta	80
2.2.2.3	Absolutna semantična globina zbirke prostorskih podatkov	80
2.2.3	Semantične razdalje	81
2.2.3.1	Uteži semantičnih razdalj.....	82
2.2.4	Semantična razpršenost	84
2.2.4.1	Izračun semantične razpršenosti zbirke prostorskih podatkov – poenostavljeni pristop	84
2.2.5	Analiza semantičnih razmerij aplikativnih konceptov v okviru semantičnega referenčnega omrežja.....	86
2.2.5.1	Analiza semantične ustreznosti OPGR	86
2.2.5.2	Analiza izrazov semantično nezdružljivih konceptov.....	86
2.3	Rezultati naloge 2.3: Izdelava idejne zasnove platforme za podporo procesu pomenskega usklajevanja	90
2.3.1	Postopek usklajevanja pomenske medopravilnosti.....	90
2.3.2	Elementi platforme za usklajevanje pomenske medopravilnosti.....	91
2.3.2.1	Semantični repozitorij.....	91
2.3.2.2	Modul za usklajevanje, harmonizacijo podatkov.....	92
2.3.2.3	Portal za humano usklajevanje in sodelovanje	92
3	Delovni sveženj (DS3): Testno pregledovanje izbranih podatkovnih shem in šifrantov GURS...	94
3.1	Metodologija	95
3.1.1	Postopek analize	95
3.1.2	Opredelitev podatkov za analizo	96
3.1.3	Ocena pogojev za semantično povezovanje	96
3.1.3.1	Sintaktična povezljivost:	96
3.1.3.2	Opis elementov kakovosti:.....	96
3.1.3.3	Obstoj semantičnih elementov:	96

3.1.4	Mere dostopnost semantičnih elementov.....	97
3.1.4.1	Tehnična dostopnost vira pomena (Sintaktična dostopnost).....	97
3.1.4.2	Organizacijska dostopnost vira.....	97
3.1.4.3	Stopnja uradnosti vira	98
3.1.4.4	Semantična bogatost vira	98
3.1.4.5	Starost vira	98
3.1.5	Umestitev šifrantov in njihovih vrednosti v UOPGR taksonomijo	98
3.2	Rezultati naloge 3.1: Testna analiza možnosti semantične povezljivosti Topografskih zbirk	99
3.2.1	Opredelitev podatkov za analizo topografskih zbirk	99
3.2.2	Ocena pogojev za semantično povezovanje topografskih zbirk.....	100
3.2.2.1	Sintaktična povezljivost topografskih zbirk.....	100
3.2.2.2	Dokumentacija elementov kakovosti topografskih podatkov	101
3.2.2.3	Obstoj semantičnih elementov topografskih podatkov	101
3.2.3	Mere dostopnost semantičnih elementov topografskih podatkov.....	104
3.2.4	Umestitev šifrantov in njihovih vrednosti DTK v UOPGR taksonomijo	105
3.2.4.1	DTK Stavbe	105
3.2.4.2	Umeščenost sloja DTK Ceste (linije) v semantični referenčni sistem	107
3.2.4.3	Umeščenost sloja DTK PokritostTal v semantični referenčni sistem	109
3.3	Rezultati naloge 3.2: Testna analiza možnosti interne semantične povezljivosti <i>nepremičninskih zbirk podatkov GURS</i>	111
3.3.1	Opredelitev podatkov za analizo	111
3.3.2	Ocena pogojev za semantično povezovanje nepremičninskih podatkov.....	111
3.3.2.1	Sintaktična povezljivost nepremičninskih podatkov.....	111
3.3.2.1.1	Izmenjava podatkov in formati.....	112
3.3.2.2	Dokumentacija elementov kakovosti ISK.....	113
3.3.2.3	Obstoj semantičnih elementov ISK.....	114
3.3.2.3.1	Šifranti za izmenjavo podatkov in semantični opisi šifer.....	116
3.3.2.3.2	Analiza semantične povednosti in bogatosti šifrantov	117
3.3.3	Mere dostopnost semantičnih elementov.....	120
3.3.4	Umestitev šifrantov in njihovih vrednosti v UOPGR taksonomijo	121
3.4	Rezultati naloga 3.3: Testna analiza možnosti semantične povezljivosti GJI zbirk podatkov	125
3.4.1	Opredelitev podatkov za analizo zbirk GJI	125
3.4.2	Ocena pogojev za semantično povezovanje	126
3.4.2.1	Sintaktična povezljivost	126
3.4.2.2	Dokumentacija elementov kakovosti	128
3.4.2.3	Obstoj semantičnih elementov	130

3.4.3	Mere dostopnost semantičnih elementov GJI.....	133
3.4.4	Umestitev šifrantov in njihovih vrednosti v taksonomijo UOPGR	134
3.4.4.1	Umestitev šifrantov GJI Kanalizacija (Linije in Točke)	134
3.4.4.2	GJI Ceste.....	135
3.5	Rezultati naloga 3.4: Testna analiza možnosti semantične povezljivosti kmetijskih, gozdarskih zbirk podatkov	137
3.5.1	Opredelitev podatkov za analizo kmetijskih, gozdarskih zbirk podatkov	137
3.5.1.1	Evidenca dejanske rabe kmetijskih zemljišč.....	137
3.5.1.2	Podatki iz pregledovalnika zavoda za gozdove	137
3.5.2	Ocena pogojev za semantično povezovanje	137
3.5.2.1	Sintaktična povezljivost kmetijskih, gozdarskih zbirk podatkov.....	138
3.5.2.2	Opis elementov kakovosti.....	139
3.5.2.2.1	Elementi kakovosti evidence dejanske rabe kmetijskih zemljišč	139
3.5.2.2.2	Elementi kakovosti slojev iz zbirke Zavoda za Gozdove RS	141
3.5.2.3	Obstoj semantičnih elementov	141
3.5.2.3.1	Semantični elementi evidence rabe kmetijskih zemljišč	142
3.5.2.4	Semantični elementi podatkov iz zbirke Zavoda za Gozdove RS	143
3.5.3	Mere dostopnost semantičnih elementov.....	144
3.5.3.1	Evidenca dejanske rabe kmetijskih zemljišč.....	144
3.5.3.2	Gozdarske zbirke podatkov Zavoda za gozdove.....	145
3.5.4	Umestitev šifrantov in njihovih vrednosti v UOPGR taksonomijo	146
3.5.4.1	Evidenca kmetijske rabe tal	146
3.5.4.2	Umestitev Gozdarske zbirke podatkov Zavoda za gozdove v UOPGR.....	148
3.5.4.2.1	Gozdni Odseki	148
3.5.4.2.2	Gozdni Sestoji.....	151
3.5.4.2.3	Gozdne Ceste	154
3.6	Skupni povezovalni test	155
3.6.1	Opredelitev slojev za primerjavo (DTK Pokrovnost tal, MK RABA, GZS Sestoji)	155
3.6.2	Primerjava mer dostopnosti semantičnih elementov	155
3.6.3	Primerjava skozi referenčni okvir UOPGR	156
3.6.3.1	Evidentiranje skupnih konceptov.....	156
3.7	Zaključek.....	158
3.7.1	Ključne ugotovitve o semantični povezljivosti analiziranih zbirk.....	158
3.7.2	Splošna dostopnost semantičnih elementov:	158
3.7.3	Predlogi za izboljšanje semantične povezljivosti	159
4	Delovni sveženj (DS4): Izdelava usmeritev, vodil za upravljanje procesa semantičnega povezovanja za GURS.....	160
4.1	Semantična kakovost poslovnih pravil pogojuje semantično kakovost podatkov	160

4.2	Sektorski podatkovni prostori in podatkovna ekonomija.....	166
4.3	<i>Mednarodni podatkovni prostori (The International Data Spaces - IDS)</i>	166
4.4	Usmeritve za interno (nacionalno) semantično povezovanje GURS	168
4.4.1	Slovenski nacionalni interoperabilnostni okvir (NIO).....	172
4.4.2	Platforma za semantično interoperabilnost (PZSI).....	173
4.4.3	NIO Centralni besednjak javne uprave.....	173
4.4.4	Metodološka navodila za pripravo izdelka medopravilnosti in vnos v Portal NIO	174
4.4.5	NIO Terminološke smernice za javno upravo in terminološki dogovori (2023).....	176
4.4.6	Seznam referenčnih terminoloških virov NIO za javno upravo RS	178
4.4.7	OPSI nacionalni portal in stičišče odprtih podatkov Slovenije	178
4.5	Usmeritve za zunanje (mednarodno) semantično povezovanje GURS	179
4.5.1	Evropska strategija za podatke (Evropska komisija, 2020) in podatkovni prostori	180
4.5.2	Povezovanje podatkov za zemljiško administracijo	180
4.5.3	INSPIRE skupni slovar konceptov	181
4.6	Ugotovitve glede pomensko zasnovane generalizacije in povečane večkratne uporabe ...	181
4.6.1	Orodja za semantično obravnavo geoprostorskih podatkov	186
4.7	Usmeritve za nadaljnje raziskave in razvoj na področju semantične paradigme	186
5	LITERATURA.....	187
6	PRILOGE	192
6.1	Vprašalnik za deležnike semantične analize	192

Uvod

Pogled na prihodnji razvoj semantične medopravilnosti kaže, da bodo odprtokodna geoprostorska programska oprema in odprti geoprostorski podatki verjetno imeli še večji vpliv v prihodnosti. Prihodnji razvoj geoinformatike bo moral vključevati semantične tehnologije za integracijo in upravljanje raznovrstnih geoprostorskih podatkov. Objaviti je treba informacije o pomenu podatkov v okviru zanesljivih kataloških storitev, da bi vrednost metapodatkov ponudili čim širši javnosti. Semantični integracijski podsistem mora nuditi temelje za shranjevanje, zagotavljanje kakovosti in uporabo pomensko obogatenih podatkovnih nizov. Toda šele z vzpostavitvijo operativne infrastrukture za prostorske informacije v Evropski uniji je postalo mogoče in treba vzpostaviti zavezujoč in do dobaviteljev nevtralen referenčni model na organizacijski in konceptualni ravni v Evropski skupnosti. Za razvoj nacionalne semantične medopravilnosti za interoperabilne klicne storitve prostorskih podatkov je treba pripraviti usmeritve in vodila za razvoj nacionalne semantične medopravilnosti geoprostorskih podatkov

Prostorski podatki, ki jih zbira in posodablja Geodetska uprava RS, se množično uporabljajo v javnem in zasebnem sektorju. Državljeni, državna in lokalna uprava in gospodarski deležniki si želijo, da so podatki poleg tehnične medopravilnosti tudi pomensko ustrezno povezljivi z uporabniškimi podatki širše domene gospodarstva, okolja in socialnega prostora.

Heterogenost virov podatkov, interakcije med različnimi domenami, nepopolnost podatkov/procesov in velike količine podatkov so pogoste lastnosti geoprostorske domene, kar predstavlja problem. Nizka stopnje interne semantične usklajenosti, uporaba neenotne in nestandardizirane terminologije, neskladen pretok podatkov med nacionalnimi sektorji, nizka stopnja mnogokratne rabe geoprostorskih podatkov in nizka stopnja čez mejne povezljivosti odprtih geoprostorskih podatkovnih zbirk zavirajo razvoj digitalne ekonomije enotnega trga EU, kar je še posebej aktualen problem za trg nepremičnin in trg gospodarske javne infrastrukture (GJI) na področju energetike, telekomunikacij in podobno. Vzroki za to so predvsem (Tričkovič – Rifelj, 2019):

- Pomanjkljivo medinstitucionalno sodelovanje pri pripravi predpisov.
- Nepoznavanje širine in povezanosti področij, ki jih pripravljavci predpisov urejajo. Pretirano opiranje na tuje jezike, zlasti angleščino.
- Prehitro spreminjanje rabe že ustaljenega izrazja (nove izraze je smiselno uvajati in uporabljati zadržano in praviloma le za nove pojme in pojave).
- Odsotnost splošno sprejetega terminološkega standarda, ki bi ga pripravljavci zakonskih besedil lahko uporabili.
- Nepoznavanje metodologij in standardov, potrebnih za informatizacijo procesov.

Cilji naloge so tako vezani na mednarodna izhodišča ter izhodišča EU za razvoj usmeritev za doseganje semantične medopravilnosti na področju upravljanja prostorskih zbirk podatkov in tehnologij za geoinformatiko v Sloveniji, z navezavo usmeritev na iniciativo »Zeleni podatki za vse« (GreenData4All) in implementacijo direktive INSPIRE z njeno navezavo na Evropski zeleni/digitalni dogovor.

Cilj naloge je tudi analizirati nekaj vzorcev podatkovnih shem v interni pristojnosti GURS in nekaj vzorcev zunanjih deležnikov geoprostorske infrastrukture v Sloveniji in na osnovi analize ugotoviti glavne ovire za semantično medopravilnost na operativni ravni.

Dodatni cilj je priprava operativnih usmeritev za doseganje več uporabnosti geoprostorskih podatkov za izboljšanje digitalne ekonomije Slovenije in enotnega trga EU, izboljšanje učinkovitosti in pretok podatkov med nacionalnimi sektorji na področju učinkovitosti trga nepremičnin in trga gospodarske javne infrastrukture (GJI) v Sloveniji.

Prav tako so pomembni cilji ozaveščanje strokovne in laične javnosti o pomenu pomenske (semantične) usklajenosti zbirk prostorskih podatkov v Sloveniji in v okviru EU in njihovo sodelovanje pri ontološkem usklajevanju.

Rešitve se kažejo v hibridni izvedbi načel povezanih podatkov, pri kateri so metapodatki in načini predstavitve podatkov bodisi implementirani, kot deli oblaka povezanih podatkov ali kot listi grafa, ki še vedno zahteva odjemalce, specifične za domeno, za iskanje in predstavitev.

Treba je boljše usklajevati poslovne in tehnične vidike povezanih podatkov. Treba je poiskati tehnično preprosto, poceni in rešitev z malo dodatnega kodiranja (z uporabo obstoječih programskih rešitev) za implementacijo za objavo povezanih podatkov. Z uporabo obstoječih orodij v procesu objavljajanja povezanih podatkov dopolnjuje cilj ponovno uporabo obstoječih programskih rešitev znotraj organizacij, ki temeljijo na splošnih arhitekturnih načelih »dobre prakse«.

Vsebina projekta je vezana na ključne elemente semantične medopravilnosti, iskanje optimalne kombinacije iz obsežnega nabora najnovejših tehnologij semantičnega spleta, semantičnih referenčnih sistemov, testiranja primernosti geoprostorskih baz za semantično povezovanje in usmerjanja nadaljnjega razvoja semantične medopravilnosti pri Geodetski upravi RS. Osnovni cilj je opredelitev optimalne kombinacije orodij za semantične obdelave geoprostorskih podatkov in oblikovanje usmeritev za prihodnji razvoj semantične medopravilnosti pri Geodetski upravi RS in vodil za upravljanje procesa semantičnega povezovanja geoprostorskih podatkov na področju geodetske geoinformatike. Izvedba raziskovalnega projekta je organizirana v štiri delovne svežnje:

- **Delovni sveženj 1 (DS1):** Pregled semantičnih spletnih tehnologij za geoprostorske podatke
- **Delovni sveženj 2 (DS2):** Pregled semantičnih referenčnih sistemov za geografski prostor
- **Delovni sveženj 3 (DS3):** Testno pregledovanje izbranih podatkovnih shem in šifrantov
- **Delovni sveženj 4 (DS4):** Izdelava usmeritev za proces semantičnega povezovanja GURS

V delovnem svežnju DS1 smo preiskali obstoječe aplikacije semantičnih spletnih tehnologij in ocenili njihovo uporabnost za semantično povezovanje geoprostorskih podatkov. Analizirali smo mogoče kombinacije semantičnih spletnih tehnologij za potencialno uporabo na slovenskih okoliščinah geodetskih baz podatkov. V okviru tega delovnega svežnja smo izvedli tudi analizo aktualnih odprtokodnih in drugih standardov na področju geoprostorskih podatkov, ki bi lahko predstavljali tudi osnovo za razvoj geoprostorske medopravilnosti v Sloveniji.

V delovnem svežnju DS2 smo pregledali različne vrste predlaganih referenčnih sistemov. V nalogi 2.1 smo izdelali konceptualno zasnovo tematskega semantičnega referenčnega sistema v slovenskem jeziku. V nalogi 2.2 smo pregledali nabor semantičnih relacij za tematske referenčne sisteme geografskega prostora in opredelili ožji nabor uporabnih relacij. V nalogi 2.3 smo izdelali idejno zasnovo za sodelovanje med deležniki pri usklajevanju geoprostorske semantične medopravilnosti v geodetski in sorodnih strokah v Sloveniji.

Povezovanje zbirk geoprostorskih podatkov je zapleten postopek, ki zahteva predhodno testno pregledovanje in analizo ustreznosti podatkovnih shem in šifrantov za pomensko povezovanje.

V delovnem svežnju DS3 smo testno pregledali omenjen nabor, vzorec podatkovnih shem baz podatkov pri Geodetski upravi RS (interni) ter vzorec podatkovnih shem pri zunanjih deležnikih kot je na primer kmetisko/gozdarska domena. Prav tako smo pregledovanje obravnavali na dveh ravneh razsežnosti; za vzorec iz zbirke geoprostorskih podatkov na občinski ravni, za primer občinskega komunalnega podjetja in zbirko podatkov državne ravni za primer vzorca podatkov gospodarske javne infrastrukture. Analizo povezljivosti smo izvajali za štiri ločene skupine podatkov in v skupnem povezovalnem testu.

V delovnem svežnju DS4 smo ocenili potencial in izdelali usmeritve za začetek internega semantičnega povezovanja GURS, za zunanje semantično povezovanje GURS in izvedena je bila analiza dodatnih možnosti uporabe ter izkoriščanja tehnologije semantičnega povezovanja, kot je na primer pomensko zasnovana generalizacija in analiz več-kratnosti uporabe procesov in podatkov.

V tem zaključnem delovnem svežnju smo v procesu sinteze povzeli ugotovitve na osnovi rezultatov posameznih delovnih sklopov in pripravili usmeritve za nadaljnje korake v razvoju semantičnega povezovanja zbirk geoprostorskih podatkov v pristojnosti GURS interno ter seveda tudi z bazami drugih deležnikov geoprostorske infrastrukture

Pregled opravljenih nalog projekta

Vsebina projekta je vezana na ključne elemente semantične medopravilnosti, iskanje optimalne kombinacije iz obsežnega nabora najnovejših tehnologij semantičnega spleta, semantičnih referenčnih sistemov, testiranja primernosti geoprostorskih baz za semantično povezovanje in usmerjanja nadaljnjega razvoja semantične medopravilnosti pri Geodetski upravi RS. Osnovni cilj je opredelitev optimalne kombinacije orodij za semantične obdelave geoprostorskih podatkov in oblikovanje usmeritev za prihodnji razvoj semantične medopravilnosti pri Geodetski upravi RS in vodil za upravljanje procesa semantičnega povezovanja geoprostorskih podatkov na področju geodetske geoinformatike. Izvedba raziskovalne projekta je organizirana v štiri delovne svežnje in sicer:

Delovni sveženj 1 (DS1): Pregled semantičnih spletnih tehnologij za geoprostorske podatke

Delovni sveženj 2 (DS2): Pregled semantičnih referenčnih sistemov za geografski prostor

Delovni sveženj 3 (DS3): Testno pregledovanje izbranih podatkovnih shem in šifrantov GURS

Delovni sveženj 4 (DS4): Izdelava usmeritev, vodil za upravljanje procesa semantičnega povezovanja

Delovni sveženj 1 (DS1): Pregled semantičnih spletnih tehnologij za geoprostorske podatke

V delovnem svežnju DS1 smo preiskali obstoječe aplikacije semantičnih spletnih tehnologij in ocenili njihovo uporabnost za semantično povezovanje geoprostorskih podatkov. Analizirali smo mogoče kombinacije semantičnih spletnih tehnologij za potencialno uporabo na slovenskih okoliščinah geodetskih baz podatkov. V okviru tega delovnega svežnja smo izvedli tudi analizo aktualnih odprtokodnih in drugih standardov na področju geoprostorskih podatkov, ki bi lahko predstavljali tudi osnovo za razvoj geoprostorske medopravilnosti v Sloveniji.

Naloga 1.1: Analiza mogočih kombinacij semantičnih tehnologij za učinkovito večkratno uporabo

Razpoložljivost in stalen razvoj semantičnih tehnologij povezanih podatkov sta omogočila uvedbo posodobljenega in razširljivega pristopa k objavi povezanih podatkov, vendar ostaja odprto vprašanje, katero od tehnologij in v kakšnem zaporedju jih je smiselno uporabiti. Izdelali smo pregled stanja semantične medopravilnosti za geoprostorske podatke pri Geodetskih upravah v EU.

V tej nalogi smo z analizo mogočih kombinacij semantičnih tehnologij iskali primeren pristop za zagotavljanje učinkovite in večkratne uporabe kombinacije nad različnimi zbirkami geoprostorskih podatkov. V nalogi smo iskali pristop, ki uporablja načelo »konfiguracije nad kodo« (configuration-over-code), za zagotavljanje, da se isto zaporedje funkcionalnih enot (računalniški cevovod-pipeline) uporablja za vse povezane projekte objave podatkov, pri čemer se komponente konfigurirajo samo, kjer je to treba. Ta vidik arhitekture uporabljenih tehnologij bo tisti, ki bo omogočil, da se zaporedje ponovi na več nizih podatkov, ki jih kombinacija semantičnih tehnologij prenaša od vira do objave v kratkem časovnem okviru.

Naloga 1.2: Analiza aktualnih odprtokodnih in drugih standardov na področju geoprostorskih podatkov in programskih rešitev

Standardi so bistveni za odprte geoprostorske podatke in številne odprtokodne geoprostorske programske rešitve izvajajo odprte standarde. Odprti standardi so tehnološko nevtralne specifikacije za strojno opremo, programsko opremo ali podatke, razvite z odprtim postopkom. Odprte standarde za geografske informacije (trenutno jih je bilo objavljenih več kot 100) tradicionalno razvijajo tri ključne mednarodne organizacije za razvoj standardov, katerih standarde smo proučili za potencialno uporabo v končnih usmeritvah te raziskovalne naloge. V tej nalogi smo pregledali trenutno stanje odprtokodnih standardov geoprostorske programske opreme in podatkov. Odprtost je spremenila način zbiranja, obdelave, analize in vizualizacije geoprostorskih podatkov.

Delovni sveženj 2 (DS2): Pregled predlogov semantičnih referenčnih sistemov za geografski prostor

V delovnem svežnju DS2 smo pregledali različne vrste predlaganih referenčnih sistemov. V nalogi 2.1 smo izdelali konceptualno zasnovo tematskega semantičnega referenčnega sistema v slovenskem jeziku. V nalogi 2.2 smo pregledali nabor semantičnih relacij za tematske referenčne sisteme geografskega prostora in opredelili ožji nabor uporabnih relacij. V nalogi 2.3 smo izdelali idejno zasnovo za sodelovanje med deležniki pri usklajevanju geoprostorske semantične interoperabilnosti v geodetski in sorodnih strokah v Sloveniji.

Naloga 2.1: Izdelava konceptualne zasnove tematskega semantičnega referenčnega sistema

Interpretacijo geoprostorskih in časovnih podatkov omogočajo dobro znani položajni in časovni referenčni sistemi, kot so koordinatni sistemi in koledarji. V tej nalogi smo obravnavali referenčne sisteme za tretjo vrsto interpretacije podatkov, in sicer tematskih. Po analogiji z geoprostorskimi in časovnimi referenčnimi sistemi bi morali semantični referenčni sistemi (SRS) omogočati preoblikovanje in projiciranje tematskih podatkov ali merjenje pomenske razdalje (imenovane semantična podobnost). Tipična aplikacija za tematski semantični referenčni sistem vključujejo preoblikovanje ali primerjavo različnih konceptualizacij med informacijskimi skupnostmi in iskanje informacij na splošno.

Naloga 2.2: Izdelava konceptualne zasnove za ocenjevanje stopnje podobnosti podatkovnih prostorov

Za obravnavo semantike geoprostorskih informacij je potrebna ustrezna formalna podlaga opredeljena s semantičnimi relacijami. To je prvi pogoj za doseganje semantične interoperabilnosti, to je zmožnosti sistemov in storitev za medsebojno delovanje prek skupnega razumevanja geografskih informacij. V tej nalogi smo opredelili nabor semantičnih relacij, ki bodo pokazale uporabnost v modelu tematskega semantičnega referenčnega sistema geografskega prostora, ko bo implementiran. Predlagali smo razširitev trenutnih meril pomenske podobnosti z upoštevanjem prostorskih razmerij med različnimi geoprostorskimi koncepti.

Naloga 2.3: Izdelava idejne zasnove platforme za podporo procesu pomenskega usklajevanja

V tej nalogi smo zasnovali spletno platformo usklajevanje pomenov geoprostorskih terminov v različnih zbirkah podatkov, ki jim je osnova prostorska podatkovna infrastruktura Geodetske uprave RS. Usklajevanje je zasnovano po tematskih skupinah deležnikov geografskega prostora in izmenjava opredelitev pomenov terminov, ki so vključeni v zbirke podatkov geodetske in sorodnih strok. Na spletni strani bi bila mogoča zbrana dokumentacija usklajevanj ter razpravni blog. Prav tako smo organizirali pregled aktualne literature ter standardov na obravnavanem področju in povezave na tematske strani semantične medopravilnosti organizacij, ki so vodilne v razvoju področja geoprostorske semantike.

Delovni sveženj (DS3): Testno pregledovanje izbranih podatkovnih shem in šifrantov GURS

Povezovanje zbirk podatkov je zapleten postopek, ki zahteva predhodno testno pregledovanje in analizo ustreznosti podatkovnih shem in šifrantov za pomensko povezovanje. V tem delovnem svežnju smo testno pregledali omenjen nabor, vzorec podatkovnih shem baz podatkov pri Geodetski upravi RS (interni) ter vzorec podatkovnih shem pri zunanjih deležnikih, kot je na primer kmetijsko/gozdarska domena. Prav tako smo pregledovanje obravnavali na dveh ravneh razsežnosti; za vzorec iz zbirke podatkov na občinski ravni, za primer občinskega komunalnega podjetja in zbirko podatkov državne ravni za primer vzorca podatkov gospodarske javne infrastrukture. Analizo povezljivosti smo izvajali za štiri ločene skupine testov in v skupnem povezovalnem testu.

Naloga 3.1: Testna analiza možnosti semantične povezljivosti **topografskih zbirk** podatkov (občinska in državna raven)

V nalogi 3.1 je testna analiza obsegala vzorec podatkov topografske za vzorec podatkov topografske tematike pri GURS (državna raven) DTK. Revidirali smo obstoječe specifikacije in zasnove novih.

Naloga 3.2: Testna analiza možnosti *interne* semantične povezljivosti nepremičninskih zbirk podatkov GURS (državna raven)

V nalogi 3.2 je interna testna analiza obsegala vzorec podatkov **nepremičninskih zbirk** pri GURS, in sicer katastrskih podatkov. Revidirali smo obstoječe specifikacije in zasnove novih.

Naloga 3.3: Testna analiza možnosti semantične povezljivosti **GJI zbirk** podatkov (državne in občinske)

V nalogi 3.3 je interna testna analiza obsegala vzorec na občinski ravni, za primer občinskega komunalnega podjetja z geoprostorsko bazo podrobnejših infrastrukturnih podatkov in zbirko podatkov državne ravni za primer vzorca podatkov gospodarske javne infrastrukture pri Geodetski upravi RS. Revidirali smo obstoječe specifikacije in zasnove novih.

Naloga 3.4: Testna analiza možnosti semantične povezljivosti kmetijskih, gozdarskih zbirk podatkov (državne)

V nalogi 3.4 je izvedena testna analiza za vzorec kmetijsko/gozdarskih zbirk podatkov. Revidirali smo obstoječe specifikacije in oblikovali zasnove novih.

Delovni sveženj (DS4): Izdelava usmeritev, vodil za upravljanje procesa semantičnega povezovanja za GURS

V delovnem svežnju DS4 smo pripravili usmeritve za pričetek internega semantičnega povezovanja GURS, za zunanje semantično povezovanje GURS in analizirali dodatne možnosti uporabe ter izkoriščanja tehnologije semantičnega povezovanja.

V tem zaključnem delovnem svežnju smo v procesu sinteze povzeli ugotovitve na osnovi rezultatov posameznih delovnih sklopov in pripravili usmeritve za nadaljnje korake v razvoju semantičnega povezovanja zbirk geoprostorskih podatkov v pristojnosti GURS interno ter seveda tudi z bazami drugih deležnikov geoprostorske infrastrukture.

V okviru naloge N4.4 se je izvajala tudi diseminacija rezultatov naloge do zaključka projekta 31.1.2025:

1. O izsledkih iz prvega delovnega svežnja smo objavili znanstveni članek v strokovni reviji Geodetski vestnik z naslovom »Uvajanje tehnologije Linked Data v geodetskih upravah: sistematični pregled literature«.

2. Vsebino izsledkov smo predstavili na mednarodnem simpoziju z naslovom »Metadata Matters, What Role Does Linked (Meta) Data Play in the Geospatial World?«, ki se je odvijal 6. Septembra 2023, v Hilversumu na Nizozemskem, v skupni organizaciji EuroSDR, GeoE3 and PLDN Kadaster.

V naslednjih poglavjih vmesnega poročila, smo predstavili rezultate dela po delovnih sklopih.

1 Delovni sveženj 1 (DS1): Pregled semantičnih spletnih tehnologij za geoprostorske podatke

Pregled stanja semantične medoprevedljivosti za geoprostorske podatke pri Geodetskih upravah v EU

Ključna ugotovitev: Relativno počasno uvajanje povezanih odprtih podatkov (Open Linked Data, OLD).

Naše ugotovitve na osnovi analiz literature kažejo na relativno počasno uvajanje povezanih odprtih podatkov z objavljanjem OLD na spletu v okviru EU. Da bi ocenili razloge za razmeroma počasno sprejemanje objave geoprostorskih podatkov kot povezanih podatkov, smo izvedli sistematičen pregled literature o objavi geoprostorskih podatkov Geodetskih uprav EU na spletu, objavljenih med letoma 2014 in 2023. Izbranih in analiziranih je bilo 19 člankov z dvanajstimi identificiranimi težavami. Ta analiza ponuja pregled ugotovljenih izzivov in najboljših praks, ki bi jih bilo mogoče uporabiti za podporo boljšega širjenja povezanih podatkov v Geodetskih upravah EU.

V raziskavi smo se osredotočili na sistematični pregled literature o objavljanju geoprostorskih podatkov v tehnologiji povezanih podatkov (linked data – LD) pri geodetskih upravah evropskih držav, odgovornih za zemljiško administracijo. Analizirali smo znanstvene študije, objavljene v letih od 2014 do 2023. Odkrivali smo vzroke za razmeroma počasno uveljavljanje LD-tehnologij za objave geoprostorskih podatkov. Metodologija temelji na analizi problemov, posegov, primerjav, izidov in okoliščin (PICOC) ter iskanju s tako imenovanim »vzratnim valjenjem snežne kepe«. Postopek presoje upravičenosti za vključitev člankov v sistematični pregled literature je predstavljen v Smernicah za prednostne postavke poročanja za sistematične preglede in metaanalize (PRISMA). V naboru devetnajstih izbranih člankov smo analizirali vsebino in identificirali dvanajst ključnih področij. S frekvenčno analizo smo pridobili pregled LD objavljenih geoprostorskih tem, ki so najpogostejše objavljene pri geodetskih upravah sedmih evropskih držav. Na podlagi petih raziskovalnih vprašanj smo ugotovili, da lahko poleg sistematične podpore razvijalcem besednjakov ter ontologij na uspeh objave LD vplivajo še zaupanje med deležniki, ozaveščenost razvijalcev in poznavanje najboljših praks LD izkušenih geodetskih uprav.

1.1.1 Uvod v analizo

Nacionalne kartografske in katastrske agencije (Geodetske uprave) shranjujejo velike količine podatkov (Big Data) in posodabljaajo uradne geopodatkovne baze v arhitekturi podatkovnega skladišča kot silosih podatkov (Crompvoets et al., 2016). Nanje se mora sklicevati in povezati z nabori podatkov drugih nacionalnih uprav in agencij ali komercialnih virov. Direktiva o odprtih podatkih in ponovni uporabi podatkov javnega sektorja (2019) Evropske komisije nalaga ustvarjanje pogojev za večkratno ponovno uporabo podatkov javnega sektorja, kar zahteva spletne tehnologije za geoprostorske storitve – lokacijske storitve in distribucijo objav podatkov z omogočeno povezljivostjo.

Infrastrukture prostorskih podatkov (SDI) so izboljšale odkrivanje podatkov in dostopnost na podlagi spletnih storitev OGC (Open Geospatial Consortium). Vendar so iskanje, dostop do in uporaba podatkov, razširjanih prek SDI, še vedno težavni za nestrokovne uporabnike (Vilches-Blázquez in Saavedra, 2019). Združevanje prostorskih podatkov iz SDI z uporabo tehnologij semantičnega spleta je pomemben korak naprej (Wiemann in Bernard, 2016).

Kot ugotavljata Crompvoets in Vancauwenberghe (2018), je večina evropskih držav od leta 1990 ustanovila strateške svetovalne odbore za opredelitev strategije SDI in sistemov za nadzor izvajanja SDI.

Povezani podatki so ključnega pomena v prostorski podatkovni infrastrukturi (SDI) (Huang et al., 2019), saj interoperabilnost omogoča programsko povezavo podatkov in informacij iz različnih virov, ki se lahko uporabijo za sprejemanje boljših odločitev in izboljšanje učinkovitosti različnih operacij. Trenutni SDI se soočajo s problemom integracije podatkov in semantične heterogenosti zaradi svoje delno ločene

organizacije podatkov. Izdelava uradnih geoprostorskih podatkov je pogosto v rokah visoko specializiranih javnih agencij, ki so tradicionalno sledile njihovi poti in vzpostavile proizvodne okvire (Ariza-López et al., 2021). Ti podatki so izolirani in jih ni mogoče deliti in integrirati med organizacijami in skupnostmi (Vilches-Blázquez & Saavedra, 2019).

Proces integracije lahko vključuje različne tehnike priprave podatkov, vključno z izbiro podatkov, čiščenjem, transformacijo in dostavo (extract, transform, load – ETL) (Folmer et al., 2019; Mallek et al., 2020). Geoportali so vidni deli SDI-jev, ki se osredotočajo na interoperabilnost z izvajanjem standardov, kot so (Open Geospatial Consortium) spletne storitve OGC za iskanje in uporabo geografskih podatkov in storitev (Owusu-Banahene, 2018).

Odrpta standardizacija Ma (2017) za spletne storitve na področju povezovanja geoprostorskih podatkov igra pomembno aplikacijsko vlogo pri implementaciji interoperabilnosti z uporabo principov povezanih podatkov in spletne objave podatkov. Interoperabilnost geoprostorskih podatkov se nanaša na zmožnost usklajevanja različnih prostorskih informacijskih sistemov, programske opreme in formatov prostorskih podatkov (Lanucara et al., 2019).

Na različnih dogodkih, kot sta konferenca ENDORSE o referenčnih podatkih in semantiki in konferenca o formalnih ontologijah v informacijskih sistemih, če omenimo samo dva, strokovnjaki razpravljajo tudi o ovirah za prevzem LD.

1.1.1.1 Motivacija za analizo

Ta študija je bila motivirana z željo po predstavitvi strukture in obsega raziskav na presečišču med semantičnimi odprto povezanimi podatki in geoprostorskimi podatki v Geodetskih upravah, spodbujeno s pričujočim projektom. Raziskovalni izziv je bil tudi določiti trenutno stanje tehnike in prepoznati znanstvenike, ki veljajo za pionirje v spletni publikaciji geoprostorskih podatkov, ki svoje raziskave osredotočajo na specifične težave Geodetskih uprav. Takšni avtorji so lahko ključni za širjenje znanja in napredek raziskovalnega področja. Nazadnje, glavno gonilo za izvedbo pregleda literature je počasno širjenje LD in neobstoj takih študij na področju Geodetske uprave. Zanima nas razumevanje ovir za sprejem LD na splošno (Attard et al., 2017; Polleres et al., 2020). Prispevek naj bi nam in strokovnjakom, ki sodelujejo pri strateškem in operativnem razvoju Geodetske uprave, omogočil vpogled v trenutno stanje objavljanja geoprostorskih podatkov kot povezanih podatkov. Zanimajo nas nedavna širitev, vprašanja in izzivi LD, objavljeni v člankih v revijah in znanstvenih poročilih, ki so na voljo v obliki odprte literature v zadnjem desetletju.

Nove tehnologije niso preproste za implementacijo, kar kažejo ocene v obliki, imenovani Gartner Group's hype cycle (O'Leary, 2008). »Pametni svet« (Smart World) je ena od štirih nedavnih najvplivnejših Gartnerjevih (Nguyen, Jump in Casey, 2022) tem. Nastajajoče tehnologije z največjim vplivom so pametni prostori, metaverzumi in digitalni dvojčki, ki so vse tudi v domeni zanimanja Geodetske uprave. Tema kritičnih omogočevalcev vključuje grafe znanja, kjer povezani podatki (LD) sodelujejo kot spodbujevalna tehnologija. Ta sistematičen pregled zbira in komentira objavljeno literaturo na področju povezovanja geoprostorskih podatkov za LD, da bi ugotovili trenutni status uporabe LD v geoprostorskih upravah (Geodetske uprave). Postavljamo vprašanje, s kakšnimi izzivi se soočajo tiste uprave, ki so že začele uporabljati semantično tehnologijo LD kot odprto povezane podatke (OLD). Zanimajo nas tudi razlogi za razmeroma počasno uveljavitev povezovalnih geoprostorskih tehnologij. Za analizo razpoložljive literature smo uporabili metodo sistematičnega pregleda literature, ki je podrobneje opisana v 2. razdelku tega sistematičnega pregleda literature. Več prejšnjih študij ponuja pregled literature na področju povezanih podatkov. S tem pregledom poskušamo zapolniti vrzel pri pregledu literature, zlasti geoprostorskih podatkov, ki jih objavlja Geodetska uprava.

1.1.1.2 Obravnava raziskovalnih vprašanj

V zadnjem desetletju je koncept povezanih podatkov pokazal močno tendenco širjenja v različne geoprostorske domene, vključno z Geodetskimi upravami. Več avtorjev je poročalo o zgodnji uspešni implementaciji geoprostorskih podatkov v obliki semantičnega spleta, kot so Goodwin, Dolbear in Hart

(2008), Hyvönen et al. (2014), Saavedra, Vilches-Blázquez in Boada (2014), Debruyne et al. (2016) Beek in Folmer (2017).

Trenutno razumevanje LD še ne ponuja dokazov o močnem pozitivnem vplivu ali izboljšanju gospodarskih rezultatov zaradi morebitnih spremenjenih poslovnih modelov Geodetskih uprav. Iz začetnih analiz razpoložljive literature se zdi, da obstajajo nekatere omejitve za širjenje obstoječega znanja o povezanih podatkih po izumu svetovnega spleta (Berners-Lee, 1996). Namen naše raziskave je identificirati glavne izzive in multidisciplinarna vprašanja, o katerih poroča literatura v zvezi z izvajanjem LD v Geodetskih upravah, in odkriti, kaj je treba storiti za povečanje njegove uporabe. Pregled temelji na naslednjih raziskovalnih vprašanjih:

RQ1: Katere so težave, ki jih študije obravnavajo pri objavljanju geoprostorskih podatkov kot LD?

RQ2: Katere rešitve in zahteve prepoznajo geodetske uprave z izkušnjami z objavljanjem LD?

RQ3: Katere prednostne podatkovne teme objavljajo geodetske uprave iz aktivnih okrajev LD?

RQ4: Katere izzive/težave so geodetske uprave ugotovile za prihodnji razvoj povezanih podatkov?

RQ5: Kakšni so razlogi za razmeroma počasno sprejemanje/razširjanje geoprostorskih podatkov LD s strani geodetskih uprav?

Sistematični pregled literature je organiziran na naslednji način. V razdelku o metodologiji razložimo fazo načrtovanja sistematičnega pregleda literature (osnutek protokola) in nato opišemo, kako je bila študija izvedena. Na koncu poročamo o rezultatih, razpravljamo o sintezi ugotovitev in odgovarjamo na raziskovalna vprašanja. Za zaključek povzemamo glavne ugotovitve.

1.1.2 Metodologija raziskave

Raziskava sledi metodologiji sistematičnega pregleda literature. Sistematični pregled (SR) je metoda, ki se uporablja za iskanje in povzemanje vseh ustreznih obstoječih dokazov, povezanih z določenim raziskovalnim vprašanjem, ki vas zanima. Ena od dejavnosti, povezanih s procesom SR, je izbor primarnih študij (Octaviano et al., 2015). V predhodni preiskavi smo opravili pregled literature o tehnologijah semantičnega spleta za prostorske podatke. Sistematično smo pregledali dokumentirane raziskovalne članke, ki opisujejo že implementirane semantične spletne povezane podatkovne aplikacije (LD). Ciljna domena je bila omejena na kontekst prostorskih podatkov in nacionalnih kartografskih in katastrskih agencij (NMCA).

1.1.2.1 Načrtovanje pregleda in protokola

Sistematične preglede je treba izvajati v skladu z vnaprej določeno iskalno strategijo. Splošni proces SR je razdeljen na tri faze: načrtovanje, delo in poročanje o pregledu. Faze, povezane z načrtovanjem pregleda, so: 1. Ugotovite potrebo po pregledu, 2. Pripravite protokol pregleda in 3. Izvedite SR. Koraki, povezani z izvedbo pregleda, so: 1. Identifikacija raziskave, 2. Izberite primarne študije, 3. Ocenite kakovost študije, 4. Izvlecite in spremljajte podatke, 5. Sintetizirajte podatke. Poročanje o pregledu je en sam korak (Kitchenham, 2004). V programskem inženiringu je mogoče raziskovalna vprašanja prilagoditi vprašanjem programskega inženiringa, kot je ocenjevanje učinka tehnologije programskega inženiringa, ocenjevanje pogostosti ali stopnje sprejemanja tehnologije, prepoznavanje stroškov in dejavnikov tveganja, povezanih s tehnologijo, prepoznavanje vpliva tehnologije o zanesljivosti, ponovni uporabi, modelih zmogljivosti in stroškov, analizi stroškov in koristi programskih tehnologij (Person et al., 2021).

Raziskovalna sinteza je skupni izraz za družino metod za povzemanje, povezovanje, združevanje in primerjavo rezultatov različnih študij o temi ali raziskovalnem vprašanju. V programskem inženirstvu je primerjava in primerjava takšnih dokazov nujna za sklepanje o empirični podpori pojava. Takšna sinteza lahko tudi osvetli pomembna področja in vprašanja, ki v prejšnjih empiričnih raziskavah niso bila ustrezno obravnavana. Raziskovalna sinteza je torej način razumevanja trditev več študij (Cruzes in Dybå, 2010).

1.1.2.2 Protokol za pregled literature

Izbor tega pregleda literature temelji na dokazih (študijah). Sledili smo raziskovalnemu pristopu PICOC (Richardson et al., 1995) in oblikovali raziskovalna merila o populacijskem problemu, posegu (intervenciji), primerjavi, rezultatih in okoliščinah (kontekst). Prvo vprašanje naj obravnava problem v določeni populaciji, vprašanje o intervenciji pa olajša iskanje natančne rešitve. Primerjalno vprašanje podaja merila za primerjavo rezultatov intervencije s stanjem prebivalstva pred intervencijo. Rezultati kažejo na učinke intervencije na probleme prebivalstva. Okoliščine opisujejo kontekst populacije s težavo (Booth et al., 2019).

Za izbor geoprostorskih raziskovalnih študij LD smo podali postavke kriterijev, ki ustrezajo vidikom pristopa PICOC:

1. Problem: Zaprti geoprostorski podatki
2. Poseg (intervencija): Objava geoprostorskih podatkov kot povezanih podatkov z uporabo semantičnih spletnih tehnologij
3. Primerjava: spletno iskanje, povezljivost
4. Rezultati: strojno berljivi, objavljeni geopodatki kot odprto povezani podatki
5. Okoliščine: Nacionalne geodetske in katastrske agencije (NMCA)

Ključne besede in sinonimi: geoprostor, "povezani podatki", publikacija, "semantični splet", "National Mapping and Cadastral Agency", NMCA

Baze podatkov: SCOPUS, IEEE, Google Scholar

Merila za izbor študija:

- Vključitev publikacij: članki v revijah, referati na konferencah, znanstvena poročila;
- Izločitev objav: dvojnikov, dokumentov, ki ne sodijo v sistem upravljanja z zemljišči;
- Jezikovne omejitve: angleški jezik;
- Časovna omejitev: leta izida od 2014 do 2023.

Preglednica 1 Vprašanja o upravičenosti pred ekstrakcijo podrobnih podatkov študij:

C1	Ali je študija pomembna in bistvena za naše raziskovalno vprašanje?
C2	Ali se študija osredotoča na geoprostorske podatke?
C3	Ali študija jasno opredeljuje problematiko objave LD?
C4	Ali študija pojasnjuje izzive objave LD?
C5	Ali študija pojasnjuje najboljše prakse objavljanja LD?
C6	Ali študija zajema področja NMCA?

Rezultati izbire in ocenjevanja študij za primernost po merilih (Preglednica 1) za podrobno pridobivanje informacij so predstavljeni v Preglednica 2.

Pridobivanje podatkov

Za povzetek informacij o študijah, izbranih za ekstrakcijo, smo ustvarili Preglednica 6 za zbiranje in sintezo informacij, opredeljenih z raziskovalnimi merili PICOC. Zbrali smo podatke o značilnostih

študij, ki jih preiskujemo tako, da smo prednostno razvrstili področje geoprostorskih podatkov (teme), geografsko pokritost in izkušnje držav, vprašanja, obravnavana v študijah, in prihodnje izzive razvoja povezanih podatkov LD v geodetskih upravah.

Raziskovalna vprašanja

Bistvena sestavina protokola je izbor raziskovalnih vprašanj, povezanih z motivacijo za raziskovanje.

Omejitve tega pregleda literature: V tej raziskavi se ne ukvarjamo s pravnimi vprašanji LD, kot so odgovornost - zasebnost - intelektualna lastnina - konkurenčno pravo.

1.1.3 Rezultati uvodne sistematične raziskave

Sistematični pregled poskuša primerjati empirične dokaze iz relativno majhnega števila študij o osredotočenem raziskovalnem vprašanju (Higgins et al., 2022). Za predstavitev postopka izbire majhnega števila študij, ki bodo vključene v pregled, smo uporabili predlogo diagrama poteka, povzeto iz smernice Priporočene postavke poročanja za sistematične preglede in metaanalize PRISMA (Page et al., 2021).

Sprva smo identificirali 133 zapisov iz treh izvornih baz podatkov: Scopus (61 zapisov, 46 %), IEEE (2 zapisa, 2 %) in Google Scholar (70 zapisov, 53 %). Poleg tega je bilo ročno dodanih 13 študij (skupaj 146 zapisov). Pred pregledom naslovov in ključnih besed je bilo odstranjenih 14 dvojniki (9 %). Pri pregledovanju naslovov in ključnih besed smo odstranili 26 (18 %) zapisov, ker niso bili povezani z našim raziskovalnim problemom.

Sto šest zapisov (prispevkov) je bilo pregledanih na podlagi povzetka (73 % od široko opredeljenih 146 zapisov). Triinšestdeset zapisov (43 %) je bilo izključenih, ker so bili nepomembni v okviru sistemov zemljiške administracije (LAS). Pregledali smo (z branjem uvoda in zaključkov prispevkov) 43 študij (29 %) glede primernosti. Tri merila primernosti so izključila štiriindvajset (16 %) zapisov: ne opisujejo domene geodetskih uprav (NMCA) 12 (8 %), ne obravnavajo kritičnih geoprostorskih podatkovnih nizov 7 (5 %) in niso pomembni za lokacijski kontekst EU 5 (3 %). Za branje celotnega besedila, ekstrakcijo informacij, analizo in sintezo je bilo izbranih 19 študij (13 % od skupno 146 prvotno identificiranih zapisov).

V procesu skrbne selekcije smo v zvezi s 43 izbranimi študijami ocenili primernost vsake s šestimi kriteriji na kratki lestvici kakovosti (Ne = 0, Delno = 0,5, Da = 1) (Preglednica 1). Za ekstrakcijo smo izbrali študije, ki so dosegle intervalno oceno [3 – 6].

Izbor študij in poročil za branje celotnega besedila in pridobivanje ključnih informacij (19 člankov), razvrščenih po letu objave, je predstavljen v Preglednica 2 Druge študije so bile izključene iz pregleda.

Preglednica 2 Podroben izbor študija in merila za oceno kakovosti

		C1	C2	C3	C4	C5	C6		
ID	Author	Releva tnost	Geospatial	Oprede ljeni problemi	Oprede ljeni izzivi	Najboljše prakse	Domena geodetske uprave	Skupni rezultatt	%
1	Saavedra et al., 2014	0.5	1	1	1	0.5	1	4.0	67%
2	Çağdaş and Stubbkjær, 2015a	1	1	0.5	0,5	0	1	4.0	67%
3	Çağdaş and Stubbkjær, 2015b	1	1	0.5	1	0	1	4.5	75%
4	Hietanen et al., 2016	0.5	1	0.5	0,5	0	1	3.0	50%
5	Beek and Folmer, 2017	1	1	0.5	0,5	1	1	5.0	83%
6	Bucher et. al., 2020	1	1	0.5	1	0.5	1	5.0	83%
7	Debruyne et.al., 2017	1	1	1	1	1	0.5	5.5	92%
8	Kyzirakos et al., 2018	1	1	0.5	1	0	0	3.0	50%
9	van den Brink et al., 2018	1	0	1	1	1	0	4.0	67%
10	Folmer et al., 2018	1	1	0	0	1	1	4.0	67%
11	Ronzhin, 2019	1	0.5	1	1	1	1	5.5	92%
12	Forlmer et al., 2019	1	1	0.5	1	1	1	5.5	92%
13	Rowland et al., 2020	1	1	0.5	0	0.5	1	3.5	58%
14	Folmer et al., 2020,	1	1	1	1	0.5	1	5.5	92%
15	Yaman et al., 2021	1	0	1	1	0.5	0.5	4.0	67%
16	Issa et al., 2021	1	1	1	1	1	0	5.0	83%
17	Rowland et al., 2021	1	0	1	1	1	1	5.0	83%
18	Rowland et al., 2022	1	1	1	1	1	1	6.0	100%
19	Yaman et al., 2022	1	0.5	1	1	1	1	5.5	92%
		1.00	0.79	0.72	0.82	0.65	0.79	4.6	76%

Iz Preglednica 2 je mogoče sklepati, da najnovejše objave izbranih študij prihajajo od Rowlandove (3), Folmerja (3) in Yamana (2). Rezultati iz, razvrščeni po letih, potrjujejo, da novejše objave prednjačijo pri oceni kakovosti in podrobnejših kriterijih izbora. To lahko kaže na rahlo pristranskost naših izbirnih kriterijev ali na priznano potrebo po pokrivanju problematičnih vprašanj objave LD s strani avtorjev, ki se bolj trudijo zagotoviti posodobljene in zanesljive informacije o svojih rešitvah za reševanje težav na področju objave LD pri geodetskih upravah.

1.1.3.1 Rezultat iskanja s postopkom »snežna kepa« iskanja

V programskem inženirstvu in informacijskih sistemih je priporočljivo izvesti sistematičen pregled literature, nadgrajen s tehniko snežne kepe, opisano v (Wohlin et al., 2022) kot hibridno strategijo iskanja. Drugotne, sekundarne študije (reference) lahko natančno pokažejo vrzeli ali izpostavijo področja, ki potrebujejo več pozornosti raziskovalcev ali praktikov. Možno je izvajati snežno kepo naprej, tj. prepoznavanje člankov, ki so citirali članke, najdene v osnovnem iskanju, ali pa sneženje nazaj (zapisi iz referenčnih seznamov) (Jalali in Wohlin, 2012).

V tem pregledu literature smo na referenčnih seznamih prvotno identificiranih člankov za SR povezanih podatkov uporabili snežno kepo za nazaj in skupno vrnili 1706 referenčnih študij. Zanimalo nas je prvih deset revij, ki najpogosteje objavljajo o geoprostorskih LD (Preglednica 3), rezultati avtorjevega skakanja nazaj pa so povzeti v Preglednica 4.

Preglednica 3 Število objav LD na revijo s tehniko iskanja s snežno kepo

Journal	Publications
International Journal of Geographical Information Science	32
Computers & Geosciences	26
Semantic Web	18
ISPRS International Journal of geo-information	16
Environmental Modelling & Software	16
Transactions in GIS	15
Communications of the ACM	12
Computers	11
Semantic Web	9
GeoInformatica	9

International Journal of Geographical Information Science je objavil približno 20 % prispevkov, ki obravnavajo, ali vsaj omenjajo geoprostorske podatke. Kljub temu so prve tri revije na zgornjem seznamu (Preglednica 3) objavile skoraj polovico člankov na tem skrajšanem seznamu desetih revij. Identificirali smo tudi najbolj uspešne avtorje v domeni povezanih podatkov, rezultati povratne snežne kepe pa so zbrani iz začetnega iskanja, kot je predstavljeno spodaj v Preglednica 4. Nekateri avtorji so vključeni v naš končni seznam devetnajstih študij (Folmer, E. in Kyzirakos, K.). Rowland, A., kot prva avtorica najkakovostnejših študij iz Preglednica 2, ni na seznamu najplodnejših avtorjev (Preglednica 4), ker je objavila nazadnje, njeno delo pa še ni dovolj uvrščeno v evidenco referenc.

Preglednica 4 Povratna iskanja (povratne snežne kepe) so štela reference na avtorja od začetnega pregleda.

Author	Publications
Janowicz, K.	26
Lehmann, J.	23
Koubarakis, M.	23
Auer, S.	20
Kraft, M.	19
Bizer, C.	17
Corcho, O.	14
Hitzler, P.	14
Bereta, K.	14
Folmer, E.	13
Kyzirakos, K.	13

Pred fazo ekstrakcije smo izvedli tudi frekvenčne analize besed v pod izboru člankov (43). Od 25.511 konceptov, samodejno zaznanih iz ustreznih člankov, je dvajset najpogostejših besed navedenih v Preglednica 5.

Preglednica 5 Frekvenčne analize besed v primernih člankih

Word	Count	%
data	7116	27.9%
spatial	2401	9.4%
linked	1491	5.8%
geospatial	1468	5.8%
information	1403	5.5%
web	1397	5.5%
semantic	1169	4.6%
RDF	946	3.7%
org	834	3.3%
level	778	3.0%
geo	756	3.0%
quality	674	2.6%
query	672	2.6%
datasets	659	2.6%
knowledge	658	2.6%
geometry	647	2.5%
GEOSPARQL	643	2.5%
ontology	624	2.4%
geographic	596	2.3%
model	579	2.3%

Po zaporedju pričakovanih splošnih izrazov, uporabljenih v začetnem iskanju, so poleg razvrščene zbirke (Preglednica 5) visoko pogosti tehnološki koncepti LD, kot so splet (5,5 %), RDF (3,7 %), GEOSPARQL (2,5 %), in ontologija (2,4 %). Koncepta RDF in GEOSPARQL sta tukaj prikazana kot potrebna za objavo LD, vendar v tem pregledu nista izrecno upoštevana. Iz izbranih študij smo pridobili informacije za nadaljnje analize in sinteze, predstavljene v razdelku 1.3.3 te naloge

Pridobivanje podatkov

Pri sistematičnem pregledu pridobivanje podatkov zajame kritične značilnosti študij v strukturirani obliki, ki temelji na informacijah iz člankov in poročil v revijah. Najpogostejša polja za ekstrakcijo v pregledih literature so opredeljena v okviru PICOC (problem, intervencija, primerjava, izid, okoliščine) (Schmidt et al., 2021). V tem koraku sistematičnega pregleda smo razvili preglednice z dokazi, v katerih smo zbrali pomembne informacije iz vsake študije (z uporabo ogrodja PICOC, kot vodila), in preglednico s povzetkom, ki je zagotovila pregled kritičnih ugotovitev iz naših dokazov pregleda na visoki ravni. V Preglednica 6 smo prikazali rezultate izvedbe kvantitativne sinteze.

1.1.4 Analize in sinteze objavljajanja povezanih podatkov pri geodetskih upravah

Raziskovalna sinteza je skupni izraz za družino metod za povzemanje, povezovanje, združevanje in primerjavo izsledkov različnih študij o temi ali raziskovalnem vprašanju. Raziskovalna sinteza je način osmišljanja zbirke študij, npr. Cruzes in Dybå (2010). Malo pozornosti se posveča sintezi raziskav v programskem inženiringu, vendar je za zaključek empirične podpore za LD potrebna primerjava in kontrast dokazov. Ta sistematični pregled literature je izpostavil številne težave, ki jih je treba rešiti, in izzive za prihodnji razvoj objavljajanja in medsebojnega povezovanja geoprostorskih podatkov, ki jih vzdržujejo NMCS.

Za razvrstitev danih ključnih značilnosti LD, odkritih v izbranih študijah, smo uvedli pet raziskovalnih in razvojnih konceptov, ki nas vodijo skozi zadnjo fazo pregleda literature: sintezo. Predlagani koncepti, sprejeti iz okvira PICOC, so "problem", "rešitev" (intervencija), "ki jo je treba izboljšati" (primerjava), "zahteva" (cilj, rezultat) in koncept "okoliščine" (kontekst).

Preglednica 6 Zbirna tabela za sintezo podatkov, zbranih iz devetnajstih izbranih študij o vprašanjih geoprostorskega objavljajanja LD

postavka LD vprašanja	opis študija	ID	štetje	%
1	interoperabilnost v skupni rabi, dostop do odprtih podatkov, ponovna uporaba, integracija - združevanje različnih virov	3, 4, 8, 9, 12, 15, 17	7	14,3 %
2	besednjak skupni tezaver za področje terminologija, večjezična terminologija, čezmejna terminologija	2, 3, 4, 6, 7, 9, 18	6	12,2 %
3	uporabniki podpore podpirajo brskanje, iskanje, poizvedovanje, analize, pregledovanje	5, 9, 10, 12, 13, 17	6	12,2 %
4	heterogenost pomensko, strukturno in sintaktično ločena; raznolikost kontekstov, definicije izrazov	1, 6, 9, 11, 14, 19	5	10,2 %
5	ontološko znanje, semantika, relacije, razlaga pomena instanc za zunanji svet	4, 7, 9, 11, 18	5	10,2 %
6	kakovostnih metrik kakovosti objave LD kot popolnost, posodabljanje, točnost	1, 11, 16, 19	4	8,2 %
7	berljivost, čitljivost s strani ljudi in strojev	4, 6, 9, 15, 18	4	8,2 %

8 veliki podatki, objavljane velikih naborov podatkov, zaprti podatki 5, 8, 9, 14 4 8,2 %
9 učinkovitost poslovanja vprašanja OLD, avtomatizacija objav, ETL, časovno in stroškovno učinkovita dostava, omejitve 8, 14, 17, 18 4 8,2%
10 podatkovni kontekst metapodatki - informacije o podatkih 2, 4 2 4,1%
11 serializacija zagotavlja podatke v različnih formatih LD 4 1 2,0%
12 razširljivost razširljiv pristop k objavi povezanih podatkov 18 1 2,0%

Preglednica 6 obravnava vprašanja objave LD v devetnajstih analiziranih študijah. Identificirali smo, ročno izluščili in analizirali dvanajst ključnih vprašanj znanja, povezanih z raziskavami na LD, navedenih v stolpcu »težave LD«. Analize pogostosti kritičnih vprašanj so predstavljene v zadnjih dveh stolpcih Preglednica 3. Preglednica je razvrščena v padajočem vrstnem redu glede na pogostost študij, ki obravnavajo določena vprašanja LD. Pogostost ključnih zadev iz devetnajstih študij se giblje od sedmih študij (14 %) za vprašanje znanja "interoperabilnost" do ene študije, ki obravnava "razširljivost" (2 %).

V naslednjih analizah so koncepti ogrodja PICOC zapisani v oklepajih. Postavke LD so interpretirane po vrstnem redu pogostosti, kot je izračunano v Preglednica 6.

Točka 1: "Interoperabilnost"

Splošni izraz »interoperabilnost« se najpogosteje pojavlja v študijah, ki opredeljujejo tudi pojem znanja s sorodnimi pojmi, kot so deljenje, dostop do odprtih podatkov, ponovna uporaba in integracija – združevanje različnih virov.

Interoperabilnost elementov smo uvrstili med zahteve (cilj, rezultat) za proces objavljane LD, ker interoperabilnost obsega vsaj tri vrste (sintaktična, strukturna, semantična), ki morajo biti vse izpolnjene, da sistemi delujejo nemoteno in interaktivno.

Točka 2: "Slovar"

Element znanja »besedišče« (vokabular) je opredeljen tudi kot »skupni tezaver za strokovno terminologijo«, »večjezična terminologija« in »čezmejna terminologija« (avtorji študij št. 2, 7 in 18).

Je najvišje uvrščena postavka med strokovnimi izrazi na LD, kar kaže na njen bistveni pomen za uspešno objavo podatkov na spletu. Menimo, da je gradnja besedišča komponenta rešitve (posega/intervencije), ki jo je treba vzpostaviti za učinkovito obogatitev podatkov iz ločene baze podatkov.

Točka 3: "Podpora"

Postavka "podpora" je v študijah opredeljena kot pomoč uporabnikom pri brskanju, iskanju, poizvedovanju, analizi in ogledu LD. Postavka ima v predstavljeni frekvenčni klasifikaciji visoko oceno, primerljivo s postavko »besedišče«, čeprav ni tehnična. Podpora uporabnikov je lahko bistvena pri razmeroma nizki razširjenosti LD v geodetskih upravah. To beležimo z oznako »za izboljšavo« (primerjava), ker lahko LD kljub nezadostni podpori dobro deluje v raziskovalnem okolju. Vendar je treba za vključitev LD v več procesov in uporabniških aplikacij geodetskih uprav zagotoviti podporne dejavnosti in orodja ter jih izjemno izboljšati.

Točka 4: "Heterogenost"

Element znanja "heterogenost" je v študijah opisan tudi kot "pomensko, strukturno in sintaktično disjunktiven", "različnost kontekstov" in kot "heterogene definicije pojmov".

Težavo "heterogenosti" prepoznavamo kot najbolj kritičen problem, ki ga je treba obravnavati, da bi dosegli dobro interoperabilnost z LD. Kar zadeva oceno v klasifikaciji, izraz "heterogenost" ustreza pogostosti besede "ontologija", ki je naslednji izraz, ki ga je treba upoštevati.

Točka 5: "Ontologija"

Pojem znanja »ontologija« je v analiziranih študijah opisan kot »znanje«, »semantika«, »relacije« in »razlaga pomena instanc za zunanji svet«.

"Ontologijo" označujemo kot rešitev (intervencijo) problema heterogenosti v projektih publikacij LD, v kombinaciji s prej obravnavano besedno zvezo "besednjak". Oba pojma skupaj tvorita osnovo za publikacijske projekte LD.

Točka 6: "Kakovost"

Element znanja "kakovost" je v študijah opisan z metrikami kakovosti, kot so "popolnost", "pogostost posodabljanja" in "natančnost". Pojavlja se v 8 % analiziranih študij, kar je na enaki ravni kot postavke znanja "berljivost", "Veliki podatki" in "učinkovitost".

Kakovost je nedvomno splošni koncept, ki ga je treba izboljšati (primerjati) na večini področij upravljanja informacij in LD.

Točka 7: "Berljivost"

Element znanja "berljivost" je berljivost podatkov s strani ljudi in strojev.

Strojna berljivost je zahteva (rezultat) in prvi pogoj v projektih objav LD. Vendar pa berljivost povezanih podatkov ni tako očitna ali praktična za nestrokovnjake. Ta razlog bi lahko bil še ena trdna ovira za zavrto širjenje objave LD v geodetskih upravah. Zato bi lahko sklepali, da bi "berljivost" "v kombinaciji z vidikom znanja" podpore "najpomembneje vplivala na širjenje LD na geoprostorskem področju.

Točka 8: "Veliki podatki" «

Veliki, izolirani nabori podatkov so kontekst upravljanja podatkov v geodetskih upravah. Postavka znanja »Big Data« se v analiziranih študijah uporablja razmeroma tako pogosto kot prej obravnavana postavka »Berljivost«. Če pa bi besedo "berljivost" umestili v zahtevani razred, bi Big Data definirali kot okoliščinski razred raziskav.

Točka 9: "Učinkovitost"

Postavka znanja »učinkovitost« je v analiziranih študijah povezana z izrazi, kot so »poslovna vprašanja odprtega LD«, »avtomatizacija objav«, »ETL«, »časovno in stroškovno učinkovita dostava« in »omejitve«. Menimo, da je to atribut LD, ki ga je treba izboljšati (primerjava), da bi dosegli širšo uporabo v geoprostorskih podatkih geodetskih uprav. Izboljšana učinkovitost bo verjetno povečala potencial razširjanja LD z izboljšano kakovostjo in podporo.

Točka 10: "Kontekst podatkov"

Element znanja »kontekst podatkov« je v študijah pogosto opisan kot razpoložljivost metapodatkov (informacije o podatkih) na primer okoliščine nastanka podatka in sledljivost (lineage).

Razpoložljivost »podatkovnega konteksta« obravnavamo kot enega od problemov za uveljavitev paradigme LD v praksi geodetskih uprav, saj so metapodatki pogosto preslabo pripravljeni, tudi s pomočjo smernic, kot je INSPIRE.

Točka 11: "Serializacija"

Izraz znanja "serializacija" je v preučeni študiji enačen z "objavo istih podatkov v različnih formatih LD". Pojavnost elementa "serialization" je redka, serializacija v različnih formatih pa podpira različne platforme za analizo. Zato menimo, da je "serializacija" pogoj (rezultat) za razširjanje LD v geodetskih upravah.

Točka 12: "Razširljivost"

Vidik "razširljivosti" znanja se nanaša na sposobnost infrastrukture in računalniškega sistema LD, da poveča zmogljivost kot odgovor na povpraševanje. Ker je ta element bolj povezan s komponentami strojne opreme sistema SDI, je pogostost tega elementa nizka. Kljub temu bi morali to točko resno

obravnavati kot zahtevo (rezultat), saj lahko razširljivost pristopa objavljavanja LD tudi bistveno izboljša učinkovitost objavljavanja LD.

Iz zgornje sinteze dvanajstih točk znanja, pridobljenih iz devetnajstih analiziranih študij, se zdi, da je treba za širšo uporabo LD v geodetskih upravah najti rešitve za premagovanje teh ugotovljenih težav. Težave heterogenosti in podatkovnih kontekstov (metapodatki) ovirajo zahteve po interoperabilnosti, kot sta berljivost in serializacija. Najpomembnejše je, da morajo geodetske uprave izboljšati kakovost, učinkovitost in podporo za objavo in ponovno uporabnost LD. Večino odgovorov na ta vprašanja je mogoče najti v prispevkih, objavljenih v zadnjih petih letih, v kratkem pa bodo predstavljeni v razpravnem delu.

1.1.5 Razprava o raziskovalnih vprašanjih

V tem delu razpravljamo o rezultatih sintezne faze pregleda literature in odgovarjamo na raziskovalna vprašanja, določena v protokolu tega pregleda literature.

5.1 RQ1: Katere težave je treba rešiti pri objavljavanju geoprostorskih podatkov geodetskih uprav kot LD?

Najpogostejše omenjena težava je heterogenost podatkov, to je pomenska, strukturna in sintaktična nedoslednost. Na vprašanje heterogenosti, nepovezanih in izoliranih podatkov je opozorilo več avtorjev (Çağdaş in Stubkjær, 2015a; Debruyne et al., 2016; Huang et al., 2019; Huerta, Schade, Granell, 2014; Mai et al., 2019; Ronzhin et al., 2019; van den Brink et al., 2018; Vilches-Blázquez & Saavedra, 2019).

Zaradi semantične heterogenosti (Bucher et al., 2020) pravnih definicij in terminologije med neodvisnimi vladnimi agencijami za upravljanje zemljišč je pristop od zgoraj navzdol oviran (ali celo nemogoč) (Ronzhin et al., 2019). Zemljiške knjige in katastri so bili določeni v skladu z različnimi podatkovnimi modeli in standardi, kar je povzročilo težave z interoperabilnostjo (Çağdaş & Stubkjær, 2015a). Tudi za katastrske proizvajalce in podatke so značilne težave glede heterogenosti (Huerta, Schade in Granell, 2014). Vse večji je razkorak med klasičnimi načini uporabe geografskih informacijskih sistemov (GIS) še danes in odprtimi, raziskovalnimi pristopi, ki se uporabljajo za pridobivanje in porabo podatkov iz grafov znanja, kot so povezani podatki (Mai et al., 2019). Heterogenost razpoložljivih prostorskih podatkov uporabnikom podatkov, spletnim aplikacijam in storitvam otežuje odkrivanje, interpretacijo in uporabo informacij v velikih in porazdeljenih spletnih sistemih (van den Brink et al., 2018). Prostorski podatki so po navadi zelo veliki (veliki podatki), kar predstavlja težave pri deljenju ali uporabi odprtih prostorskih podatkov prek spleta (van den Brink et al., 2018).

Naša raziskava je razkrila tudi opredelitev termina »okolščine podatka«, angleško »data context« kot enega od problemov pri uveljavljanju paradigme LD v praksi geodetskih uprav, saj so metapodatki pogosto preslabo pripravljeni. To pomeni, da tudi metapodatki, ki so standardizirani in urejeni z direktivami, kot je "Infrastruktura za prostorske informacije v Evropski skupnosti (INSPIRE)," 2007, niso pomensko definirani znotraj kontekstno dovolj bogatega referenčnega sistema.

5.2 RQ2: Kakšne so rešitve in zahteve za objavo LD pri geodetskih upravah?

Identificirali smo šest bistvenih rešitev in njihovo kombinacijo za uspešno objavo LD v geodetskih upravah, navedenih v Preglednica 6: besednjaki, ontologije, metapodatki in rešitve, ki uporabljajo standardizirane tehnologije, ki izpolnjujejo zahteve, kot sta berljivost in serializacija. Oblikovanje besednjakov za ustvarjanje in objavo LD zahteva razumevanje področja znanja in predvidene uporabe (Çağdaş in Stubkjær, 2015b; Ronzhin et al., 2019). Vzpostaviti je treba standardni nabor dogovorjenih ontologij, besednjakov in nizov ključnih besed geografskih metapodatkov za opis odnosov med prostorskimi značilnostmi (Wiemann & Bernard, 2016). V zvezi s standardi je OGC razvil arhitekturo, ki podpira njegovo vizijo geoprostorske tehnologije in interoperabilnosti podatkov, imenovano OGC Abstract, ki zagotavlja konceptualno podlago za večino specifikacij OGC (Ariza-López et al., 2021). En objekt iz resničnega sveta ima lahko različne predstavitve podatkov in serializacije, ki včasih izvirajo iz različnih organizacij ali izpostavljajo različne vidike razpoložljivih podatkov. (Schleidt et al., 2020).

Serializacija domene LD se uporablja v RDF/XML, Turtle, JSON-LD in HTML (Beek & Folmer, 2017; Hietanen et al., 2016; Schleidt et al., 2020).

5.3 RQ3: Katere so prednostne teme podatkov geodetskih uprav za objavo v obliki LD?

Geodetske uprave zagotavljajo različne količine podatkov LD in geoprostorske teme na področju objave LD. Zanimajo nas najpogostejše javno objavljene geoprostorske teme. Zato smo v

Preglednica 7 povzeli podatke LD, ki so na voljo na portalih geodetskih uprav (Preglednica 8) izbranih držav.

Preglednica 7 Geoprostorske teme izbranih držav, ki so najpogostejše objavljene kot LD na portalih geodetskih uprav.

Geoprostorska tema	Število tematik
Upravne enote	7
Poštni enote	1
Zemljepisna imena	1
Imena krajev	3
Ulice	1
Naslovi	2
Identifikatorji stavb	3
Topografija (malo merilo)	1
Topografija (veliko merilo)	1
Državne ceste	1
Kataster	1
Koordinatni referenčni sistem (CRS)	1
Postajališča javnega prevoza	1
Skupaj	24

Glede na Preglednica 7 so najpogostejše objavljene teme upravne enote (7), sledijo imena krajev (3), identifikatorji zgradb (3) in naslovi (2). Sklepamo, da dajejo Geodetske uprave prednost temam objave, ki omogočajo posredno sklicevanje (poizvedba, navigacija) za zunanje uporabnike. Predlagamo lahko IGN France, ki je prepoznal potrebo po in ustvaril besednjak za opis koordinatnih referenčnih sistemov (CRS) v strojno berljivi obliki (van den Brink et al., 2018).

Preglednica 8 Pregled geodetskih uprav in naslovov njihovih portalov povezanih podatkov (LD)

Država	LD portal geodetske uprave	Spletni naslov
NL	PLDN linked data environment	https://data.labs.kadaster.nl/kadaster/-/overview
UK	Ordnance Survey Linked Data Platform	https://data.ordnancesurvey.co.uk/
IE	OSI's National Map (Prime2) data	https://osi.ie/blog/linked-data/
FI	Linked Data Finland	https://www.ldf.fi/datasets.html
FR	Ontology of IGN administrative units	http://data.ign.fr/def/geofla/20210218.htm
CH	The Linked Data Service	https://ld.geo.admin.ch
NO	Kartverkets LOD	https://rdf.kartverket.no/api/1.0#/administrativeunit

Najproduktivnejše države glede objavljajanja LD so Nizozemska (upravne enote, topografija, naslovi in zgradbe, topografija velikega merila, kataster), Švica (upravne enote, postajališča javnega prevoza, imenik ulic, imenik naslovov in drugo) oziroma Finska (zemljepisna imena, krajevna imena, upravne enote, zgradbe). V povprečju so geodetskih uprav sedmih izbranih držav že objavile približno tri geoprostorske teme kot LD.

5.4 RQ4: Katere izzive/težave so geodetske uprave ugotovile za prihodnji razvoj povezanih podatkov?

Izzivi, s katerimi se geodetske uprave soočajo pri izvajanju geoprostorskega objavljajanja LD, so večplastni in segajo od tehnoloških do vprašanj, povezanih s politiko. Ti izzivi, navedeni v izločenih dokumentih, vključujejo odločanje o prioritetah razvoja, sledenje tehnološkemu trendom, upoštevanje EU in državnih politik o izvajanju odprtih podatkov ter nadomeščanje, kompenziranje finančnih stroškov souporabe podatkov. Poleg tega lahko spretnosti osebja, ki dela na lokalnih državnih projektih, ali strokovnjakov v geodetskih upravah in razpoložljivost smernic najboljše prakse predstavljajo izzive pri izvajanju objavljajanja LD.

Naše raziskovalno vprašanje o izzivih geodetskih uprav v zvezi z geoprostorsko objavo LD bi lahko podprli s koncepti, izraženimi v študijah izbranih avtorjev in drugih (

Preglednica 9).

Preglednica 9 Izzivi za geodetske uprave v zvezi z geoprostorsko objavo LD

Izzivi za geodetske uprave v zvezi z geoprostorsko objavo LD Avtorji
Povezljivost končnih točk LD Mai et al., 2019
Uporaba ontologij za izpeljavo ustreznih GIS predstavitev Mai et al., 2019
Večjezično okolje EU in širše Çağdaş in Stubkjær, 2015b
Ustvarjanje ontologij iz obstoječih modelov UML Ronzhin et al., 2019
Razširljiv pristop za zmanjšanje časa in stroškov, povezanih z objavo Rowland et al., 2022
Avtomatizacija je proces pretvorbe vhodnih geoprostorskih podatkov v povezane podatke. Kyzirakos et al., 2018
Ustvarjanje tematskih zemljevidov s kombiniranjem atributov iz povezanega odprtega podatkovnega oblaka z geometrijo, shranjeno v strežniku prostorske zbirke podatkov (SDS) Owusu-Banahene, 2018
Izbira med dvema pristopoma pri obravnavanju geometrij: datoteke GML, GeoJSON ali strukturirani povezani podatki, ki temeljijo na RDF, van den Brink et al., 2018
Zagotavljanje ustrezne kakovosti podatkov za modele metapodatkov Beyza et al., 2020

Težave, ki jih navajajo zgoraj omenjeni avtorji, so v izbranih študijah delno rešene, saj gre večinoma začetne in izvirne raziskave, ki poskušajo odgovoriti na izzive. Manj študij ponuja stabilne proizvodne rešitve, ki so lahko še vedno omejene na njihov lokalni primer uporabe (Ma, 2017). Predstavljajo vzdržljiv izziv za druge geodetske uprave, ki delujejo v drugačnem kontekstu.

5.5 RQ5: Kakšni so razlogi za relativno počasno sprejemanje geoprostorske LD v geodetskih upravah?

Zadnje raziskovalno vprašanje, RQ 5, je najbolj splošno in zato zahteva podrobnejši pregled različnih vidikov in predlogov za izboljšanje razširjanja LD v geodetskih upravah. Ustrezni elementi iz pregledanih študij so zbrani v podporo argumentu za počasno širjenje LD.

Nacionalne kartografske in katastrske agencije so „uradni“ državni organi, za katere je država določila nacionalni interes, da pooblasti javni organ za izdelavo zemljevidov, ki podpirajo splošni namen. Prehod na politiko odprtih podatkov predstavlja izziv za poslovni model nacionalnih kartografskih in katastrskih agencij (NMCA), oziroma geodetskih uprav (Folmer et al., 2020). Zato bi morala biti povezani podatki (LD) del poslovnih razprav, vključno z izjavami o strateški viziji in ambicijah. Uvedba pristopa povezanih podatkov v domeno zemljiške administracije lahko naredi nabore podatkov o zemljiški administraciji bolj dostopne prek spleta (Çağdaş & Stubkjær, 2015a). Geodetske uprave zbirajo zanesljive, zaupanja vredne in dobro vzdrževane geoprostorske informacije, vendar niso edini javni subjekti, ki morajo to početi. Sodelujejo z drugimi državnimi organi, kot so statistika, državne agencije in občine.

Na podlagi dobro vzpostavljenega sodelovanja javnih subjektov lahko pričakujemo nadaljnje merljive prihranke stroškov in povečanje učinkovitosti za upravo z uporabo geoprostorske platforme LD, ki temelji na nacionalnih standardih (Coumans, 2017). Okvir za sodelovanje kot platforma bo prispeval k interoperabilnosti in integraciji različnih vrst informacij javnega sektorja in drugih geoprostorskih informacij (Crompvoets et al., 2016). Razumevanje podatkov geodetskih uprav je zapleteno in zahteva znanje z vsaj dveh področij: geoinformacije in zemljiške administracije (Ronzhin et al., 2019).

Na žalost je veliko ponudnikov podatkov v zadnjih letih opustilo podporo in vzdrževanje javnih končnih točk SPARQL, s čimer je ogroženo zaupanje med potrošniki in ponudniki odprtih povezanih podatkov. Zaradi birokracije in zapletene zakonodaje, ki je značilna za javne storitve, so javni organi veliko počasnejši pri sprejemanju novih tehnologij, kot so povezani podatki, kot zasebne organizacije in podjetja (Mouzakitis et al., 2017).

Čeprav tehnologije povezanih podatkov in semantičnega spleta niso nove, je bilo njihovo sprejemanje v različnih panogah počasno, morda zaradi omejenega orodja za podporo tem prizadevanjem (Rowland, Folmer in Beek 2020). Sodobna orodja za povezane odprte podatke pogosto ne morejo pravilno oglaševati uporabnosti naborov podatkov potencialnim uporabnikom, kar ovira ponovno uporabo (Folmer et al., 2019; A. Rowland et al., 2021; Rowland et al., 2022). Najboljše prakse podatkov na spletu podajajo *splošne* smernice za objavo podatkov na spletu, ne pa posebnih smernic o prostorskih podatkih (van den Brink et al., 2018). V zadnjih nekaj letih so bili povezani odprti podatki sorazmerno počasi sprejeti, deloma zaradi pomanjkanja dovolj značilnih in funkcionalnih možnosti za brskanje in pregledovanje. (Folmer et al., 2018).

Pomanjkanje znanja o upravljanju grafov znanja in podatkovnih nizov povezav je omejevalni dejavnik za shemo uvedbe "petih zvezdic" (Tim Berners-Lee, 2010), ki se uporablja tudi za odprto objavo geoprostorskih podatkov. In to so:

1. *Omogočite dostop do podatkov na spletu v kateri koli obliki.*
2. *Dajte podatke na voljo kot strukturirane podatke (npr. Excel).*
3. *Uporabljajte nelastniške formate (npr. CSV namesto formata Excel).*
4. *Uporabite URI-je za označevanje stvari, da lahko ljudje pokažejo na vaše podatke.*
5. *Povežite svoje podatke z drugimi podatki, da zagotovite kontekst.*

Semantic Computing Research Group (SeCo), avtorji platforme Linked Data Finland (Hyvönen et al., 2014), so založnikom podatkov predlagali dve dodatni zvezdici:

6. *Naboru podatkov zagotovite eksplicitno shemo metapodatkov, ki pojasnjuje nabor podatkov.*
7. *Preverite podatke glede na shemo za boljšo kakovost.*

Ta celoviti pristop omogoča ponovno uporabo podatkov za boljše razumevanje značilnosti nabora podatkov in ovrednotenje kakovosti podatkov za predvideni namen (Hyvönen et al., 2014). Zato je premostitev ozaveščenosti strokovnjakov o razširjeni ponovni uporabi podatkov v povezani obliki za spletno objavo ključnega pomena za usmeritev prihodnjih raziskav. V tem kontekstu bi bilo zanimivo vedeti, kako deliti lastništvo in odgovornost za vzdrževanje grafa znanja (Ronzhin et al., 2019).

Primerjalno merilo GeoSPARQL lahko izvajalcem pomaga izboljšati rešitve za shranjevanje RDF (Jovanovik et al., 2021).

Velitchkov (2021) je zastavil anketno vprašanje "Kakšne so ovire za uporabo povezanih podatkov?" manjši skupini strokovnih delavcev LD in predstavili fin na konferenci ENDORSE (2021). Najbolj zanimivi odgovori strokovnjakov za LD so, da LD zahteva od ljudi razmišljanje na nov način, da je LD preveč akademska in da založniki sami ne uporabljajo LD. Poleg tega LD zaradi pomanjkanja poklicnega povpraševanja ne uspe privabiti prodajalcev IT.

Poleg tehničnih rešitev je treba pred operativno uvedbo rešiti več netehničnih vprašanj na institucionalni, politični, pravni in družbeni ravni (Wiemann in Bernard, 2016). Zato je bil okvir za učinkovito upravljanje zemljišč (FELA) razvit kot krovni vodnik politike (strokovna skupina Združenih narodov za upravljanje in upravljanje zemljišč, 2019). V Evropi obstaja šest sklopov instrumentov upravljanja nacionalnih SDI: strukture kolektivnega odločanja, strateško upravljanje, dodeljevanje nalog in odgovornosti, ustvarjanje trgov, med organizacijska kultura, upravljanje znanja ter ureditev in formalizacija infrastrukture (Crompvoets in Vancauwenberghe, 2018), ki jih je primerno sprejeti neposredno v LAS, vključno z založniško vloge geodetskih uprav na področju povezanih podatkov LD. Z odgovorom na vsa raziskovalna vprašanja smo zagotovili bolj strukturiran pregled težav, rešitev in izzivov, s katerimi se geodetske uprave soočajo v zvezi z geoprostorskim objavljanjem LD.

Če povzamemo, sprejetje LD s strani geodetskih uprav zahteva celosten pristop, ki obravnava tehnične in ne tehnične izzive. Z izvajanjem ustreznih smernic, orodij in strokovnega znanja lahko geodetske uprave zagotovijo, da so njihova prizadevanja za objavljanje LD uspešna in izpolnjujejo potrebe njihovih deležnikov. Podobno lahko obravnava institucionalnih, političnih, pravnih in družbenih vprašanj zagotovi, da so založniška prizadevanja LD usklajena z nacionalnimi prednostnimi nalogami in se lahko dolgoročno vzdržujejo.

1.1.6 Sklepi

V tem pregledu literature smo iskali trenutno stanje uporabe povezanih podatkov LD v geodetskih upravah in izzive, s katerimi se le-te soočajo. Rezultat sistematičnega pregleda je bilo zadnjih devetnajst izbranih študij za pridobivanje informacij v celotnem besedilu. Če bi vseh prvotno identificiranih 132 študij vključili v podrobno pridobivanje podatkov, bi to bistveno vplivalo na kakovost predstavljenega dela. Razlogi za ne vključitev posameznih dokumentov so v metodologiji sistematičnega pregleda literature, ki je zaradi sicer velike obremenjenosti namenjena izčrpnemu preučevanju omejenega števila primernih člankov. S podporo recenzentskega protokola smo ocenili kakovost študij s pomočjo šestih kriterijev za izbiro študije, kot primerne za odgovarjanje na raziskovalna vprašanja.

Odgovori na pet raziskovalnih vprašanj pričujoče raziskave so povzeti takole.

Odgovori na RQ1: **Težave**

Več avtorjev je izpostavilo vprašanje heterogenosti, konteksta, metapodatkov in ločenih podatkov kot glavne težave pri objavljanju povezanih geoprostorskih podatkov geodetskih uprav (Çağdaş in Stubkjær, 2015a; Debruyne et al., 2016; Huang et al., 2019; Huerta, Schade, Granell, 2014; Mai et al., 2019; Ronzhin et al., 2019; van den Brink et al., 2018; Vilches-Blázquez & Saavedra, 2019).

Odgovori na RQ2: **Rešitve in zahteve**

Na podlagi analiz devetnajstih člankov smo identificirali šest bistvenih rešitev in njihovo kombinacijo za uspešno objavo LD v navedenih geodetskih upravah: slovarje, ontologije, metapodatke in rešitve, ki uporabljajo standardizirane tehnologije, ki izpolnjujejo zahteve, kot sta berljivost in serializacija. Pri razvojnih projektih LD je bistveno zagotoviti sistematično podporo ustvarjalcem in uporabnikom pri razvoju slovarjev in ontologij ter razširljivosti pristopov za obravnavo velikih podatkovnih nizov.

Odgovori na RQ3: **Prednostne teme podatkov**

V povprečju so geodetske uprave sedmih obravnavanih držav že objavile približno tri geoprostorske teme kot LD. Najpogostejše objavljene teme so upravne enote (7), sledijo imena krajev (3), identifikatorji stavb (3) in naslovi (2).

Odgovori na RQ4: **Izzivi**

Kot je navedeno v analiziranih prispevkih, izzivi vključujejo določanje razvojnih prioritet, sledenje tehnološkemu trendom, upoštevanje EU in državnih politik o uvajanju odprtih podatkov ter nadomestilo finančnih stroškov deljenja podatkov. Poleg tega lahko spretnosti osebja, ki dela na lokalnih državnih projektih, ali strokovnjakov v geodetskih upravah in razpoložljivost smernic najboljše prakse predstavljajo izzive pri izvajanju objavljanja LD.

Odgovori na RQ5: **Razlogi za počasno širjenje LD**

Javni organi veliko počasneje kot zasebne organizacije in podjetja sprejemajo nove tehnologije, kot so povezani podatki (Mouzakitis et al., 2017). Čeprav tehnologije povezanih podatkov in semantičnega spleta niso nove, je bilo njihovo sprejemanje v različnih panogah počasno, morda zaradi omejenega orodja za podporo tem prizadevanjem (Rowland, Folmer in Beek 2020). Pomanjkanje znanja o upravljanju grafov znanja in podatkovnih nizov povezav omejuje shemo uvajanja »petih zvezdic« (Tim Berners-Lee, 2010), ki velja tudi za odprto objavo geoprostorskih podatkov. Sodobna orodja za povezane odprte podatke pogosto ne morejo pravilno oglaševati uporabnosti naborov podatkov potencialnim uporabnikom, kar ovira ponovno uporabo (Folmer et al., 2019; A. Rowland et al., 2021; Rowland et al., 2022). V zadnjih nekaj letih so bili povezani odprti podatki sorazmerno počasi sprejeti, deloma zaradi pomanjkanja dovolj značilnih in funkcionalnih možnosti za brskanje in pregledovanje. (Folmer et al., 2018). Velitchkov (2021) predpostavlja, da založniki sami ne uporabljajo LD, LD pa zaradi pomanjkanja strokovnega povpraševanja ne uspe privabiti prodajalcev IT.

Čeprav pregled ni obravnaval pravnih in zapletenih tehničnih vprašanj, nameravajo avtorji podrobneje raziskati LD, da bi razumeli, katere kombinacije pristopov, protokolov in standardnih semantičnih programskih tehnologij so najučinkovitejše za objavo geoprostorskih podatkov geodetskih uprav v spletnih ekosistemih.

Pomembno je omeniti, da so lahko zaključki sistematičnega pregleda literature omejeni z zastarelimi informacijami v nekaterih analiziranih člankih. Poleg tega študija ni obravnavala tehnologij, standardov, procesov ali smernic najboljše prakse za objavo LD pri geodetskih upravah, kar bo predmet naših prihodnjih preiskav.

1.2 Naloga 1.1: Analiza mogočih kombinacij semantičnih tehnologij za učinkovito večkratno uporabo (R1.1)

V prejšnjem poglavju smo izvedli pregled stanja semantične medopravilnosti za geoprostorske podatke pri Geodetskih upravah v EU. Ugotovili smo ključni nabor semantičnih tehnologij, in sicer smo na podlagi analiz devetnajstih člankov identificirali šest bistvenih rešitev in njihovih kombinacij za uspešno objavo LD v navedenih geodetskih upravah, kot so našteje v nadaljevanju:

Spletni slovarji, ontologije, metapodatki in rešitve, ki uporabljajo standardizirane tehnologije, ki izpolnjujejo zahteve, kot sta *standardizirana* strojna in humana berljivost ter *serializacija* (edinstvena identifikacija - dodeljevanje edinstvene serijske številke). Pri razvojnih projektih LD je bistveno zagotoviti *sistematično podporo* ustvarjalcem in uporabnikom pri razvoju slovarjev in ontologij ter *razširljivosti* pristopov za obravnavo velikih podatkovnih nizov.

Najpogostejši tehnološki koncepti LD, kot so *splet* (5,5 %), RDF (3,7 %), GEOSPARQL (2,5 %) in ontologija (2,4 %). Ker je Nizozemska najbolj napredna EU država na področju semantične medopravilnosti geoprostorskih podatkov, v nadaljevanju predstavili analizo kombinacij semantičnih tehnologij za učinkovito večkratno uporabo na primeru Nizozemske geodetske uprave imenovane Kadaster.

1.2.1 Uvod

Nizozemska agencija za kataster, zemljiško knjigo in kartiranje, je kot geodetska uprava (www.kadaster.nl), verodostojen vir informacij v zvezi z upravnimi in prostorskimi podatki, ki obravnavajo nepremičnine in lastninske pravice na Nizozemskem. Geodetska uprava vzdržuje več velikih ključnih registrov nizozemske vlade, vključno z osnovnim registrom za naslove in zgradbe (nizozemska kratica: BAG), osnovnim registrom za topografijo (nizozemska kratica: BRT) in osnovnim registrom za topografijo velikega obsega (nizozemska kratica: BGT); vsi so na voljo kot odprti podatki. Poleg tega organizacija aktivno objavlja in vzdržuje ta in druga geoprostorska sredstva kot povezane (odprte) podatke in kot del tega prizadevanja in v duhu nenehnih inovacij je bilo več geoprostorskih baz podatkov zdaj ponovno objavljenih na spletu kot povezani, odprti podatki (Open Linked Data – OLD). V času, ko je bil ta dokument prvotno predložen za predstavitev na 4. mednarodni delavnici o geoprostorskih povezanih podatkih na konferenci EWSC 2021 [1], sta bili dve od teh geoprostorskih baz objavljeni kot povezani podatki. Od oktobra 2021 so bili ob uporabi tega novega pristopa objavljeni štirje osnovni registri in številni zunanji nabori podatkov, povezani s temi registri: s poudarkom na učinkovitosti in ponovljivosti tega pristopa.

Oblikovanje in izvajanje tega novega pristopa znotraj nizozemske geodetske uprave Kadaster je motiviralo povečano povpraševanje po povezanih podatkovnih produktih, tako interno kot eksterno, ki jih je treba dostaviti na časovno in stroškovno učinkovit način. Razpoložljivost in stalen razvoj povezanih podatkovnih tehnologij in standardov sta omogočila uvedbo posodobljenega in razširljivega pristopa k objavi povezanih podatkov, da bi zadostili tej zahtevi.

Namen naslednjega razdelka je podrobneje orisati problemski kontekst, ki je vodil zasnovo in implementacijo tega pristopa znotraj Geodetske uprave Kadaster, nato pa orisati tako novosti v pristopu modeliranja kot tudi v arhitekturi, uporabljeni za objavo teh naborov podatkov.

V nadaljevanju bomo razpravljali tudi o širši viziji, ki vodi izvajanje tega pristopa v obravnavani organizaciji. Po tem, ko je s tem pristopom na voljo več naborov odprtih podatkov in s prvo implementacijo grafa znanja, ki je na voljo v Kadastru, bo ta dokument ponudil tudi oceno in perspektivo pridobljenih izkušenj o pristopu, predstavljenem v tem dokumentu.

1.2.2 Tehnološki kontekst ustvarjanja povezanih geoprostorskih podatkov

Upravni geoprostorski podatkovni nizi so organizirani kot podatkovni silosi, imenovani *ključni registri*; katerih integracija je tudi na splošno slaba. Uporabniki, ki želijo združiti te silose za določen namen, se

morajo pogosto zateči k uporabi posebnih orodij, ki podpirajo to integracijo, ali pa prenosu celotnih nizov podatkov in izvajanju integracije celotnih nizov podatkov.

Za izvedbo semantične integracije mora uporabnik uporabiti posebna orodja GIS (na primer QGIS ali ArcGIS, modul Interoperability) ali pa prenesti celoten register ključev (kodiranja) in nato izvesti integracijo. Cilj lažje uporabniške integracije osnovnih registrov je razlog za razvoj projektov integracije.

Ena od rešitev težave integracije, je *zagotoviti te podatkovne silose kot povezane, odprte podatke* in izvesti integracijo z uporabo povezovalnih semantičnih tehnologij.

Dejansko obstaja vse večje povpraševanje po povezanih podatkovnih storitvah tako znotraj geodetskih uprav kot zunaj njih, kar prinaša potrebo po tem, da so povezani podatki na voljo hitreje in na najbolj (stroškovno) učinkovit način. To povpraševanje, skupaj z naraščajočo razpoložljivostjo in nenehnim razvojem povezanih podatkovnih tehnologij in standardov, je gonilna sila za inovacijo povezanega objavljanja podatkov v organizaciji geodetske uprave.

Opisan pristop izpolnjuje omenjeno potrebo na več načinov. Prvič, ta *pristop* v skladu s splošnimi arhitekturnimi načeli uporablja obstoječe knjižnice (npr. `rdf-validate-shacl` (<https://github.com/zazuko/rdf-validate-shacl>) in `hdt-cpp` (<https://github.com/rdfhdt/hdt-cpp>)), ki gradita na odprtokodnih projektih kot na primer *Comunica* (<https://comunica.dev>) in *ClioPatria* (<https://cliopatria.swi-prolog.org>) in se zato *izogiba potrebi po razvoju lastnih rešitev po meri*, za izpolnjevanje tega povpraševanja. Podobno ta pristop uporablja obstoječe komercialne izdelke, kjer so na voljo, v interesu zmanjšanja stroškov vzdrževanja.

Drugič, ta pristop uporablja *načelo konfiguracije nad kodo*, ki zagotavlja, da se isti » informacijski cevovod« (pipeline) uporablja za vse povezane projekte objavljanja podatkov, pri čemer je treba le konfiguriranje komponent, kjer je to treba. Vidik arhitekture sistema je tisti, ki omogoča, da se ta *pristop lahko ponovi na več naborih podatkov*, pri čemer izvajanje od vira do objave poteka v kratkem časovnem okviru.

Tretjič, implementacija vseh pomembnih komponent v tej zasnovi je izvedena z mislijo na *pretočni pristop*. V praksi to pomeni, da so vsi *povezani podatkovni modeli čim bližje izvirnemu modelu* in da lahko izbrani viri podpirajo funkcijo pretakanja v interesu dostave podatkov v realnem času.

Čeprav ta funkcija pretakanja še ni delujoča funkcija v opisani arhitekturi, je ta vključitev v zasnovu pomembna v viziji širše slike za ta pristop, kot je razloženo v nadaljevanju.

To so razmeroma zapleteni, povezani nabori podatkov, od katerih ima vsak kompleksen podatkovni model in so po svojem količinskem obsegu veliki. Dejansko vsak nabor podatkov vsebuje med 800 milijonov in 1 milijardo *trojčkov*.

Ta pristop poudarja izboljšano stroškovno učinkovitost in učinkovitost virov ter krepi poslovni primer za povezane podatke v organizaciji, kot je Geodetska uprava [3]. V razdelkih, ki sledijo, so predstavljeni koncepti in arhitektura, ki podpirajo ta nov pristop k objavi povezanih podatkov znotraj Nizozemske Geodetske uprave, vključno s standardi, tehnologijami in ustreznimi odločitvami v zvezi s tem med objavo naborov geoprostorskih podatkov s strani Geodetske uprave.

1. Izvirni viri geoprostorskih podatkov
2. *Model, graf znanja* proti informacijskemu modelu

Prvi od novih dodatkov k *objavi povezanih podatkov (LD)* Geodetske uprave Kadaster je *izrecna razlika med modelom znanja in informacijskim modelom*, pri čemer oba sestavljata večji povezani podatkovni model za vsak nabor podatkov. Ta ločitev odraža dejstvo, da *mora biti povezan podatkovni model sposoben opisati pomen podatkov zunanjemu svetu z uporabo modela znanja*, hkrati pa v informacijskem modelu opisati vidike modela, specifične za organizacijo.

Ta ločitev omogoča *optimizacijo informacijskega modela* glede na notranje zahteve organizacije, vključno s posebnimi modeli in procesi v zvezi s sredstvom, hkrati pa še vedno omogoča optimizacijo povezanega modela znanja za učinkovito podporo zunanjim, skupnostnim standardom objave, ki so potrebni za namene odkrivanja in medoperabilnosti. [4].

Ker so tako notranji kot zunanji vidiki pomembni za prizadevanja Geodetske uprave Kadaster pri objavljanju podatkov, ta novi pristop k objavljanju povezanih podatkov bolje izvaja organizacijske zahteve za povezane nize podatkov.

Dejansko obstajajo številni posebni razlogi v kontekstu geodetske uprave, da je bila ta delitev potrebna pri oblikovanju novega pristopa k objavljanju.

- a) Prvič, *informacijski model* za dano bazo podatkov vsebuje posebne in notranje informacije v zvezi z danim sistemom. Te informacije vključujejo lastnosti trenutnih informacijskih sistemov, ki se uporabljajo v bazi in v bližini dane baze, pravila, specifična za organizacijo, ki se nanašajo na bazo, kot tudi tehnične podrobnosti.

Na primer, dejstvo, da imena občin ne smejo biti daljša od 80 znakov, ni izrecno kodirano v zakonodaji, ampak je prej omejitvev, ki jo nalagajo notranji sistemi, ki jih geodetska uprava uporablja za shranjevanje teh informacij.

Vključitev informacije o omejitvi dolžine zapisa v model znanja bi bila nepravilna: ta največja dolžina zapisa ni del semantike za imena občin. Vendar pa bi bilo tudi izpuščanje teh informacij nepravilno, ker povezani podatkovni model ne bi mogel izključiti vrednosti, ki jih notranji sistemi za registracijo pri Geodetski upravi ne sprejemajo.

Kot take so te informacije modelirane v informacijskem modelu danega sredstva in predstavljene z uporabo »*jezika za omejitve oblik*« (*Shapes Constraint Language, SHACL*) [5], ki služi za omejevanje danega modela na podlagi interno določenih pravil in odnosov za dano sredstvo.

Obstaja več tehničnih pristopov, ki bi se lahko uporabili za formalizacijo takšnih omejitev v informacijskem modelu, na primer shema XML ali shema JSON, vendar je *SHACL najbolj optimiziran za podatke v obliki grafov in povezane podatke*. Zato je bil SHACL izbran kot predstavitveni jezik v tem projektu.

- b) Drugič, *model znanja* za isto bazo podatkov opredeljuje vsako generično in izmenljivo znanje, ki ga je pomembno obdržati v organizaciji, vendar ga je treba deliti z drugimi. Model znanja tudi *olajša ponovno uporabo zunanjih povezanih podatkovnih modelov* s kontekstom, specifičnim za organizacijo.

Na primer, *zunanjó prednostno oznako SKOS (skos:prefLabel)* lahko uporabite za predstavitev imen javnih prostorov (model znanja), hkrati pa tudi kodirate zgoraj omenjeno omejitev števila znakov v zapisu (informacijski model).

V implementaciji modela znanja v Geodetska upravi Kadaster vsi povezani podatkovni modeli med drugim uporabljajo besednjake in ontologije RDF(S) [6], OWL [7] in/ali SKOS [8], ki služijo temu, da se model poveže z zunanjim svetom, vedno v interesu ponovne uporabe in interoperabilnosti.

Kot je prikazano na spodnji sliki, informacijski model in model znanja nista popolnoma neodvisna drug od drugega in sta dejansko preslikana drug v drugega, ko definirata in preoblikujeta model.

Ta proces na primer preslika

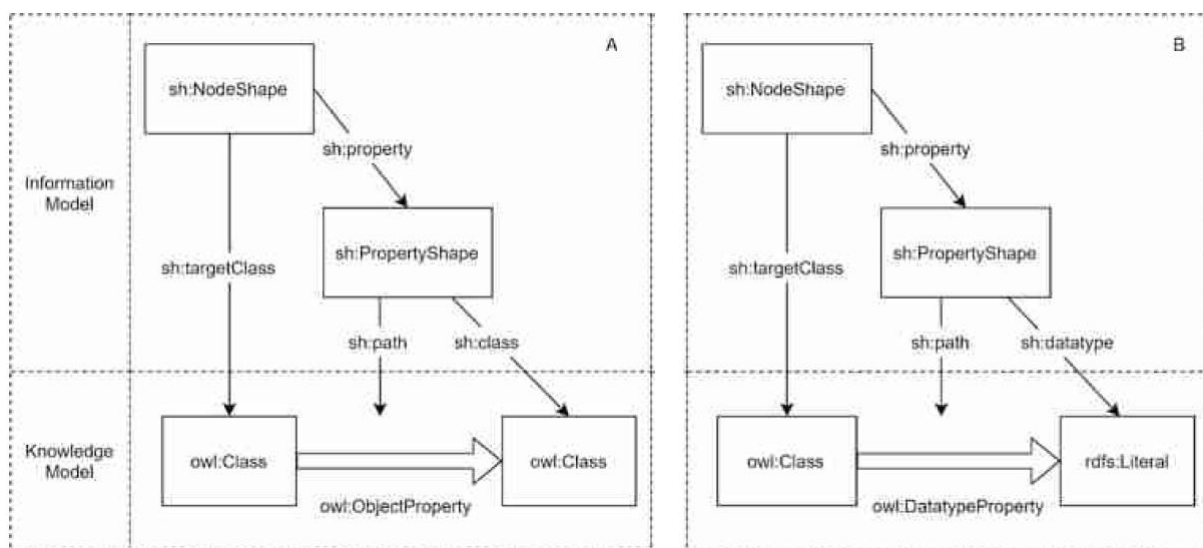
- oblike SHACL, definirane v informacijskem modelu,

v ustrezne

- razrede OWL ali literale RDF, definirane v modelu znanja.

Obstajata dve različici tega postopka preslikave, ena je preslikava lastnosti objekta v dveh modelih (Slika 1A), druga pa preslikava lastnosti podatkovnega tipa v modelih (Slika 1B).

Obe različici bi morali biti dokončani med pretvorbo podatkovnega modela v povezane LD podatke.



Slika 1 Preslikava modela lastnosti objekta/podatkovnega tipa. Sheme sledijo drugi obliki. (A) (vir: Kadaster)

Slika 1 ponazarja, kako se preslikava med informacijskim modelom in modelom znanja izvaja v praksi glede na *owl:ObjectProperty*; (B) ta slika ponazarja, kako se preslikava med informacijskim modelom in modelom znanja izvaja v praksi glede na *owl:DatatypeProperty*.

- V prvi različici (**Error! Reference source not found. A**) je postopek skoraj enak, razen dejstvo, da so vozlišča SHACL in oblike lastnosti, definirane za lastnosti objekta v danem podatkovnem/informacijskem modelu, preslikane v ustrezne razrede OWL in lastnosti objekta, definirane v modelu znanja.
- V drugi različici (**Error! Reference source not found. B**) so lastnosti podatkovnega tipa definirane v podatkovnem modelu s preslikavo ustreznega vozlišča SHACL in oblik lastnosti za vsak podatkovni tip v ustrezni razred OWL, lastnost podatkovnega tipa in literal RDFS, definiran v modelu znanja.

Da bi podprli boljše preskušanje, validacijo nastalega modela, je bil v procesu modeliranja uporabljen tudi korak validacije SHACL. Ta korak zagotavlja, da so oblike za vsak objekt in lastnost tipa podatkov v podatkovnem modelu v celoti preverjene glede na podatke primerka.

Ta korak validacije vključuje številne najboljše prakse v zvezi z modeliranjem informacijskega modela, osredotočenega na *uporabo zaprtih oblik vozlišč* [9,10].

To zagotavlja, da je model tako specifičen, kot je treba za zagotovitev smiselne validacije modela, hkrati pa še vedno omogoča validacijo pravih, a redkih primerkov podatkov.

Preglednica 10 prikazuje primere *uporabe informacijskega modela poleg modela znanja za izvajanje posebnih najboljših praks* za podatkovni model in na primer podatke. Upoštevajte, da *model znanja (stolpec 3) določa pravilno semantiko/pomen podatkov*. Podatkovni / Informacijski model (stolpec 2) določa dodatne strukturne kriterije, ki – strogo gledano – niso del pomena podatkov.

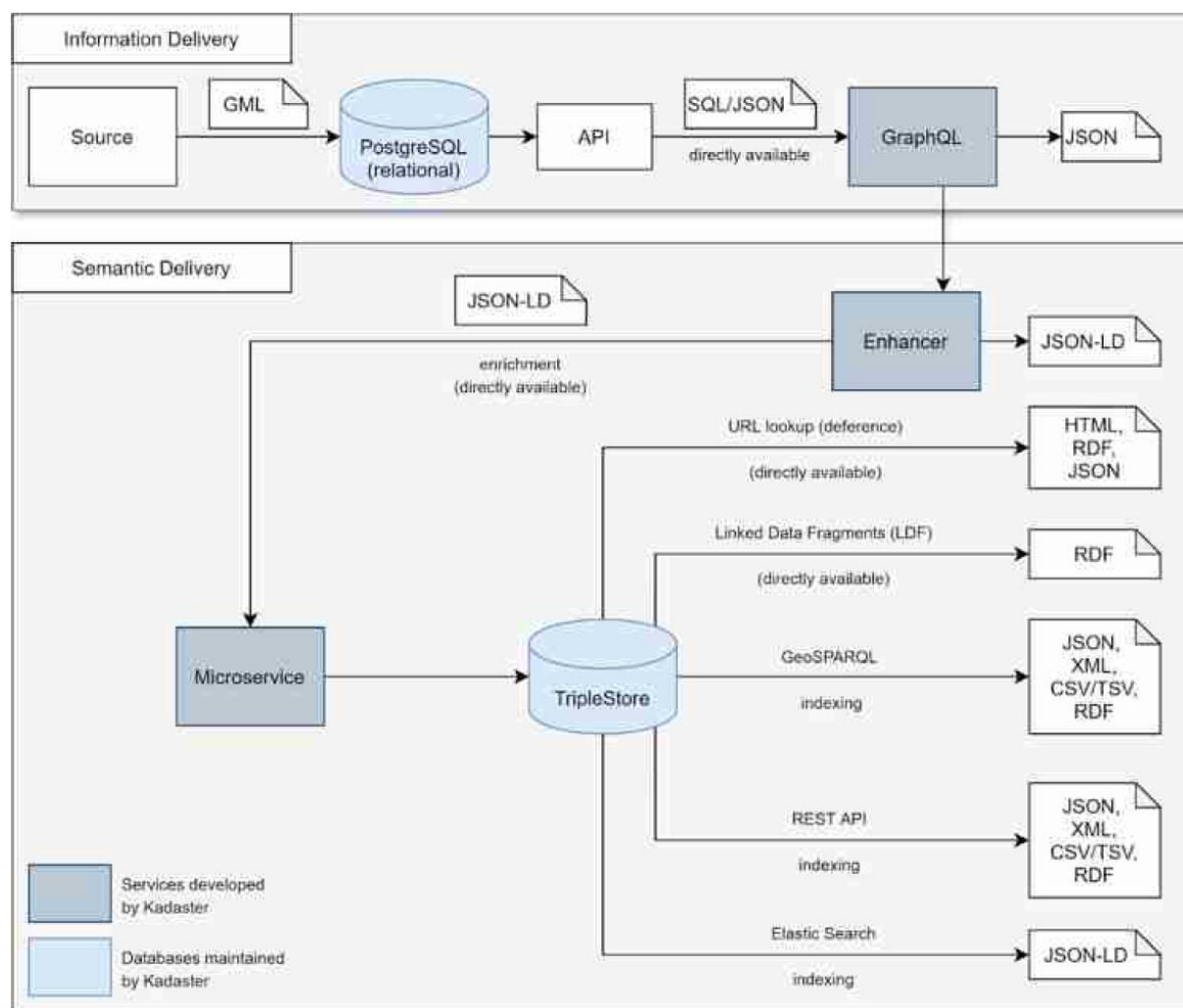
Preglednica 10 Najboljše prakse, uporabljene za validacijo modela znanja in informacij (vir: Kadaster)

Best Practice	Information Model	Knowledge Model	Example
The IRI strategy is correctly implemented.	- Regular expression check on IRI paths.	-	- The unique identifier for a building.
Linguistic content is readable for human users.	- Correct length (sh:minLength, sh:maxLength). - Valid characters (sh:pattern). - Correct language tags (sh:languageIn).	- Published as human-readable content (e.g., skos:prefLabel). - Semantic characterisation of the intended purpose (e.g., skos:example, skos:definition).	- The human-readable definition of a concept. - The name of a public space.
Best Practice	Information Model	Knowledge Model	Example
Subsumption relation is correct and easy-to-use.	- Emit a warning if a class has more than one parent (subsumption relation cannot be shown/navigated as a tree).	- Semantic characterisation of the subsumption relation (rdfs:subClassOf).	- A building is a geospatial object.
Geospatial object must have geometry.	- Emit an error if a building has no geometry.	- Use international OGC standards to semantically characterise the geometry (and its coordinate-reference system).	- Every building is published with exactly one geometry.

1.2.3 Načrtovanje in razvoj podpornih arhitektur

Proces pretvorbe relacijskih podatkov za dano geoprostorsko sredstvo (zbirko podatkov) v povezane podatke se zaključi v več korakih, izvedenih med *ekstrakcijo, pretvorbo in nalaganjem (ETL)* proces.

Ta proces je prikazan v orisu arhitekture na spodnji sliki (Slika 2) in podrobno opisan, kot sledi. Pri podrobnem opisu arhitekture, uporabljene v tem pristopu, bodo občasno vključene kratke razprave o trenutnem izvajanju in področjih za izboljšave z namenom ponuditi pogled na nadaljnji razvoj te arhitekture.



Slika 2 Arhitektura, ki podpira proces ETL, ki zagotavlja povezane podatke (vir: Kadaster)

Prvi korak naloži relacijske podatke iz vira, v tem primeru iz nizozemske platforme za dostop do geostoritev (PDOK), v zbirko podatkov *PostgreSQL* po koraku indeksiranja v geografskem označevalnem jeziku (GML) [11].

Končna točka GraphQL (<https://graphql.org/>) se nato uporabi za dostop do podatkov, dostavljenih prek API-ja iz baze podatkov *PostgreSQL* po dostavi in potrditvi podatkovnega modela od končnega uporabnika.

V praksi se ta korak izvede z razširitvijo tipskih definicij tako, da so objekti v podatkovnem modelu pravilno opisani v *GraphQL*, z razširitvijo *razreševalcev (semantic solver)*, ki omogočajo poizvedovanje po objektih s pravimi parametri, in končno z dodajanjem zahtevanih poizvedb SQL ustreznim razreševalcem.

Upoštevajte, da ta pristop sam po sebi ni omejen na relacijske vire podatkov, saj lahko končna točka GraphQL lahko dostavi tudi iz drugih vrst virov. Poleg tega so vse končne točke GraphQL, ki so na voljo za vsak podatkovni silos, *skladne s standardom Apollo GraphQL*.

Skladnost s tem standardom olajša združevanje končnih točk in zato se v praksi, ko se različne končne točke GraphQL sklicujejo druga na drugo, ker ključni registri vsebujejo medsebojne odnose, se na primer *Apollo uporablja za združevanje teh podatkovnih silosov*.

V procesu pridobivanja podatkov iz vira so podatki na voljo temu »cevovodu čim bližje viru«, da se izognemo morebitnim nepotrebnim kopijam podatkov in zagotovimo, da je *aktualnost podatkov čim bližje podatkom v realnem času*.

Pristop "blizu vira"

Ta pristop "blizu vira" zagotavlja, da pri modeliranju povezanih podatkovnih modelov za vsak nabor podatkov, ki jih je treba pretvoriti, ni narejena nobena nepotrebna ali nepregledna transformacija za podatkovni model in je kot tak povezan podatkovni model popolnoma prepoznaven strokovnjakom na področju.

V trenutni implementaciji arhitekture za vsak nabor podatkov Geodetska uprava pridobi podatke za transformacijo na različnih stopnjah oddaljenosti do vira preprosto na podlagi trenutne metode dostave, ki jo izvaja sam vir.

Ti načini dostave segajo od objav GitHub v določenem časovnem okviru (BAG) do prenosa kopije podatkov, dostavljenih vsako četrtno (BGT) in do možnosti dostopa do podatkov neposredno iz samega objekta (BRT).

Kratkoročno je verjetno, da bo dostava teh virov centralizirana v informacijskem objektu, prek katerega bo lahko ta »cevovod« dostopal do vira.

Ko je model na voljo kot končna točka GraphQL, lahko po njem poizveduje programska rutina »Enhancer«, ki je *interno razvita komponenta, ki omogoča poizvedovanje JSON-LD na podlagi končne točke GraphQL*.

Za to izvedbo so potrebni naslednji koraki.

- a) Prvič, Enhancer ima nabor vnaprej določenih poizvedb s specifičnim časom in/ali parametri *ostranjenja* (paginacije) za vsak objekt tako, da je predmet dostavljen kot končna točka, do katere lahko dostopa *mikrostoritev za dostavo rezultatov JSON-LD* [12].

Konfiguriranje vnaprej določenih poizvedb, ki se uporabljajo v Enhancerju, trenutno deluje z dodajanjem poizvedbe GraphQL *pod imenik/query* v Enhancerju. Ta poizvedba vrne podatke primerka na podlagi objektov, definiranih v podatkovnem modelu za določen nabor podatkov, in se lahko uporabi pri polni obremenitvi nabora podatkov, kot so povezani podatki.

Posebna oblika te končne točke temelji na sprejemni glavi *aplikacije/n-quads* ali *aplikacije/ld+json*.

Upoštevajte, da je do končne točke Enhancer mogoče dostopati tudi z zahtevo POST, ki vsebuje poizvedbo GraphQL.

- b) Drugič, določiti je treba *sklic na ustrezno lokacijo konteksta JSON-LD* za določen nabor podatkov, kar se naredi za vsak nov nabor podatkov, ki gre skozi postopek ETL. Vsak ključ v kontekstu JSON-LD se nanaša na attribute in/ali objekte v definicijah vrst GraphQL, ki podpirajo Enhancer s pretvorbo teh atributov v povezan format podatkov. Kot je razvidno iz prejšnjega razdelka, se lahko SHACL uporablja tako za validacijo podatkovnega modela z uporabo vzorčnih podatkov kot tudi za validacijo transformiranih podatkov primerka z uporabo podatkovnega modela. Znotraj te arhitekture je dejansko potreben *korak preverjanja SHACL*, da se zagotovi veljavnost podatkov, dostavljenih iz Enhancerja.

- c) Končno je v procesu pretvorbe dostave GraphQL v povezane podatke treba *paketno poizvedovati po celotnem naboru objektov z uporabo vnaprej določenih poizvedb*.

Te paketne poizvedbe so vsebovane v minimalnem delu kode, imenovanem *mikrostoritev*, kot je prikazano na sliki (Slika 2).

Celoten postopek ETL uporablja *čim manj kode po meri*, zato postopek uporablja *Apache Airflow* (<https://airflow.apache.org/>) programska oprema kot 'obravnavalec', ki vodi podatke skozi celoten proces ETL.

Kot je prikazano na sliki (Slika 2), mikrostoritev pridobi podatke iz Enhacerja in ponavlja ta postopek, dokler niso vsi podatki pridobljeni kot JSON-LD iz izvirne končne točke GraphQL.

V praksi je ta mikrostoritev naloga znotraj Airflowa, v kateri je Enhancer obravnavan zaporedno z omejeno kodo, rezultati pa so zbrani kot povezani podatki.

Ko so vsi podatki potrjeni in naloženi v shrambo trojčkov, ki je v tem primeru primerek *TriplyDB* (<https://triplydb.com>), je mogoče uporabiti (instancirati) različne storitve, vključno z *ElasticSearch*, podatkovnim brskalnikom, končno točko SPARQL [13] za uporabo v podatkovnih zgodbah.

Te je mogoče uporabiti znotraj vmesnika same shrambe trojčkov. Zaradi boljše dostopnosti povezanih podatkovnih modelov so podatkovni modeli za posamezen register ključev tudi vizualizirani z orodjem *Weaver* (<https://kadasster.wvr.io/bag2-0/home>).

Čeprav trenutna arhitektura še ne vključuje pretočnega pristopa k povezani objavi podatkov, je bila sama arhitektura zasnovana tako, da bo ta pristop sčasoma mogoč.

Ključni dejavnik pri zagotavljanju te funkcije pretakanja za povezane podatke pri Geodetski upravi Kadaster je povečana razpoložljivost teh nizov podatkov na Datahub v zemljiški administraciji.

Ker je vizija, da je to *Datahub* osrednji vir vseh naborov katastrskih podatkov, bi moralo Datahub izpolnjevati tudi številna načela, vključno s ponudbo podatkov v skoraj realnem času.

Za zagotavljanje teh načel v svoji zasnovi bo verjetno za vsak nabor podatkov vključen *pretočni API*, ki omogoča poizvedbe na podlagi vleka in na podlagi naročnine.

Povezovanje »cevovoda«, opisanega v tem poglavju, s tem pretočnim API-jem prek GraphQL bo omogočilo transformacijo v realnem času in dostavo povezanih podatkov za vsak osnovni register v isto okolje Datahub.

1.2.4 Vizija za integracijo geoprostorskih podatkov

Medtem, ko so napredek v tehnologijah in standardih povezanih podatkov ter povečano povpraševanje po teh storitvah sprožili potrebo po posodobljenem pristopu k zagotavljanju povezanih geoprostorskih podatkov, je ta pristop zdaj tudi v središču ambicija Geodetske uprave po dostavi grafa znanja [14].

V arhitekturi rešitve, predstavljeni v prejšnjem razdelku, je prvi korak k temu grafu znanja narejen tako, da so različni ključni registri na voljo kot povezani podatki.

Ti nabori podatkov, ki si prizadevajo ohraniti povezani podatkovni model čim bližje izvirnemu podatkovnemu modelu, so še vedno specifični za domeno.

To ima prednost za izolirane povezane nabore podatkov, ker so podatkovni modeli prepoznavni in dostopni strokovnjakom za področje, tehnika upravljanja pa je jasna pri upravljanju teh podatkov s strani lastnika nabora podatkov; dejavnikov, ki so pomembni za interno upravljanje in lastništvo podatkov v različnih formatih znotraj Geodetske uprave.

Pristop „blizu vira“, za objavo povezanih podatkov pa ni uporaben, če si prizadevamo narediti podatke, vsebovane v ključnih registrih, bolj dostopne neizkušenemu uporabniku, ker *uporaba zahteva nekaj znanja o domeni*.

V praksi neizkušene uporabnike takšnih podatkovnih virov bolj zanimajo informacije o njihovih domovih ali bivalnem okolju in si prizadevajo za interakcijo s podatki na način, ki jim je znan glede na njihove okoliščine.

Kot tak se pri zagotavljanju *grafa znanja, ki je rezultat združevanja registrov ključev* in drugih naborov podatkov, transformiranih z uporabo prej opisane arhitekture rešitve, za *graf znanja Geodetske uprave* uporablja nov, uporabniku prijazen podatkovni model.

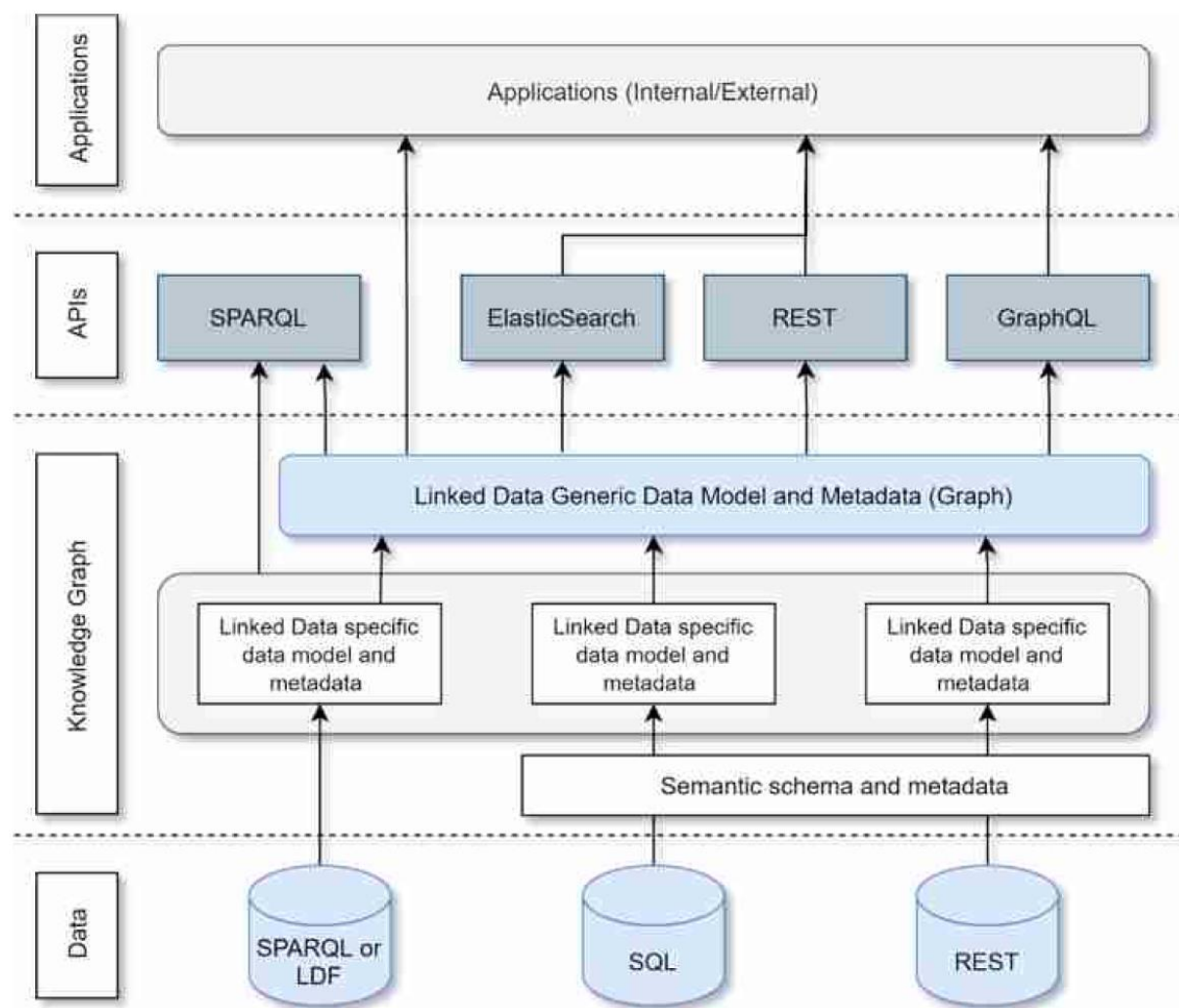
Vsebina grafa znanja so povezani nabori podatkov za vsak ključni register, digitalni katastrski načrt in drugi ustrezni nabori podatkov, *osredotočeni na temo stavbe*.

Tehnični postopek združevanja ločenih naborov podatkov za oblikovanje grafa znanja je prikazan na sliki (Slika 3). Kot prikazuje slika (Slika 3), je graf znanja Geodetske uprave dostavljen z *ustvarjanjem plasti na vrhu ključnih registrov v njihovi povezani podatkovni obliki*.

Kombinacija teh povezanih registrov podatkov se trenutno izvaja z definiranjem podatkovnega modela z uporabo specifikacij *schema.org* v zvezi z *zgradbami*.

Uporaba *schema.org* v prvi objavi grafa znanja geodetske uprave je narejena iz treh razlogov.

- a) Prvič, z arhitekturo grafa znanja kot sloja na ločenih, povezanih naborih podatkov, ki sestavljajo graf, je *izvor izvirnih naborov podatkov po potrebi še vedno na voljo končnemu uporabniku* grafa znanja.
- b) Drugič, uporaba specifikacij *schema.org* poteka v interesu *ponovne uporabe obstoječih standardov* skupnosti ter v interesu podpore zunanje odkritosti in interoperabilnosti.
Poleg tega je specifikacija *schema.org* v zvezi s pojmom »stavba« na splošno veliko bližje temu, kako neizkušeni uporabniki dojemajo svoje okolje.
Dejansko lahko uporabnik pri iskanju podatkov o hiši razmišlja o poštni številki in hišni številki in ne le o identifikacijski številki v informacijskem sistemu.
Kot tak se podatkovni model izkaže za veliko bolj uporabniku prijaznega kot kompleksni podatkovni modeli, ki definirajo ključne registre.
Dostop do trenutnega grafa znanja je na voljo prek *storitev REST, GraphQL, GeoSPARQL [15] in ElasticSearch*, pri čemer jih aplikacije tretjih oseb uporabljajo za dostavo geoprostorskih podatkov končnemu uporabniku [16].



Slika 3 Vizija implementacije grafa znanja Geodetske uprave (vir: Kadaster).

Čeprav je bila uporaba specifikacije schema.org uporabna pri razvoju dokaza koncepta za graf znanja Geodetske uprave Kadaster, je malo verjetno, da bo ta specifikacija končna specifikacija, uporabljena za podatkovni model.

To je večinoma posledica dejstva, da *prihodnja ponovitev grafa znanja* verjetno ne bo vsebovala le informacij o stavbah, zgradbah, *specifikacija schema.org pa (še) ni dovolj široka*, da bi dosledno modelirala vse ustrezne geoinformacijske objekte (npr. ceste, infrastruktura, topografske značilnosti). V prihodnjih iteracijah grafa znanja bo Geodetska uprava Kadaster tesno sodeloval z nacionalnimi organi za standardizacijo in uporabniškimi skupinami pri definiranju podatkovnega modela, ki je semantično dovolj bogat, da natančno združuje in odraža kompleksnost in bogastvo informacij, vsebovanih v ključnih registrih, hkrati pa združuje podatke na tak način, da se ohrani uporabniku prijazna narava *kombiniranega podatkovnega modela*.

1.2.5 Vrednotenje in pridobljena spoznanj

Pri izvajanju pristopa, opisanega v tem dokumentu, so bili obravnavani različni izzivi v zvezi z *objavo povezanih podatkov Geodetske uprave*. Ker je bil začetni pristop razvit v kontekstu veliko manj

povezanih podatkovnih zbirk in izdelkov, je bilo pogosto treba *razviti infrastrukturo, treba za objavo povezanih podatkov v organizaciji*.

V preteklosti se je nabralo precejšnja količina podedovane infrastrukture, katere vzdrževanje je zahtevalo veliko virov. Sčasoma ta infrastruktura tudi ni odražala razvoja, ki je bil sprejet v širši povezani podatkovni skupnosti v zvezi s podatkovnimi zbirkami in tehnologijami.

Zahteva po bolj razširljivem pristopu zaradi omrežnih učinkov, povezanih z večjo razpoložljivostjo povezanih podatkov, je zahtevala, da se postopek objave povezanih podatkov pri Geodetski upravi inovira.

Za zagotovitev te razširljivosti in učinkovitosti virov se je novi pristop naslonil na *obstoječe standarde skupnosti in komercialne tehnološke rešitve za razvoj časovnice objave* v skladu s splošnimi arhitekturnimi načeli, ki poudarjajo potrebo po *vključitvi obstoječih tehnologij, kjer koli je to mogoče*.

Kot je navedeno v okoliščinah problema, je prenova obstoječega pristopa v korist pristopa, opisanega v tem dokumentu, privedla do objave in vzdrževanja več velikih in zapletenih naborov podatkov s strani relativno majhne notranje ekipe v Geodetski upravi.

To se doseže z večjim poudarkom na skupnosti in *ponovni uporabi komercialnih standardov* in izdelkov, ki zmanjšujejo količino sredstev, porabljenih za razvoj in *vzdrževanje povezane infrastrukture za objavo podatkov*.

Poleg načel, kot je *konfiguracija nad kodo*, kot tudi *objava nabora podatkov "bližje viru"* z upoštevanjem *pretočnega pristopa*, novi pristop znatno zmanjša kompleksnost in intenzivnost virov prejšnjih pristopov.

Čeprav je bil prejšnji pristop k objavi povezanih podatkov uspešen v času njegovega začetka, je razvoj povezanih tehnologij sčasoma spodbudil potrebo po inovacijah in navsezadnje razvoj pristopa, razvitega in predstavljenega tukaj.

Uspešna objava različnih ključnih registrov in zunanjih naborov podatkov ponuja priložnost za razmislek o pristopu, ki ga je ubrala Geodetska uprava Kadaster.

Tu je mogoče identificirati številne "naučene lekcije", od katerih je vsako mogoče grobo razvrstiti v dve temi.

Prva tema

Prvo temo je mogoče povzeti kot *potrebo po boljšem ujemanju med poslovnim in tehničnim vidikom povezanih podatkov* in vključuje dve pridobljeni lekciji.

- a. Prvič, pristop, predstavljen v tem dokumentu, je *tehnično lahka, poceni in nizko kodirana* (z uporabo obstoječih programskih rešitev) *izvedba za objavo povezanih podatkov*. To se veliko bolje ujema s poslovnim vidikom in načeli arhitekture, ki se na splošno upoštevajo v organizacijah.

Na primer, *nizki stroški implementacije* dopolnjujejo dejstvo, da končni uporabniki trenutno uporabljajo te storitve relativno v majhnem obsegu.

Poleg tega uporaba obstoječih orodij v objavljanju dopolnjuje *ponovno uporabo obstoječih programskih rešitev* v organizacijah, ki temeljijo na splošnih načelih arhitekture »dobre prakse«.

- b. Druga lekcija, ki se je je treba naučiti v okviru te teme, se nanaša na *splošno pomanjkanje upravljanja*, ki obkroža obstoječe pristope k povezani objavi podatkov v organizacijah.

Pristop, predstavljen v tem dokumentu, jasno *razlikuje med objavo izoliranih povezanih naborov podatkov – iz silosa*, torej naborov podatkov, ki so modelirani in implementirani čim bližje izvirnemu podatkovnemu modelu, kolikor je tehnično mogoče, *in vizijo za implementacijo grafa znanja kot plasti na vrhu teh nizov podatkov*.

To razlikovanje z vidika upravljanja omogoča lastniku ločenega nabora podatkov, da *ostane odgovoren za povezano (LD) podatkovno različico svojega silosa*, ker v samih podatkih (modelu) ni prišlo do bistvenih sprememb, kar *omogoča, da graf znanja spada pod drug model upravljanja*. Ta razdelitev podpira učinkovito upravljanje povezanih podatkov v organizaciji, tako s tehničnega kot poslovnega vidika.

Druga tema

Druga tema se nanaša na *vprašanje podatkovnega modeliranja grafov znanja* v vladnem kontekstu.

Prva implementacija grafa znanja pri Geodetski upravi Kadaster je bila izvedena z razvojem podatkovnega modela, ki temelji na schema.org. Kot je bilo omenjeno v prejšnjem razdelku, obstajajo različne *pomanjkljivosti tega podatkovnega modela*, kar je privedlo do priznanja, da je ta *integrirani podatkovni model potreben in bi ga morala razviti in vzdrževati Geodetska uprava sama* – to bo bolj verodostojen vir, kot pa je organ za standarde.

Modeliranje podatkov v vladnem kontekstu je seveda lahko nekoliko zapleteno zaradi različnih pravnih in upravljavskih zahtev, ki obkrožajo podatkovni model.

Razvoj novega, integriranega podatkovnega modela za graf znanja lahko traja dolgo in vprašanje, kdo naj bo odgovoren za razvoj, dostavo in vzdrževanje modela, ostaja odprto.

Kot taka je lahko največja ovira za dostavo integriranega podatkovnega modela v vladnem kontekstu *kompleksnost okoli dejavnosti modeliranja podatkov*; ovira, ki bi lahko odložila široko uporabo teh grafov v prihodnjih letih.

S tem pregledom dobrih praks zaključujemo predstavitev rezultatov Naloge 1.1 Analiza mogočih kombinacij semantičnih tehnologij za učinkovito večkratno uporabo (R1.1) in nadaljujemo s predstavitvijo standardov na področju semantičnih tehnologij v okviru Naloge 1.2.

1.3 Naloga 1.2: Analiza aktualnih odprtokodnih in drugih standardov (R1.2)

Standardi so bistveni za odprte geoprostorske podatke in številne odprtokodne geoprostorske programske rešitve izvajajo odprte standarde. Odprti standardi so tehnološko nevtralne specifikacije za strojno opremo, programsko opremo ali podatke, razvite z odprtim postopkom. Odprte standarde za geografske informacije (trenutno jih je bilo objavljenih več kot 100) tradicionalno razvijajo tri ključne mednarodne organizacije za razvoj standardov, katerih standarde smo proučili za potencialno uporabo v končnih usmeritvah te raziskovalne naloge. V tej nalogi smo pregledali trenutno stanje odprtokodnih standardov geoprostorske programske opreme in semantičnih podatkov. Odprtost podatkov je spremenila način zbiranja, obdelave, analize, vizualizacije in objavljanja geoprostorskih podatkov.

1.3.1 »Plačljivi« ISO standardi za področje obravnave geoprostorskih metapodatkov

ISO standardizacija geoprostorskih podatkov je izvedena in dopolnjena v štirih dopolnjujočih se delih, opredeljenih v nadaljevanju.

ISO 19115 Geografske informacije — Metapodatki

ISO 19115-1:2014 1. del: Osnove – sprememba 2 (Part 1: Fundamentals — Amendment 2, 2020)

Ta 1. del standarda ISO 19115 opredeljuje *shemo*, treba za *opisovanje* geografskih informacij in storitev s pomočjo *metapodatkov*. Zagotavlja informacije o identifikaciji, obsegu, kakovosti, prostorskih in časovnih vidikih, vsebini, prostorski referenci, prikazu, distribuciji in drugih lastnostih digitalnih geografskih podatkov in storitev.

Ta del standarda ISO 19115 se uporablja za:

- *katalogiziranje* vseh vrst virov, dejavnosti klirinške hiše in popoln opis naborov podatkov in storitev;
- geografske *storitve*, nizi geografskih podatkov, serije nizov podatkov ter posamezne geografske značilnosti in lastnosti elementov.

Ta 1. del standarda ISO 19115 opredeljuje:

- obvezne in neobvezne/pogojne odseki metapodatkov, entitete metapodatkov in elemente metapodatkov;
- *minimalni* (obvezni) nabor metapodatkov, potrebnih za uporabo večine metapodatkovnih aplikacij (*odkrivanje* podatkov, ki določa *primernost* podatkov za uporabo, *dostop* do podatkov, *prenos* podatkov ter *uporabo* digitalnih podatkov in storitev);
- *neobvezni* metapodatkovni elementi, ki omogočajo obsežnejši standardni opis virov, če je to treba;
- *metoda za razširitev/prilagoditev* metapodatkov, da ustrezajo posebnim potrebam.

Čeprav se ta del standarda ISO 19115 uporablja za digitalne podatke in storitve, je njegova načela mogoče razširiti na številne druge vrste virov, kot so *zemljevidi*, *diagrami* in *besedilni dokumenti* ter ne *geografski podatki*. Določeni pogojni metapodatkovni elementi morda ne veljajo za te druge oblike podatkov.

ISO19115-2:2019/Amd 1:2022 Del 2: Razširitve za pridobivanje in obdelavo — sprememba 1
(Geographic information — Metadata — Part 2: Extensions for acquisition and processing — Amendment 1)

Ta dokument razširja del 1. Osnove (ISO 19115-1:2014) z opredelitvijo sheme, potrebne za *izboljšan opis pridobivanja in obdelave* geografskih informacij, vključno s posnetki. Vključene so *lastnosti merilnih sistemov* ter *numeričnih metod* in računskih postopkov, ki se uporabljajo za pridobivanje geografskih informacij iz podatkov, pridobljenih z njimi. Ta dokument zagotavlja tudi kodiranje XML za pridobivanje in obdelavo metapodatkov, s čimer se *razširijo sheme XML*, opredeljene v ISO/TS 19115-3.

ISO 19115-3:2023 3. del: Implementacija sheme XML za temeljne koncepte (Part 3: XML schema implementation for fundamental concepts)

Standard opredeljuje integrirano izvedbo XML standardov ISO 19115-1 in ISO 19115-2 z opredelitvijo naslednjih dejstev:

- nabor shem XML, potrebnih za validacijo dokumentov instanc metapodatkov, ki so v skladu z elementi konceptualnega modela, opredeljenimi v ISO 19115-1 in ISO 19115-2; in
- nabor pravil ISO/IEC 19757-3 (Schematron), ki izvajajo omejitve preverjanja v modelih UML ISO 19115-1 in ISO 19115-2, ki niso potrjeni s shemo XML.

Ta del opisuje postopek, uporabljen za *ustvarjanje shem XML iz konceptualnih modelov* geografskih informacij ISO, povezanih z metapodatki. Sheme XML so generirane *neposredno iz konceptualnega modela UML* (8.5).

ISO 19115-4 4. del: Implementacija sheme JSON za osnove metapodatkov (2024 v razvoju) (Part 4: JSON schema implementation of metadata fundamentals)

4. del standarda opisuje *postopek, uporabljen za ustvarjanje sheme JSON iz konceptualnih modelov* geografskih informacij ISO, povezanih z metapodatki.

Aplikacijska shema *ne spreminja semantike izvornih konceptualnih modelov*, ampak *dodaja konstrukcije* in prilagaja strukturo, treba za učinkovito implementacijo JSON, in elemente UML za olajšanje dela programske opreme za preoblikovanje JSON. Sheme JSON so bile samodejno ustvarjene iz modela UML za implementacijo JSON v skladu s pravili ISO 19115-1, ISO 19115-2 in 19157-1.

Naša analiza aktualnih odprtokodnih in drugih spletnih standardov na področju odprtih geoprostorskih podatkov in programskih rešitev pa podrobneje obravnava predvsem odprte standarde najbolj uveljavljenih organizacij za standardizacijo objavljanih podatkov na spletu, kar navajamo v nadaljevanju.

- Standardi konzorcija W3C Semantic web (Web of Data)
- OGC standardi

1.3.2 Standardi konzorcija W3C Semantic web

Semantični splet je pobuda konzorcija W3C za predstavitev znanja na svetovnem spletu na *strojno berljiv način*, tako da ga lahko *stroji razumejo in uporabljajo* za inteligentne aplikacije. Glavna področja Semantičnega spleta so: povezani podatki (Linked Data), besednjaki, poizvedbe, sklepanje in vertikalne aplikacije.

Poleg klasičnega »spleta dokumentov« W3C pomaga graditi tehnološki sklad za podporo »spletu podatkov«, vrsti podatkov, ki jih najdete v spletnih bazah podatkov. Končni cilj spleta podatkov je računalnikom omogočiti bolj učinkovito delo in razviti sisteme, ki lahko podpirajo zaupanja vredne interakcije prek omrežja.

Izraz "semantični splet" se nanaša na vizijo W3C o spletu povezanih podatkov. Tehnologije semantičnega spleta ljudem omogočajo *ustvarjanje podatkovnih shramb na spletu, ustvarjanje besednjakov in pisanje pravil za ravnanje s podatki*. Povezane podatke podpirajo *standardizirane tehnologije*, kot so na primer RDF, OWL, SPARQL, RDFa, JSON-LD, SKOS, SHACL.

Konzorcij svetovnega spleta (W3C) vzdržuje standarde za RDF, vključno s temeljnimi koncepti, semantiko in specifikacijami za različne formate. Prva sintaksa, definirana za RDF, je temeljila na razširljivem označevalnem jeziku (XML). Zdaj se pogosteje uporabljajo *druge sintakse*, vključno s trojnim jezikom Terse RDF (Turtle), zapisom objektov JavaScript za povezane podatke (JSON-LD) in N-trojčki.

1.3.3 Standardi OGC

Odprti geoprostorski konzorcij (OGC) je zveza strokovnjakov, ki se zavzema za izboljšanje dostopa do geoprostorskih ali lokacijskih informacij. Povezuje ljudi, skupnosti in tehnologijo za reševanje globalnih izzivov in vsakodnevnih potreb.

Standardi OGC so sredstvo za medopravilnost geoprostorskih informacij in jih uporablja na tisoče organizacij po vsem svetu ter so predstavljeni v milijonih vrstic kode. Podpirajo jih mednarodne organizacije, uporabljajo se v predlogih in izvajajo za pospešitev procesa inovacij.

OGC je središče geoprostorskih inovacij, sodelovanja in standardov. Je mednarodna članska organizacija, ki podpira raznoliko skupnost več kot 500 podjetij, vladnih agencij, raziskovalnih organizacij in univerz, ki vsi skupaj delajo, da bi bile informacije o lokaciji »POŠTENE« (FAIR)—najdljive, dostopne, medopavilne in ponovno uporabne (FAIR – Findable, Accessible, Interoperable, and Reusable).

Izvedbeni standardi se razlikujejo od abstraktnih specifikacij. Izvedbeni standardi so napisani za bolj tehnično občinstvo in podrobno opisujejo strukturo vmesnika med komponentami programske opreme.

Šteje se, da je *specifikacija vmesnika* na ravni izvedbe s podrobnostmi skladna, če se pri izvajanju dveh različnih inženirjev programske opreme, ki se ne zavedata drug drugega, nastale komponente medsebojno vklopijo in delujejo na tem vmesniku.

Izvedbeni standardi OGC

Vse sheme (XSD, shema JSON itd.), ki podpirajo odobreni standard implementacije, je mogoče najti v uradnem repozitoriju shem OGC.

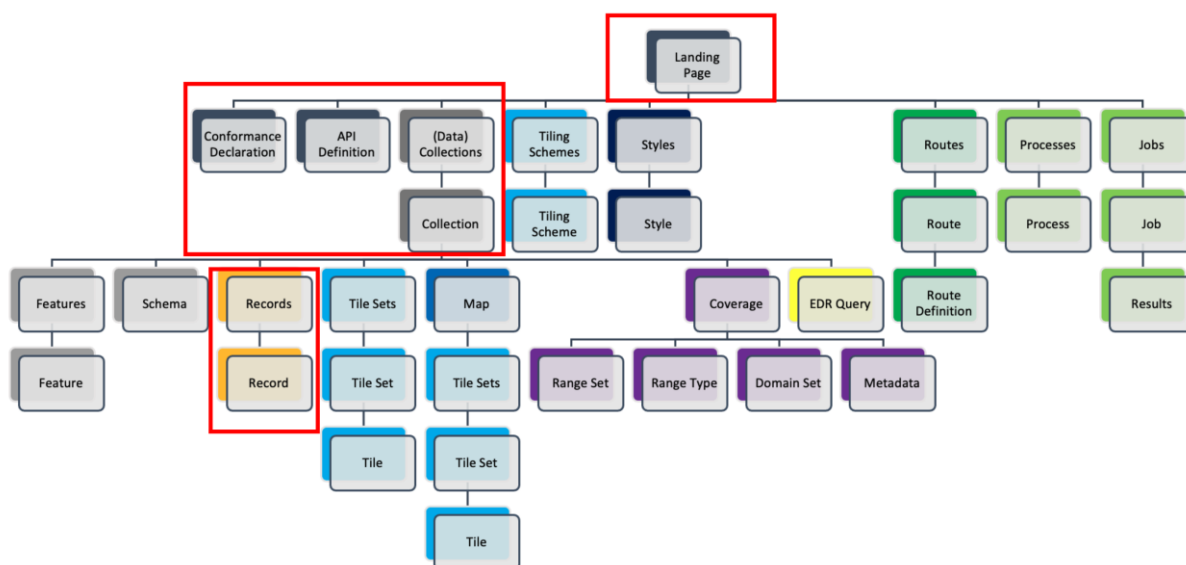
Abstraktna specifikacija OGC

Tehnični odbor OGC (TC) je razvil arhitekturo v podporo svoji viziji geoprostorske tehnologije in medopravilnosti podatkov, imenovano "OGC Abstract Specification". *OGC abstraktna specifikacija* zagotavlja konceptualno podlago za večino dejavnosti razvoja specifikacij OGC. Odprti vmesniki in protokoli so zgrajeni in se sklicujejo na abstraktne specifikacije, kar omogoča medopravilnost med različnimi blagovnimi znamkami in različnimi vrstami sistemov prostorske obdelave. Abstraktna specifikacija zagotavlja *referenčni model za razvoj standardov* za izvajanje OGC arhitekture.

OGC API - Records za odkrivanje geoprostorskih podatkov in storitve

Geodetski upravi predlagamo uporabo standardov za izboljšano odkrivanje geoprostorskih podatkov. Eden od pomembnejših je OGC API Zapisi (Records).

OGC API - Records (Slika 4) je večdelna specifikacija, ki ponuja možnost ustvarjanja, spreminjanja in poizvedovanja po metapodatkih na spletu. Del 1, Core (osnutek) zajema dostop samo za branje do zapisov in zmožnosti preprostega poizvedovanja. Zmogljivosti za bogatejše poizvedbe ali ustvarjanje, posodabljanje ali brisanje zapisov bodo določene v dodatnih delih.



Slika 4 Velika slika na OGC API – je vključno z Records (rdeči okvirčki) za izboljšanje vidnosti geoprostorskih podatkov in storitev (vir:OGC)

OGC API Records zagotavlja:

- Standardiziran dostop do metapodatkov: dostop do zapisov metapodatkov in poizvedb za geoprostorske podatke in storitve, kar uporabnikom olajša odkrivanje ustreznih informacij.
- Dosledno odkrivanje podatkov: zapisi metapodatkov sledijo dosledni strukturi in obliki.
- Bogatejši metapodatki: podrobni metapodatki o geoprostorskih virih, vključno z informacijami o kakovosti podatkov, prostorskem obsegu, časovni pokritosti in več.

- Medopravilnost: do metapodatkov je mogoče preprosto dostopati in jih razumeti širok nabor aplikacij in storitev
- Spletni dostop: standardne zahteve in odgovori HTTP

OGC API Records podpira naslednje izhodne formate:

HTML, XML (ISO19139, ISO19115-3), JSON, RSS, DCAT, Turtle, Schema.org.

OGC Kataloška storitev za splet (Catalogue Service for the Web - CSW)

Kataloška storitev za splet (CSW) je OGC standard za izpostavljanje kataloga geoprostorskih zapisov v XML na internetu (prek HTTP). Katalog je sestavljen iz zapisov, ki opisujejo geoprostorske podatke (npr. KML), geoprostorske storitve (npr. WMS) in povezane vire.

CSW je en del (ali "profil") kataloške storitve OGC, ki definira skupne vmesnike za odkrivanje, brskanje in poizvedovanje po metapodatkih o podatkih, storitvah in drugih potencialnih virih. Različica 2.0 specifikacije je bila izdana maja 2004. Najnovejša izdaja je 2.0.2, ki je bila objavljena leta 2007. Med drugimi operacijami omogoča tudi pridobivanje »obiranje ali žetev« (harvesting) s tem, da ustvarite/posodobite metapodatke tako, da od strežnika zahtevate, da metapodatke 'povleče' od nekod.

Dejavnost skupnosti OGC v zadnjih letih na področju povezanih (LD) podatkov

Dejavnost "NetCDF" (2022)

Open Geospatial Consortium (OGC) išče javne komentarje o kandidatu za standard OGC »*Encoding Linked Data Graphs in netCDF Files Standard*« (netCDF-LD). NetCDF-LD je pristop za sestavo opisov povezanih podatkov z uporabo metapodatkov in struktur v datotekah netCDF.

NetCDF je format za kodiranje matrično usmerjenih znanstvenih podatkov, zlasti na številnih področjih znanosti o Zemlji in vesolju. Običajna praksa je pomoč pri medopravilnosti datotek netCDF v znanostih o Zemlji z uporabo konvencije o podnebnju in napovedovanju (CF) in tudi konvencije o atributih za odkrivanje podatkov (ACDD). Več skupnosti opredeljuje dodatne konvencije netCDF za opisovanje semantike v zvezi z njihovimi različnimi domenami. Ko se širi sočasna uporaba več – in morda nasprotujočih si opredelitev, se povečuje težava pri iskanju skupnega mehanizma za preverjanje in interpretacijo metapodatkov. Kandidat za standard netCDF-LD ponuja rešitev za ta problem, vdelano v datoteke netCDF.

Povezani podatki (LD) so najboljša praksa konzorcija svetovnega spleta za kodiranje in objavljanje metapodatkov za izpostavljanje in povezovanje podatkov znotraj posameznih nizov podatkov in med več nizi podatkov. LD tudi omogoča, da postanejo opisi naborov podatkov bolj uporabni z izvajanjem semantično omogočenih poizvedb med nabori podatkov. Povezani podatki uporabljajo W3C *Resource Description Framework (RDF)* za semantično izražanje informacij in odnosov. Objavljanje opisov datotek netCDF z metodami LD daje uporabniški skupnosti priložnost, da izkoristi izboljšano izpostavljenost in povezavo LD ter posledično naredi objavljene podatke bolj dostopne, interoperabilne in ponovno uporabne (FAIR).

Kandidat za standard netCDF-LD zagotavlja standard za kodiranje semantike povezanih podatkov v datoteke netCDF in interpretacijo datotek netCDF kot grafov RDF z minimalnimi ali majhnimi spremembami obstoječih datotek netCDF. Datoteke NetCDF, ki sprejemajo standard netCDF-LD in so sestavljene v grafe RDF, se lahko uporabljajo z drugimi tehnologijami povezanih podatkov za združevanje podatkov, zmogljivo poizvedovanje in odkrivanje podatkov. NetCDF-LD omogoča tudi zajem izboljšanih metapodatkov netCDF, kar omogoča, da se informacije v datotekah netCDF povežejo

z objavljenimi opredelitvami in nadzorovanimi besednjaki, ki se uporabljajo za izražanje vsebine, s čimer se izboljša semantična medopravilnost in doslednost.

Dejavnost Veliki povezani geoprostrski podatki 2017 (Big Linked Geodata) delavnica tipa lokacijske moči (Location Powers)

Poudarek delavnice je bil raziskati, kako lahko pristopi velikih podatkov (BigData) pomagajo pri učinkovitem izkoriščanju povezanih geografskih podatkov.

Količina podatkov, ki jih ustvarja človeštvo, eksponentno narašča, pri čemer je bilo 90 % podatkov, ki so trenutno na svetu, ustvarjenih v zadnjih dveh letih [1]. Poleg tega naj bi se skupna količina globalnih podatkov v naslednjem desetletju povečala za 40 % iz leta v leto [1]. Večina teh podatkov ima lokacijsko komponento. *Razvoj semantičnega spleta* omogoča povezovanje podatkov na podlagi geografskih informacij na način, ki zagotavlja dragocen vpogled. Zmožnost pravilne uporabe prostorske razsežnosti na področju velikih podatkov je zato ključna in bo imela koristi od kombinacije standardov OGC in pristopa semantičnega spleta.«.

Cilj: pospeševanje uporabe računalniških tehnik "povezanih" in "velikih" podatkov, ki se uporabljajo za geoprostorske podatke.

O dogodkih »lokacijske moči« (Location Powers)

Dogodki »Location Powers« združujejo strokovnjake zasebnega in javnega sektorja z vsega sveta in so zasnovani kot interaktivne razprave med skupnostmi, pri čemer je vsak od njih do zdaj pripravil bele knjige, ki vodijo industrijo. Poleg tega, da so mesto učenja za vse o moči lokacije, so vrhovi zasnovani tudi za pomoč pri informiranju o prihodnji usmeritvi in delu v konzorciju za nadaljnji razvoj ustreznih in uporabnih standardov.

Dva organa za standarde, W3C in OGC tesno sodelujeta pri ustvarjanju spleta prostorskih in povezanih podatkov saj sta za to področje združena v delovno skupino »Spatial Data on the Web Working Group« (SDWWG).

na povezavi:

<https://www.locationpowers.net/events/1703delft/index.php#conversation>

Nizozemske praktične izkušnje (dogodek »Veliki povezani podatki« iz leta 2022):

Zahvaljujoč standardizaciji OGC in aktivni lokalni geodomeni imajo na Nizozemskem veliko aplikacijskih shem in podatkov GML. Upali so, da bodo lahko podatke in modele samodejno preoblikovali v povezana podatkovna standarda RDF in OWL.

Za GML podatke je bilo to uspešno in pravzaprav precej enostavno zaradi GML-jevega objektnih lastnosti modela, ki se ujema s povezanim podatkovnim modelom. Toda za UML modele je bilo to problematično. Izkazalo se je, da je prišlo do spremembe paradigme, povezane z načelom "Vsakdo lahko reče karkoli o čemer koli". Uspelo jim je sicer preslikati lastnosti modela UML v standardizirane povezane podatkovne besednjake in pretvoriti UML v OWL, vendar dobljeni modeli niso bili prijazni do povezanih podatkov, zato so to pot opustili.

Osnovne topografske baze kot povezani podatki

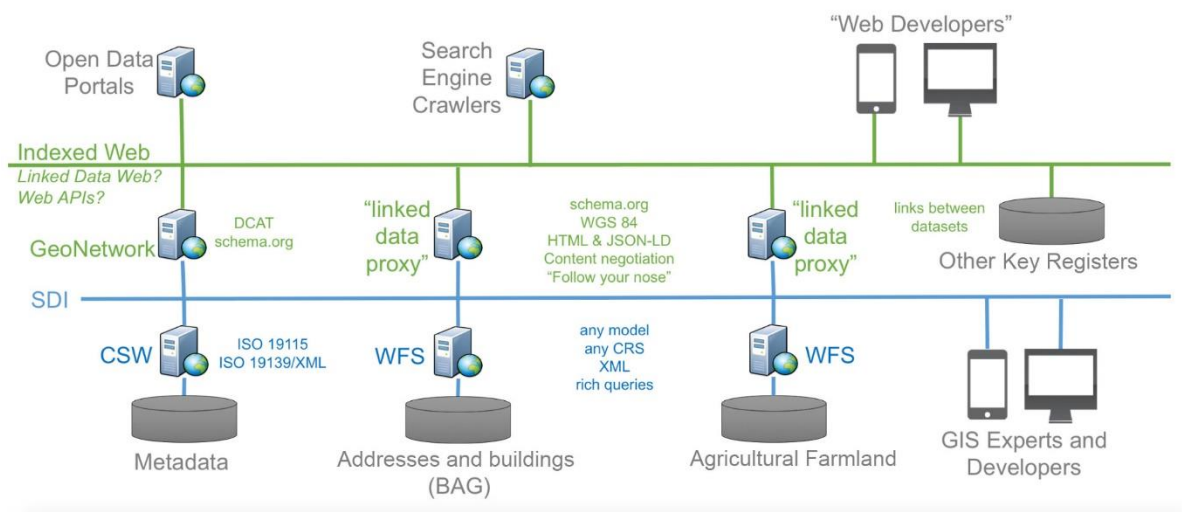
Eksperimentiranje s povezanimi podatki je postalo resnejše, ko so od nizozemskega Ministrstva za infrastrukturo in okolje dobili proračun za pilotni projekt osnovnega registra topografije velikega merila kot povezanih podatkov. Te podatke so povezali z drugimi vladnimi (negeografskimi) nabori podatkov

ter dobro znanimi spletnimi viri, jih shranili v shrambo trojčkov in uporabili GEOSPARQL kot posrednik med podatki in spletnim mestom, ki je preprosto za uporabo. To je jasno pokazalo, da povezovanje podatkov pomaga končnim uporabnikom pri brskanju po povezanih podatkih, pomaga pa tudi razvijalcem – spletno mesto je bilo preprosto ustvariti, ker so bili podatki že povezani.

Medtem je v nizozemski LD skupnosti potekalo veliko zanimivih eksperimentov. Kadaster je postal aktiven, kar je sčasoma vodilo do objave nabora treh uradnih baz kot povezanih podatko. Davčni urad je uporabljal te povezane podatke za boj proti goljufijam. Razvita je bila tako imenovana »pralnica LOD«, ki lahko dela z ogromnimi količinami povezanih podatkov. Povezani podatki so bili uporabljeni za objavo nizozemske zakonodaje in imigracijskih postopkov. Proizvajalec mikročipov NXP ga je uporabil za upravljanje podatkov o izdelkih podjetij. Izdelani so bili obsežni besednjaki za podatke o kulturni dediščini in grajenem okolju. In povezani podatki so bili uporabljeni v številnih raziskovalnih projektih o distribuciji energije (CERISE), farmaciji (OpenPHACTS).

V istem obdobju se je fokus začel širiti. Nekateri člani skupnosti so se pritoževali o kompleksnosti povezanih podatkov, trojnih shramb in končnih točk SPARQL. Z objavo priročnih API-jev pa so želeli nuditi podatke na voljo širšemu krogu razvijalcev. Oblikovali so šest stopenjsko (razširjeno) lestvico medopravilnosti za razvijalce (standardnemu naboru so dodali raven spletni servis v JSON API oziroma spletno storitev s kontekstom JSON-LD API).

To je pripeljalo do ideje za preskusno platformo Geonovum, v kateri so izbranim strankam zagotovili sredstva za raziskave in preizkus objave prostorskih podatkov na spletu – to niso bili le povezani podatki, ampak je bil poudarek širši, na drugih spletnih standardih in najboljših praksah. Eden od rezultatov je bila metoda in orodje za objavo povezanih podatkov in spletnih podatkov poleg trenutnega SDI (Slika 5).



Slika 5 Preskusna platforma Geonovum (vir: Van den Brink 2022)

Eksperimentirali so tudi z objavo velikih prostorsko povezanih podatkov. To je predstavljalo težavo večini trenutno razpoložljivih trgovin trojčkov. Združevanje prostorskih in administrativnih komponent (npr. katera bližnja podjetja imajo več kot 100 zaposlenih?) v eni poizvedbi je povzročala velike težave pri delovanju. Prav tako je manjkala podpora za GeoSPARQL. Geodetska uprava Kadaster je ta vprašanja uspešno reševala z upoštevanjem naslednjih načel in pristopov ter standardov:

- Podatki ustrezajo standardom W3C
- »Pristop« do URI-jev, semantike, izvora v skladu z najboljšimi praksami W3C

- Med seboj povezani podatki
- Najdljivi (indeksirani) prek iskalnikov (schema.org)
- Razvijalcem prijazen API
- Portal za razvijalce
- Vizualizacija povezanih podatkov
- Samodejni ETL (ekstrakcija, transformacija, nalaganje) povezanih podatkov in API-jev
- Posebni API-ji za na primer Zakon o varstvu okolja
- Končna točka GeoSparql
- Oblikovalnik poizvedb SPARQL

2016 je nizozemski PLDN organiziral tematski dan povezanih geografskih podatkov v Evropi. Kadaster je predstavil svojo platformo povezanih podatkov, ki objavlja več povezanih podatkovnih nizov, vključno z naslovi, zgradbami, topografijo in katastrskimi parcelami. JRC je predstavil zamisli za povezane podatke INSPIRE. Predstavljene so bile francoske, finske in švedske pobude za povezane podatke. In nenazadnje se je skupnosti predstavila delovna skupina za prostorske podatke na spletu.

Decembra istega leta so bili povezani geografski podatki spet tema. Tokrat so se Irska, Švica in Belgijska Flandrija izkazale za precej napredne pri povezani objavi geopodatkov.

INSPIRE kot povezani podatki

Konceptualni podatkovni modeli INSPIRE:

- temelji na skupnem okvirju modeliranja in med področnega usklajevanja,
- prostorski informacijski objekti in njihove lastnosti ter razmerja za 34 podatkovnih tem,
- upravljajo se v skupnem skladišču UML.

Pomisleki o izvedljivosti in prednostih povezanih podatkov INSPIRE (Lutz, JRC EC, 2022)

- Konvencionalna orodja in skupnosti SDI (OGC) pogosto niso pripravljeni na izkoriščanje povezanih podatkov in semantike.
- Orodja in pravila skupnosti semantičnega spleta niso zelo jasni za splošno skupnost in orodja spletnih razvijalcev.
- Pogosto se trdi, da imajo povezani (geoprostorski) podatki koristi, vendar pogosto *manjkajo resnični dokazi o koristih* (v primerjavi s konvencionalnimi pristopi SDI).

Zato je treba

- ugotoviti možnost izvedbe in prednosti predstavitve INSPIRE v RDF,
- izdelati osnutek smernic za objavo podatkov INSPIRE v RDF,
- izdelati vrednostne predloge geoprostorskih podatkov kot povezanih odprtih podatkov.

Nastajajoče smernice in predlagani besednjaki INSPIRE RDF bodo dokumentirani v repozitoriju (Guidelines for the RDF encoding of spatial data).

OGC standard za kodiranje "JavaScript Object Notation" (JSON)

JavaScript Object Notation (JSON) standardna notacija postaja vse bolj priljubljena za kodiranje podatkov v spletnih aplikacijah. JSON je sestavljen iz nizov predmetov, opisanih s pari, ime/vrednost. Ta standard OGC opisuje kodiranje GeoJSON in JSON-LD za metapodatke opazovanja Zemlje (EO) za nize podatkov (granule). Ta standard je mogoče uporabiti za kodiranje metapodatkov na podlagi profila metapodatkov opazovanja Zemlje za opazovanja in meritve (O&M) OGC 10-157r4 ali kot kodiranje konceptualnega modela poenotenega modela metapodatkov za granule (Unified Metadata Model for Granules - UMM-G).

Kodiranje GeoJSON, je opredeljeno kot zgoščevanje predlaganega kodiranja JSON-LD skozi normativni kontekst z nekaterimi razširitvami. Zato je mogoče kodiranje JSON-LD uporabiti tudi za druga kodiranja RDF, vključno z RDF XML in RDF Turtle.

Ta dokument ne predpostavlja nobenih predpostavk glede "storitvenih" vmesnikov, prek katerih se dostopa do metapodatkov, in velja enako dobro za storitveno usmerjeno arhitekturo kot tudi za "na vire usmerjeno arhitekturo" ali RESTful arhitekturo. Dokumentiran pristop je mogoče uporabiti v kombinaciji z naslednjimi tehnologijami:

- razširitve OGC OpenSearch,
- Povezana podatkovna platforma W3C,
- OASIS searchRetrieve,
- OASIS OData.

JSON je človeku berljiv in zlahka razčlenljiv. Vendar je JSON brez sheme. Dokumenti JSON in GeoJSON ne vključujejo eksplicitne definicije strukture objektov JSON, ki jih vsebujejo. Zato ta standard temelji na normativnem kontekstu JSON-LD, ki omogoča, da je vsaka lastnost eksplicitno definirana kot URI. Poleg tega je kodiranje JSON definirano s shemo JSON [OR18], ki omogoča preverjanje primerkov glede na te sheme.

OGC standardizirano kodiranje GeoJSON

GeoJSON je format za kodiranje zbirk preprostih geografskih značilnosti skupaj z njihovimi neprostorskimi atributi z uporabo JSON. Informacijski Objekti GeoJSON lahko predstavljajo geometrijo, funkcijo ali zbirko funkcij. GeoJSON podpira naslednje *tipe geometrije*, izpeljane iz *specifikacije OGC Simple Features*: Point, LineString, Polygon, MultiPoint, MultiLineString, MultiPolygon in GeometryCollection. *Funkcije v GeoJSON* vsebujejo geometrijski objekt in dodatne lastnosti, *zbirka funkcij* pa predstavlja seznam funkcij.

GeoJSON(-LD) format za serializacijo povezanih podatkov (Linked Data)

GeoJSON(-LD) je besednjak in kontekst JSON-LD za GeoJSON. Specifikacija formata definira JSON-LD 1.1 format, ki temelji na JSON za serializacijo povezanih podatkov. Sintaksa je zasnovana za enostavno integracijo v nameščene sisteme, ki že uporabljajo JSON, in zagotavlja gladko pot nadgradnje iz JSON v JSON-LD. Predvsem naj bi bil način za uporabo povezanih podatkov v spletnih programskih okoljih, za gradnjo interoperabilnih spletnih storitev in shranjevanje povezanih podatkov v mehanizme za shranjevanje, ki temeljijo na JSON. Opisuje nadnabor funkcij, opredeljenih v JSON-LD 1.0 [JSON-LD10], in razen kjer je navedeno drugače, dokumenti, ustvarjeni z različico 1.0 te specifikacije, ostanejo združljivi z JSON-LD 1.1.

OGC Standard GeoJSON(-LD) za metapodatke opazovanj Zemlje (EO) (Dataset Metadata GeoJSON(-LD) Encoding Standard)

Ta standard OGC definira kodiranje GeoJSON [NR2] in JSON-LD [NR3] metapodatkov Opazovanja Zemlje (EO) za nabore podatkov (tj. izdelke ali granule). Standard zagotavlja modele dokumentov za izmenjavo informacij, ki opisujejo nabore podatkov EO, znotraj in med različnimi organizacijami.

Standardni model je izpeljan iz konceptualnih modelov, opredeljenih v profilu metapodatkov opazovanja Zemlje za opazovanja in meritve (O&M) OGC 10-157r4 [OR1] in poenotenega modela metapodatkov za granule (UMM-G) [OR2],

W3C standardizirano kodiranje RDF (Resource Description Framework - RDF)

Kodiranje RDF je standardni *model za izmenjavo podatkov na spletu*. Kodiranje RDF ima funkcije, ki olajšajo združevanje podatkov, tudi če se osnovne sheme razlikujejo, in posebej podpira razvoj shem skozi čas, ne da bi bilo treba spremeniti vse uporabnike podatkov.

Kaj je RDF?

Ogrodje/okvir za opis virov (RDF) je splošno ogrodje za *predstavitev med seboj povezanih podatkov na spletu*. Stavki RDF se uporabljajo za *opisovanje in izmenjavo metapodatkov*, kar omogoča standardizirano izmenjavo podatkov na podlagi relacij. RDF se uporablja za *integracijo podatkov iz več virov*. Primer tega pristopa je spletno mesto, ki prikazuje spletne kataloge proizvajalca in povezuje izdelke z ocenami na različnih spletnih mestih in s trgovci, ki prodajajo izdelke. Semantični splet temelji na uporabi ogrodja RDF za *organiziranje informacij na podlagi pomenov*. Stavki RDF izražajo razmerja med viri, na primer naslednje:

- dokumenti
- fizičnimi predmeti
- ljudmi
- abstraktnimi pojmi
- podatkovnimi objekti

Zbirke povezanih izjav RDF sestavljajo usmerjen graf, ki preslika razmerja med entitetami. Zbirko izjav RDF o povezanih entitetah lahko uporabite za izdelavo *grafa RDF*, ki prikazuje, kako so te entitete povezane.

Kodiranje RDF tako razširja povezovalno strukturo spleta za *uporabo URI-jev za poimenovanje razmerja* med stvarmi in obema koncema povezave (to se običajno imenuje "trojček"). Z uporabo tega preprostega modela omogoča, da se strukturirani in pol strukturirani *podatki mešajo, izpostavijo in delijo med različnimi aplikacijami*.

Ta *povezovalna struktura* tvori *usmerjen, označen graf*, kjer robovi predstavljajo *imenovano povezavo* med dvema viroma, ki ju predstavljajo vozlišča grafa. Ta grafični pogled je najlažji možni mentalni model za kodiranje RDF in se pogosto uporablja v lahko razumljivih vizualnih razlagah.

Kako deluje RDF?

RDF je standardni način za podajanje izjav o virih. Izjava RDF je sestavljena iz treh komponent, imenovanih **trojka**:

- **Subjekt** je vir, ki ga opisuje trojka.
- **Predikat** opisuje razmerje med subjektom in predmetom.
- **Objekt** je vir, ki je povezan s predmetom.

Subjekt in objekt sta vozlišči, ki predstavljata stvari. Predikat je lok, ker predstavlja razmerje med vozlišči.

Standard RDF predvideva *tri različne vrste vozlišč*:

1. **Enotni identifikator vira (URI)** je standardizirana oblika za identifikacijo vira, abstraktnega ali fizičnega. Enotni lokator virov (URL) je vrsta URI-ja, ki se pogosto uporablja v stavkih RDF. Ko je W3C leta 2014 posodobil specifikacijo RDF na različico 1.1, je dodal *internacionalizirani identifikator virov (IRI)* kot vrsto vozlišča. IRI so podobni in dopolnjujejo URI, kar *omogoča uporabo mednarodnih naborov znakov*.
2. **Literal** je določena *podatkovna vrednost* in je lahko niz, datum ali številka vrednost. Dobesedne vrednosti so izražene z uporabo formata URI ali IRI.
3. **Identifikator praznega vozlišča** je znan tudi kot anonimni vir ali bnode. Predstavlja subjekt, o katerem ni znano nič drugega kot odnos. Identifikatorji praznih vozlišč uporabljajo posebno sintakso za njihovo identifikacijo.

Več stavkov RDF o isti entiteti bodo vsi veljavni trojčki RDF. Imajo isti subjekt, vendar različne predikate in objekte. Pri gradnji grafa RDF iz teh trojčkov je subjekt lahko prikazan enkrat, z več puščicami, ki se razvejajo iz subjekta in predstavljajo različne predikate in različne objekte.

Kako se uporablja RDF?

Stavke RDF je mogoče vključiti v spletne strani Hypertext Markup Language (HTML) ali shraniti v ločene datoteke in jih povezati s podatki v spletni vsebini. Ko je bil RDF prvič določen, so bili stavki RDF vključeni v dokumente XML, povezane s spletno vsebino. Čeprav je bila v preteklosti sintaksa XML edina izbrana sintaksna možnost, standardi za kodiranje stavkov RDF zdaj vključujejo naslednje *tri sintakse*:

- **Turtle** je najbolj priljubljena sintaksa besedila za stavke RDF. W3C ga opisuje kot "kompaktno in naravno besedilno obliko", ki vključuje *okrajšave* za pogosto uporabljene vzorce.
- **JSON-LD** uporablja *sintakso JSON* za stavke RDF.
- **N-Triples** je podmnožica sintakse Turtle, zasnovana kot enostavnejša besedilna oblika za stavke RDF za *izboljšano enostavnost uporabe s strani ljudi*, ki pišejo izjave. Enostavnejša oblika tudi olajša programom ustvarjanje in razčlenjevanje stavkov RDF.

Poizvedovalni jezik RDF se uporablja za dostop in upravljanje informacij, shranjenih v grafih RDF. Poizvedovalni jeziki RDF morajo biti sposobni razčlenjevanja trojčkov RDF ter interpretacije in ustvarjanja rezultatov, povezanih z vsebino trojčkov, kot tudi odnosov med trojčki.

Prednosti RDF

Semantični splet je odvisen od odprtega in medopravilnega standarda za izmenjavo podatkov in metapodatkov. To je tisto, kar RDF zagotavlja in je razlog, zakaj je bil prvič standardiziran. Prednosti RDF vključujejo naslednje:

- Okvir RDF *spodbuja dosledno skupno rabo metapodatkov* o internetnih virih.
- Standardne sintakse RDF za opisovanje in poizvedovanje podatkov omogočajo programski opremi, ki uporablja metapodatke, lažje delo.
- *Standardna sintaksa* in zmožnost poizvedb omogočata aplikacijam lažjo izmenjavo informacij.

- Iskalci dobijo *natančnejše rezultate z iskanjem na podlagi metapodatkov*, kot bi jih dobili z indeksi, pridobljenimi z zbiranjem celotnega besedila.
- Inteligentni *programski agenti imajo natančnejše podatke za delo in so bolj natančni* pri tem, kar zagotavljajo uporabnikom.
- Za razliko od relacijskih baz podatkov lahko *podatki RDF zagotovijo veliko bolj zanimive informacije* o entiteti in njenih odnosih z drugimi entitetami.

Omejitve standardizirane notacije RDF

S svojimi odprtimi specifikacijami in široko uporabnostjo je RDF uspešen standard za objavo semantičnih spletnih podatkov in sestavljanje semantičnih informacij. RDF ima nekaj omejitev, od katerih so nekatere posledica njegove priljubljenosti.

- Standardizacija besedišča za opis virov RDF je lahko težavna. *Pobuda za metapodatke »Dublin Core«* je pomembno prizadevanje za standardizacijo elementov in izrazov RDF za skupne entitete in attribute.
- Izbira najprimernejše oblike sintakse za stavke RDF *zahteva vajo*. Različni formati so morda lažji za uporabo za posebne izvedbe, zlasti za sisteme, ki generirajo stavke RDF.
- Izbira najprimernejšega poizvedovalnega jezika RDF za aplikacijo je odvisna od značilnosti poizvedovalnega jezika in posebnih potreb aplikacije.

RDF Turtle

Besedilna sintaksa za RDF, imenovana Turtle (Terse RDF Triple Language), ki omogoča, da se *graf RDF v celoti zapiše v kompaktni in naravni besedilni obliki, z okrajšavami* za običajne vzorce uporabe in tipe podatkov. Turtle zagotavlja ravni združljivosti s formatom N-Triples [N-TRIPLES] kot tudi sintakso trojnega vzorca priporočila SPARQL (W3C Recommendation). Komentariji so lahko podani za oznako

' # ', in se nadaljujejo do konca vrstice.

Notacija RDF XML

Notacija RDF XML opisuje, kako kodirati grafe RDF v XML sintaksi.

SPARQL – W3C standard za poizvedovalne jezike RDF

SPARQL, rekurzivna kratica za protokol SPARQL in poizvedovalni jezik RDF, je *standard za poizvedovalne jezike RDF*, ki ga je razvil W3C. Poizvedbeni jeziki, skladni s SPARQL, delujejo samo na trojčkih RDF in ne podpirajo poizvedb na shemah RDF. **Sheme RDF** so razširitve osnovnega besednjaka RDF, ki definirajo standardizirane vire in odnose.

GeoSPARQL – geografski poizvedovalni jezik za podatke RDF

Standard OGC GeoSPARQL podpira *predstavljanje in poizvedovanje* geoprostorskih podatkov v semantičnem spletu. *GeoSPARQL definira besednjak za predstavitev geoprostorskih podatkov v kodiranju RDF* in definira razširitev poizvedovalnega jezika SPARQL za obdelavo geoprostorskih podatkov. Poleg tega je GeoSPARQL zasnovan za prilagoditev sistemov, ki temeljijo na kvalitativnem prostorskem sklepanju, in sistemov, ki temeljijo na kvantitativnih prostorskih izračunih.

“Trgovine RDF trojčkov” (RDF Triple Store)

Trgovine RDF trojčkov (RDF triplestore) je vrsta baze podatkov o grafih, ki spletno shranjuje podatke kot mrežo objektov (z materializiranimi povezavami med njimi) in uporablja sklepanje za odkrivanje novih informacij iz obstoječih odnosov. Njegena prilagodljiva in dinamična narava omogoča povezovanje različnih podatkov, njihovo indeksiranje za semantično iskanje in obogatitev z analizo besedila za izdelavo velikih grafov znanja (knowledge graph - KG).

Shranjuje **semantična dejstva**, zapisana v notaciji RDF. Zaradi tega so RDF triplestores prednostna izbira za upravljanje visoko povezanih podatkov. Trgovine trojčkov so na primer bolj prilagodljive in cenejše kot relacijska zbirka podatkov.

Podatkovna baza RDF, pogosto imenovana “baza podatkov semantičnega grafa”, je prav tako sposobna obravnavati močne semantične poizvedbe in uporabljati sklepanje za odkrivanje novih informacij iz obstoječih odnosov.

Označeni graf lastnosti (Labeled Property Graphs – LPG)

Označeni graf lastnosti (LPG) je vrsta baze podatkov grafov. V zbirkah podatkov v slogu LPG je graf sestavljen iz vozlišč in odnosov. Vsako vozlišče ali razmerje ima oznako ID, eno ali več "oznak", ki opisujejo njegov tip ali razred, in nabor "lastnosti", ki predstavljajo vrednosti, in ustrezen ključ, ki omogoča sklicevanje. Intuitivno sta dve vozlišči vedno povezani z razmerjem, da tvorita večjo strukturo grafa. Za baze podatkov v slogu LPG ne obstaja standardiziran poizvedovalni jezika, vendar je jezik Cypher v praksi najbolj razširjen.

Besednjak podatkovnega kataloga DCAT

Besednjak podatkovnega kataloga DCAT (Data Catalog Vocabulary - DCAT) je v razvojni fazi W3C *Delovnega osnutka* (z dne 7.3.2023), ki je objavljen za pregled in preprosto zgodovinsko referenco. Standardni dokument W3C je razvit z zaporednimi delovnimi osnutki, s pričakovanjem, da bo sčasoma postal kandidati za Priporočila. Vsebinsko je besednjak DCAT za *objavljanje podatkovnih katalogov na spletu*, ki je bil prvotno razvit v okviru vladnih podatkovnih katalogov, kot je data.gov, vendar je uporaben tudi v drugih kontekstih. Imenski prostor za izraze DCAT je <http://www.w3.org/ns/dcat#>. Predlagana predpona za imenski prostor DCAT je dcat.

Zgodovina DCAT

Prvotni besednjak DCAT 1 je bil razvit in gostovan na Inštitutu za raziskave digitalnih podjetij (DERI), nato ga je izpopolnila interesna skupina eGov in končno **standardizirala leta 2014** [VOCAB-DCAT-1] delovna skupina za povezane podatke vlade (GLD).

Drugo priporočeno revizijo DCAT, **DCAT 2** [VOCAB-DCAT-2], je razvila delovna skupina za izmenjavo naborov podatkov kot *odgovor na nov niz primerov uporabe in zahtev* [DCAT-UCR], zbranih na podlagi izkušenj ljudi z besediščem DCAT iz časa prvotne različice in nove aplikacije, ki niso bile upoštevane v prvi različici.

Tretjo različico **DCAT 3**, je razvila delovna skupina za izmenjavo naborov podatkov ob upoštevanju nekaterih bolj perečih primerov uporabe in zahtev med tistimi, ki so v prejšnjem standardizacijskem krogu ostale neobravnavane.

Profil aplikacij GeoDCAT-AP (Aplikacijski Profil – AP za podatkovne portale v Evropi)

GeoDCAT-AP je *razširitev* DCAT-AP za opis naborov geoprostorskih podatkov, nizov naborov podatkov in storitev. Zagotavlja vezavo sintakse RDF za združevanje elementov metapodatkov, opredeljenih v osrednjem profilu standarda ISO 19115:2003, in tistih, opredeljenih v okviru direktive INSPIRE. Njegov osnovni primer uporabe je *omogočanje iskanja po naborih prostorskih podatkov, serijah podatkov in storitvah na splošnih podatkovnih portalih*, s čimer je zagotovljeno izboljšano iskanje po geoprostorskih informacijah prek mednarodnih meja in sektorjev. To lahko dosežemo z izmenjavo opisov podatkovnih nizov med podatkovnimi portali.

Specifikacija GeoDCAT-AP ne nadomešča Uredbe o metapodatkih INSPIRE niti tehničnih smernic za metapodatke INSPIRE, ki temeljijo na standardih ISO 19115 in ISO19119. Njegov namen je podati upravljavcem geoprostorskih metapodatkov možnost, da dosežejo več povezljivosti z zagotavljanjem dodatne sintaksne vezave RDF.

(vir: https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en/)

SHACL (W3C) Jezik omejitev oblik (Shapes Constraint Language - SHACL)

SHACL je jezik za preverjanje grafov RDF glede na nabor pogojev. Ti *pogoji* so podani kot oblike in drugi konstrukti, izraženi v obliki grafa RDF. Grafi RDF, ki se uporabljajo na ta način, se v SHACL imenujejo "*grafi oblik*", grafi RDF, ki so potrjeni glede na graf oblik, pa se imenujejo "*grafi podatkov*". Ker se grafi oblik SHACL uporabljajo za *preverjanje*, *ali podatkovni grafi izpolnjujejo nabor pogojev*, jih je mogoče obravnavati tudi kot *opis podatkovnih grafov*, *ki izpolnjujejo te pogoje*. Takšni opisi se lahko uporabljajo za različne namene poleg *preverjanja*, vključno z gradnjo uporabniškega vmesnika, ustvarjanjem kode in *integracijo podatkov*.

W3C Spletni ontološki jezik OWL

Web Ontology Language (OWL) je jezik semantičnega spleta, jezik znanja o stvareh, zasnovan za predstavitev bogatega in kompleksnega *znanja o stvareh, skupinah stvari in odnosih med stvarmi*. OWL je jezik, ki temelji na računalniški logiki, tako da lahko znanje, izraženo v OWL, izkoristijo računalniški programi, npr. za preverjanje doslednosti tega znanja ali za eksplicitno implicitno znanje. Dokumenti OWL, znani kot *ontologije*, se lahko objavijo v svetovnem spletu in se lahko *nanašajo na druge ontologije OWL* ali se nanje nanašajo iz njih. OWL je del nabora tehnologij semantičnega spleta W3C, ki vključuje RDF, RDFS, SPARQL itd.

SWRL (Semantic Web Rule Language)

Jezik pravil semantičnega spleta SWRL je neposredna razširitev OWL 2 DL s pod naborom RuleML in je zasnovan kot jezik pravil semantičnega spleta. Ontologija OWL v abstraktni sintaksi vsebuje zaporedje aksiomov in dejstev. Aksiomi so lahko različnih vrst, npr. aksiomi podrazreda in aksiomi enakovrednega razreda. Predlaga se, da se to dejstvo razširi z aksiomi pravil.

SKOS (W3C) sistem organizacije preprostih znanj (simple knowledge organization system SKOS)

Sistem SKOS razvija specifikacije in standarde za podporo uporabi *sistemov organizacije znanja (KOS)*, kot so *tezavri*, *klasifikacijske sheme*, *sistemi predmetnih oznak in taksonomije* v okviru semantičnega spleta.

W3Cstandard SKOS gradi most med svetom sistemov organizacije znanja - vključno s besednjaki (tezavri), klasifikacijami, predmetnimi naslovi, taksonomijami in folksonomijami - in skupnostjo spletnih povezovalcev podatkov (Linked Data), kar prinaša koristi obema. Knjižnice, muzeji, medijske hiše, vladni portali, podjetja, aplikacije za družabna omrežja in druge skupnosti, ki upravljajo velike

zbirke knjig, zgodovinskih artefaktov, novic, poslovnih glosarjev, vnosov v spletne dnevnike in drugih predmetov, lahko zdaj uporabljajo sistem organizacije preproste znanja (SKOS) za izkoriščanje moči povezanih podatkov LD.

GraphQL jezik za vmesnike

Grafi so močno orodje za modeliranje številnih pojavov v resničnem svetu, ker so podobni našim naravnim miselnim modelom in verbalnim opisom osnovnega procesa. GraphQL je odprtokodni jezik za poizvedovanje in obdelavo podatkov za API-je in mehanizem za izvajanje poizvedb. GraphQL je poizvedbeni jezik za API-je in strežniško izvajalno okolje za izvajanje poizvedb z uporabo sistema tipov, ki ga določite za svoje podatke. GraphQL ni vezan na nobeno specifično zbirko podatkov ali mehanizem za shranjevanje, temveč je podprt z vašo obstoječo kodo in podatki.

Z GraphQL svojo poslovno domeno modeliramo kot z grafom za definiranje sheme. Znotraj svoje sheme definiramo različne vrste vozlišč in kako se med seboj povezujejo/povezujejo. Na odjemalcu to ustvari vzorec, ki je podoben objektu usmerjenemu programiranju: vrste, ki se sklicujejo na druge vrste. Ker GraphQL na strežniku definira samo vmesnik, ga lahko uporabljate s katerim koli zaledjem (novim ali starejšim!). Storitve GraphQL je ustvarjena z definiranjem vrst in polj na teh vrstah, nato pa zagotavljanjem funkcij za vsako polje na vsaki vrsti.

Poimenovanje stvari je zahteven, a pomemben del gradnje intuitivnih API-jev, zato si vzemite čas in skrbno premislite, kaj je smiselno za vašo problemsko domeno in uporabnike. Vaša ekipa bi morala razviti skupno razumevanje in soglasje teh pravil poslovne domene, ker boste morali izbrati intuitivna, trajna imena za vozlišča in razmerja v shemi GraphQL. Razmišljajte o svoji shemi GraphQL kot o izraznem skupnem jeziku za vašo ekipo in uporabnike. Če želite ustvariti dobro shemo, preučite vsakdanji jezik, ki ga uporabljate za opis svoje organizacije.

Sloj poslovne logike

Kje opredeliti trenutno poslovno logiko? Kje bi morali izvajati validacijske in avtorizacijske preglede? Odgovor: znotraj namenske plasti poslovne logike. Plast poslovne logike bi morala delovati kot edini vir resnice za uveljavljanje pravil poslovne domene.

Včasih pri delu s podedovanimi viri podatkov ugotovimo, da ne ti odražajo popolnoma načina, kako stranke uporabljajo podatke. V teh primerih bolje zgraditi shemo GraphQL, ki opisuje, kako odjemalci uporabljajo podatke, namesto zrcaljenja podedovane sheme baze podatkov. Zgradite svojo shemo GraphQL, da izrazite "kako" namesto "kaj". Nato lahko izboljšate svoje podrobnosti implementacije, ne da bi pokvarili obstoječi vmesnik s starejšimi odjemalci.

GraphQL omogoča deklarativno pridobivanje podatkov, kjer lahko odjemalec natančno določi, katere podatke od API-ja potrebuje. Namesto več končnih točk, ki vrnejo ločene podatke, strežnik GraphQL izpostavi eno samo končno točko in se odzove s točno tistimi podatki, ki jih zahteva odjemalec. Ker lahko strežnik GraphQL pridobi iz ločenih virov podatkov in predstavi podatke v poenotenem grafu, ni vezan na nobeno specifično bazo podatkov ali mehanizem za shranjevanje.

Apollo GraphQL

Sistem Apollo GraphQL je zgrajen na osnovnem odprtokodnem odjemalcu in strežniku GraphQL podjetja. Platforma ponuja razvijalska orodja in storitve za pospešitev razvoja, varno infrastrukturo in prilagajanje podatkovnega grafa več ekipam. Cilj Apolla je zagotoviti en sam vir informacij za strukturo grafa s sledenjem shemi v centralnem registru. Register sheme deluje kot središče, ki usklajuje vse sisteme in orodja, ki sestavljajo graf, in ohranja zasebnost sheme.

Platforma Apollo uporabnikom pomaga zgraditi, poizvedovati in upravljati podatkovni graf: poenoteno podatkovno plast, ki aplikacijam omogoča interakcijo s podatki iz katere koli kombinacije povezanih shramb podatkov in zunanjih API-jev. Podatkovni graf je med odjemalci aplikacije in zalednimi storitvami ter olajša izmenjavo podatkov med njimi.

Grafi in variante

Graf v Apollo Studiu predstavlja povezan podatkovni graf organizacije. Vsak graf ima eno ali več različic, ki ustrezajo okoljem, kjer graf deluje, kot sta prikazovanje in produkcija. Vsaka različica ima svojo shemo GraphQL, kar pomeni, da se lahko sheme med okolji razlikujejo.

Razporejeni grafi

Razporejene grafe delijo z drugimi člani organizacije. Primerni so za vse primere uporabe, razen za lokalni razvoj.

Grafi razvoja

Razvojni grafi (ali razvijalski grafi) so vidni samo primarnemu uporabniku v njegovem razvojnem okolju. Uporabljajo se lahko za ponavljanje grafa v razvojnem okolju, ne da bi motili ali opozarjali druge uporabnike.

Apache Airflow

Apache Airflow je odprtokodno orodje za programsko ustvarjanje, razporejanje in spremljanje delovnih tokov. Je ena najbolj robustnih platform, ki jih uporabljajo podatkovni inženirji za usmerjanje delovnih tokov ali cevovodov. Z lahkoto lahko vizualizirate odvisnosti svojih podatkovnih cevovodov, napredek, dnevnike, kodo, sprožilne naloge in stanje uspeha.

Apache Airflow je za avtomatizacijo delovnih tokov, ki temelji na Pythonu in se uporablja za nastavitvev in vzdrževanje zmogljivih podatkovnih cevovodov. Samo po sebi ni orodje ETL, vendar upravlja, strukturira in organizira cevovode ETL z uporabo usmerjenih acikličnih grafov (Directed Acyclic Graphs - DAG). Apache Airflow je že leta nedvomno najbolj priljubljen odprtokodni projekt za podatkovni inženiring.

Elasticsearch

Elasticsearch je distribuirani iskalnik in analitični mehanizem, zgrajen na Apache Lucene. Od izdaje leta 2010 je Elasticsearch hitro postal najbolj priljubljen iskalnik in se pogosto uporabljajo za analitiko dnevnikov, iskanje po celotnem besedilu, varnostno obveščanje, poslovno analitiko in primere uporabe operativnih obveševalnih podatkov.

Elasticsearch je zasnovan z Javo in je baza podatkov NoSQL. To pomeni, da shranjuje podatke na nestrukturiran način in da ne morete uporabiti SQL za poizvedovanje po njih.

WeaVER

Vizualizacijsko orodje WeaVER za preučevanje več prostorskih značilnosti.

3D Weaver

3D Weaver profesionalcem omogoča shranjevanje, ogled in upravljanje velikih, visokokakovostnih 3D sredstev v spletnem brskalniku. Pretvarja velike 3D-modele za progresivne funkcije spletnega gledanja, kar zagotavlja, da se 3D-sredstva naložijo hitro in v izboljšani obliki. 3D Weaver je optimiziran za potrebe strokovnjakov za kulturno dediščino, saj jim omogoča dodajanje opomb in pridobivanje meritev (razdalja, višina, površina, koti) v 3D modele.

Uporabniki lahko naložijo 3D podatke v orodje in uporabijo integriran spletni urejevalnik za dodajanje metapodatkov v skladu z vnaprej določenimi standardi, vključno z informacijami o kakovosti in licenciranju. Recenzenti lahko ocenijo vsak model, ki so ga uporabniki označili za objavo, in ga odobrijo.

Zaključek pregleda spletnih standardov za semantiko geoprostorskih podatkov

V tem razdelku smo izvedli pregled najpomembnejših odprtih spletnih standardov, ki zadevajo domeno semantike geoprostorskih podatkov. Razdelek predstavlja ključne rezultate analiz (R1.2), ki smo jih opravili v sklopu naloge Naloge 1.2: Analiza aktualnih odprtokodnih in drugih standardov.

Za zaključek poglavja navajamo še možnost ocenjevanja standardov za doseganje medopravilnosti s Splošno metodo ocenjevanja standardov in specifikacij (The Common Assessment Method for Standards and Specifications, CAMMS).

CAMSS je evropski vodnik za ocenjevanje in izbiro standardov in specifikacij za projekt e-uprave, referenca pri gradnji arhitekture in sredstvo za utemeljitev izbire standardov in specifikacij glede na potrebe in zahteve medopravilnosti. Popolnoma je usklajen z evropsko standardizacijsko uredbo 1025/2012.

Metoda CAMSS je okvir, zasnovan za ocenjevanje in spodbujanje uporabe odprtih in široko podprtih standardov in specifikacij in je lahko pomembno orodje za javne uprave, saj jih vodi pri izbiri in implementaciji rešitev IKT, ki so interoperabilne, varne in usklajene z ustreznimi okviri in predpisi.

<https://interoperable-europe.ec.europa.eu/collection/common-assessment-method-standards-and-specifications-camss>

2 Delovni sveženj 2 (DS2): Pregled predlogov semantičnih referenčnih sistemov za geografski prostor

V Delovnem svežnju 2 (DS2) smo v okviru naloge 2.1 obravnavali referenčne sisteme za tretjo vrsto interpretacije podatkov, in sicer tematskih. Po analogiji z geoprostorskimi in časovnimi referenčnimi sistemi bi morali semantični referenčni sistemi (SRS) omogočati preoblikovanje in projiciranje tematskih podatkov ali merjenje pomenske razdalje (imenovane semantična podobnost). Tipična aplikacija za tematski semantični referenčni sistem vključujejo preoblikovanje ali primerjavo različnih konceptualizacij med informacijskimi skupnostmi in iskanje informacij na splošno.

V okviru naloge 2.2 Izdelava konceptualne zasnove za ocenjevanje stopnje podobnosti podatkovnih prostorov smo opredelili nabor semantičnih relacij, ki bodo pokazale uporabnost v modelu tematskega semantičnega referenčnega sistema geografskega prostora, ko in če bo implementiran v strojno obliko. Predlagali smo razširitev trenutnih meril pomenske podobnosti z upoštevanjem prostorskih razmerij med različnimi geoprostorskimi koncepti. V nadaljevanju tega razdelka (DS2) najprej predstavljamo rezultate zasnove semantičnega referenčnega sistema.

2.1 Rezultati naloge 2.1: Izdelava konceptualne zasnove tematskega semantičnega referenčnega sistema (SRS)

Področje semantičnega rudarjenja oziroma umetne inteligence je v razvojni fazi, pri čemer se kaže kot najbolj obetaven pristop uporabe ontologij. Največji izziv je razvoj globalno povezljive in pomensko bogate specifične domenske ontologije. Vendar so obstoječi sezname ontologij nepovezani in nepopolni, ustreznost obstoječih ontologij pa je težko oceniti. Ustrezno opredeljena ontologija z dobro definirano preslikavo na relacijsko podatkovno bazo je jedro semantičnega rudarjenja. Cilj je uporaba pomenskega rudarjenja s strani uporabnikov, ki niso računalniški strokovnjaki, zato je treba podatke taksonomija oziroma ontologije prikazati na grafičen način tako, da bo uporabnik podatke razumel in znal sestaviti povpraševanje. Z ustreznim označevanjem dokumentov s pomočjo metapodatkov lahko dosežemo, da se semantično rudarjenje izvaja nad vsemi razpoložljivimi viri (Rupnik, Ojstršek, 2016).

Za semantično modeliranje je mogoče uporabiti pomensko povezane ključne besede obravnavane vsebine, gruč (grozde, klastre) ključnih besed, saj semantični »grozdi« lahko pomagajo zagotoviti, da je vsebina bolj razumljiva za najverjetnejše uporabnike.

Semantični model podatkov (na primer taksonomija) lahko izboljša semantično kakovost podatkov z zagotavljanjem jasne in dosledne strukture za organiziranje in predstavitev podatkov. Hierarhična struktura omogoča preprosto razumevanje in interpretacijo podatkov ter učinkovitejše poizvedovanje in obdelavo.

S kartiranjem tem in podtem obravnavane vsebine (na primer geografskega prostora) v referenčni semantični sistem lahko zagotovi, da je vsa obravnavana vsebina pomensko vzajemno povezljiva. To pomaga iskalnikom bolje razumeti okoliščine (kontekst) in namen, za katerega je bila obravnavana vsebina ustvarjena.

Geografski prostor je sestavljen iz enot, ki niso zgolj fizični objekti, ampak v pomembnem obsegu ustvarjeni v človeškem umu (Smith in Mark, 1998). Predstavljena je ontologija prostora geografskih razsežnosti (OPGR), na osnovi katere je mogoče izvajati pomensko integracijo podatkov zbirk prostorskih podatkov. OPGR vsebuje semantični slovar prostorskih izrazov. Pri izdelavi modela konceptualnega omrežja so upoštevani izsledki raziskav ontološkega modeliranja. V nadaljevanju poglavja so izvedene opredelitve mer za ocenjevanje semantike zbirk prostorskih podatkov. Izvedena je

analiza semantičnih razmerij aplikativnih konceptov v okviru OPGR. V sklepnem delu poglavja so nakazani metode in postopki semantičnega povezovanja zbirk prostorskih podatkov.

Razvita je bila metodologija posrednega povezovanja zbirk prostorskih podatkov. Metoda temelji na posredovalnem sistemu, ki omogoča prenašanje pomena med podatkovnimi zbirkami. Posredovalni sistem je sestavljen iz semantičnega referenčnega sistema ter semantičnih razmerij med semantičnim referenčnim sistemom in katalogi obravnavanih zbirk podatkov. Za razvoj semantičnega referenčnega sistema je bila razvita metodologija izdelave ontologije prostora geografske razsežnosti. Za ocenjevanje semantičnih razmerij pa je bila razvita metoda ocenjevanja semantike zbirk prostorskih podatkov.

Za razvoj metodologije posrednega povezovanja zbirk prostorskih podatkov so bile uporabljene nekatere predhodno razvite metode obravnavanja semantike zbirk podatkov oziroma modifikacije in njihova nadgradnja v novo razvito metodologijo. Razvita metodologija se sestoji iz štirih delov:

- metode izdelave ontologije prostora geografskih razsežnosti,
- metode ocenjevanja semantike zbirk prostorskih podatkov,
- metode ocenjevanja skladnosti med koncepti posameznih katalogov prostorskih podatkov ter
- metode izvedbe semantične povezave skladnih konceptov zbirk prostorskih podatkov.

2.1.1 Prostorski objekti geografskega obsega

Prostorski objekti geografskih razsežnosti so objekti, ki so večji od človeškega telesa in jih v celoti ni mogoče zaznati v okviru enega zaznavnega dejanja (Egenhofer in Mark, 1995). Geografski prostor je sestavljen iz enot, ki niso zgolj fizični objekti, ampak v pomembnem obsegu ustvarjeni v človeškem umu (Smith in Mark, 1998). V splošnem se ločuje med predmetnimi objekti in nepredmetnimi objekti geografskega obsega. Predmetni objekti obstajajo v fizični stvarnosti neodvisno od človeškega uma; primeri predmetnih objektov geografskega obsega so hiše, ceste, mostovi, utrdbe, gore, reke in morja. Nepredmetni objekti so ustvarjeni v človeškem umu; primeri nepredmetnih objektov so pravni objekti, na primer zemljiška posest, politične enote razdelitve prostora (Bittner in Frank, 1999).

Znanje o predmetnih objektih je pridobljeno z opazovanji in meritvami. S tehnikami opazovanj in meritev je mogoče pridobiti vedenje, znanje o približkih lastnosti in lokacij prostorskih objektov. Opazovanja in meritve lokacije predmetnih geografskih objektov podajajo znanje o približku lokacije v okviru lokalnih, regionalnih razdelitev, ki so bile ustvarjene ali izbrane s pomočjo opazovalnega ali merskega postopka (Bittner, 1999). Razlikovati je mogoče dve skupini predmetnih objektov geografskega obsega:

- predmetni objekti/pojavi z določenimi mejami,
- predmetni objekti/pojavi z nedoločenimi mejami.

Pojavi z določenimi mejami imajo ostre meje, njihovo lokacijo je mogoče natančno izmeriti, in sicer v okviru omejitev podrobnosti oziroma občutljivosti meritev. Primeri pojavov z določenimi mejami so hiše, mostovi, ceste in podobno. Položaja pojavov z nerazločnimi, nedoločenimi mejami ni mogoče natančno izmeriti, kajti meje so nerazločne, nedoločene, se premikajo ali jih ni mogoče opazovati. Primeri pojavov z nerazločnimi mejami so reke, gore, otoki. Nepredmetni objekti so konstrukt človeškega uma. Ustvarjeni so z opredelitvami. Ker ljudje ustvarjajo in določajo pravne objekte z nedvoumnimi opredelitvami, je znanje o lastnostih in lokacijah pravnih objektov natančno in popolno. Primer pravnega objekta je zemljiška posest in administrativno-politične enote razdelitve prostora. Za pravne objekte je jasna predstavitev natančne lokacije možna in nujno potrebna, za fizične objekte je lahko poznana in predstavljena le približna lokacija.

2.1.1.1 *Koncept lokacije v ontologiji prostora*

V zvezi z ontološkim in epistemološkim ujemanjem je mogoče ločevati med formalizacijo in predstavnostjo natančnih in približnih lokacij. Pojem lokacije se nanaša na razmerja med posameznimi enotami in referenčnim okvirom. Pojem lociranja znotraj referenčnega okvira dovoljuje abstrakcijo, ki posamezne individuum nadomesti z ustreznimi razredi. Skupina individuumov, ki ima isto lokacijo znotraj izbranega referenčnega okvira, tvori ustrezeni razred. Lokacijska abstrakcija pomeni obravnavo ustreznih razredov posameznih individuumov, ki si delijo isto lokacijo, ne pa obravnavo posameznih individuumov. Razlikovati je mogoče med natančno in približno lokacijo.

V primeru natančne lokacije vsebujejo vsi ustrezeni razredi iste lokacije samo posameznike, ki so natančno solocirani (Casati in Varazi, 1994, 1995). Dva individuum sta natančno solocirana, če zasedata isti del prostora. Predmetni objekti ne morejo biti solocirani. Dva različna predmetna objekta ne moreta imeti natančno iste lokacije v istem trenutku časa. Solocirani so lahko le objekti, ki so ontološko različni. Za veliko razredov geografskih objektov (gore, doline, področja onesnaženja, Alpe) ni mogoče poznati natančne lokacije, temveč je za takšne objekte mogoče poznati le približno lokacijo. Pojem približne lokacije temelji na referenčnih okvirih, ki predstavljajo regionalne razdelke obravnavanega prostora. Približna lokacija je opredeljena v smislu razmerij med prostorskimi objekti in razdelki obravnavanega prostora.

2.1.2 *Pravila za modeliranje ontologije prostora geografskih razsežnosti*

Pri razvoju formalne ontologije in njenih kategorij, razredov, je treba ohraniti čim manjše število razredov najvišje ravni. To zagotavlja jasnost razumevanja členjenja ontologije in enostavnost njene uporabe. Vsaka uspešna komunikacija zahteva jezik, ki je zgrajen nad jedrom skupnih konceptov. Podrobnejša opredelitev formalne ontologije je: ontologija je členjena, strukturirana, omejevalna zbirka nedvoumno opredeljenih konceptov. Obstajajo štiri temeljna pravila za modeliranje formalne ontologije (Kuhn, 1996): ontologija je zbirka skupnih konceptov; koncepti ontologije so nedvoumno opredeljeni, zbirka konceptov je omejevalna: koncepti, ki niso opredeljeni v ontologiji, niso uporabni; zbirka ima model členitve, kar pomeni, da ontologija vsebuje razmerja med koncepti.

Ontologije je mogoče graditi za tematske vsebine geografskih objektov kakor tudi za njihove razsežnostne koncepte. Ontologije razsežnostnih konceptov prostorskih objektov so geometrija, topologija ter simbologija predstavitev geografskih objektov (Rodriguez et al, 1996). Zbirka skupnih konceptov – ontologija domene – je torišče primerjav aplikacijskih ontologij, ki so izdelane za vsak geografski podatkovni niz. Primer področne zbirke skupnih zasnov je domena topografskega kartiranja, ki predstavlja prvo ontološko raven. Za tovrstno domeno so podane opredelitve topografskih objektov, kot so ceste, železnice, zgradbe in podobno (Ravi, 1995).

Druga raven so aplikacijske ontologije stvarnih geografskih podatkovnih nizov. V teh ontologijah so uporabljene besedne oznake za snemane oziroma kartirane zasnove (cesta, zgradba in podobno), vendar njihov natančni pomen v splošnem ni enak pomenu zasnov iz področne ontologije, v kateri so uporabljene enake ali podobne besedne oznake. Zaradi teh razlik je za vsak podatkovni niz, ki ga obravnavamo v postopku integracije, treba izdelati aplikacijske ontologije. Vključevalna pravila predstavljajo kriterije za zajem podatkov o objektih, določajo, kateri objekti bodo posneti. Predstavitvena pravila opredeljujejo, kako bodo objekti predstavljeni. Pravila poenostavljanja določajo, kako bodo stvarni objekti poenostavljeni v predstavitev objektov. Pravila združevanja opredeljujejo, kako bodo stvarni objekti združevani v predstavitev objektov.

2.1.2.1 *Opredelitev pravil semantične podobnosti*

Problem semantične medopravilnosti je odkrivanje semantično podobnih objektov, ki pripadajo različnim podatkovnim zbirkam (Rodriguez et al., 1999). Semantična razmerja med razredi ontologije domene in razredi ontologije aplikacije opredeljujejo semantiko obravnavanega področja obravnave.

Tako obstajata dve osnovni semantični razmerji (Sheth in Kashyap, 1993): prvo je razmerje enakovrednih razredov ontologije domene in razredov ontologije aplikacije; drugo je razmerje povezave dveh ali več razredov ontologije domene, tako da tvorita oziroma tvorijo v ontologiji aplikacije en sestavljeni agregatni razred.

Ti dve semantični razmerji določata, da mora biti ontologija domene dovolj bogata z zasnovami, da lahko obvlada vse koncepte ontologije aplikacij. Na osnovi dveh predhodno opredeljenih razmerij so opredeljena semantična razmerja med objektnimi razredi različnih podatkovnih nizov (Sheth in Kashyap, 1993). Objektni razredi različnih aplikacijskih ontologij so semantično enakovredni, ekvivalentni, če ustrezajo istim razredom v ontologiji domene. Objektni razredi različnih aplikacijskih ontologij so semantično sorodni, če ustrezajo razredom v ontologiji domene, ki so drug drugemu podrazred ali nadrazred. Objektna razreda A in B iz različnih aplikacijskih ontologij sta semantično ustrezna, če je razred A enakovreden z razredom v ontologiji domene, ki predstavlja en del sestavljenega razreda B v ontologiji aplikacije.

Semantična vsebina zbirk podatkov je povečana, če se pri modeliranju zbirke prostorskih podatkov uporabljajo koncepti višjih ravni za modeliranje kompleksnih prostorskih objektov, ki so podobni konceptom, s katerimi ljudje obravnavajo geografsko stvarnost. Druga razsežnost povečane semantičnosti zbirke podatkov je modeliranje razmerij med geografskimi pojavi v okviru zbirke podatkov, s čimer se zagotavlja njena pravilnost (Kemppainen, 1995).

2.1.3 Opredelitev ontologije prostora geografskih razsežnosti

Da bi postale geografske informacije bolj koristne in uporabne, je bilo treba razviti metodo za gradnjo ontologije z osredotočenjem na človeške aktivnosti v prostoru. Metoda temelji na človeških aktivnostih, povezuje aktivnosti z objekti domene. Oblikovana je hierarhija konceptov. Pri oblikovanju se uporablja zdravorazumsko mišljenje, uporabljajo se zapisana besedila, ki opisujejo človeške aktivnosti. Konceptualna povpraševanja v zbirkah prostorskih podatkov so po naravi hierarhična

(Frank, 1999). Rosch (1987) ločuje dve ravni kognitivne klasifikacije: osnovno in napredno. Vsaka družina zdravorazumskih kategorij je hierarhično organizirana v drevo, ki ima na svojem začetku zelo splošne kategorije, ki v nadaljevanju postajajo specializirane in organizirane po različnih načelih. Na področju človeških aktivnosti je vzročnost posledica želja in hotenj posameznikov (Dowty, 1998). V metodologiji je predpostavljena hierarhična razdelitev v naslednje ravni:

- raven 0: izhodišče ontologije,
- raven 1: inducirana kategorija,
- raven 2: superrazred,
- raven 3: razred,
- raven 4: objekt/shema (pojav),
- raven 5: atribut (objekta/sheme pojava).

Določene razlike človeških kognitivnih sistemov so posplošene tako, da je mišljenje splošno veljavno. Splošno, zdravorazumsko mišljenje torej obstaja neodvisno od različnosti človeških kognitivnih aktivnosti (Smith, 1992). Casati et al. (1998) je odkrival ontologijo ljudske, zdravorazumske geoprostorske stvarnosti z uporabo formalnih logičnih metod. Za GIS je bilo treba razviti metode transformacije kvantitativnih prostorskih podatkov v vrsto kvalitativnih predstavitev prostorskih pojavov geografskih razsežnosti, ki so prilagodljive razumevanju GIS izvedencev in laikov. Za osnovo tem metodam lahko služi le teorija ontologije zdravorazumskega pojmovanja geoprostorske stvarnosti (Smith in Mark, 2001). Metodo izdelave ontologije prostora geografskih razsežnosti je zaradi potrebe po hierarhičnosti modela, ki povečuje enoličnost strukture in hitrost iskanja konceptov, razdeljena v tri faze: določitev izhodišča ontologije, izdelava inducirane ravni ontologije ter stvarnih ravni ontologije prostora.

Z metodo, opisano v nadaljevanju, je izdelano konceptualno omrežje domene prostora geografskih razsežnosti v naravnem jeziku kakor tudi v obliki formalno opredeljenih izrazov na osnovi ontologije konceptov predstavitve znanja privzeto po Sowa (2000b). Najprej so predstavljeni temeljni koncepti predstavitev znanja, primerni za opredelitev znanja, simbolov prostora. Univerzalna ontologija predstavitve znanja o prostoru geografskih razsežnosti je zasnovana kot sistem človeških aktivnosti (namenov) in prostorskih pojav (pojavnih oblik – shem), ki so vzrok (pojav) ali posledica (objekt) aktivnosti.

2.1.3.1 Določitev izhodišča ontologije prostora

Za določitev izhodišča ontologije prostora je določena omejitev koncepta »prostor« s parametrom razsežnost. Z obravnavano ontologijo je za željeno obravnavati objekte geografskega prostora, zato je parameter razsežnosti omejen na geografsko razsežnosti, ki sta jo opredelila Egenhofer in Mark (1995): »Prostorski objekti geografskega obsega (razsežnosti) so objekti, ki so večji od človeškega telesa in jih ni mogoče v celoti zaznati v okviru enega zaznavnega dejanja.« Izhodišču ontologije je določena oznaka ravni 0.

2.1.3.2 Izdelava inducirane ravni ontologije prostora

Pristop induciranja konceptov je privzet po metodologiji konstruktivne indukcije, ki je metoda klasifikacije v induktivnem logičnem programiranju. Uporablja se pri obravnavi nekaterih vrst problemov, kjer se je pokazalo, da obstoječi atributi in relacije ne zadoščajo za razumljiv opis danega koncepta. Problemi takšne vrste se rešujejo s samodejnim ali ročnim dodajanjem novih, vmesnih konceptov (Robnik-Šikonja, 1997).

Z izdelavo inducirane ravni ontologije prostora so povezani deli splošne ontologije predstavitve znanja, ki ga je razvil Sowa (2000a). Tri izbrane kategorije so formalno zapisane v obliki trditve [1], [2] in [3]. Aktivnosti ljudi v prostoru geografskih razsežnosti so opredeljene s kategorijo »namen«. Namen je hotenje, ki pojasnjuje stanje, situacijo:

$$\exists \text{ Namen} = \text{pojmovno} \wedge \text{posredovalno} \wedge \text{pojavno} \quad [1]$$

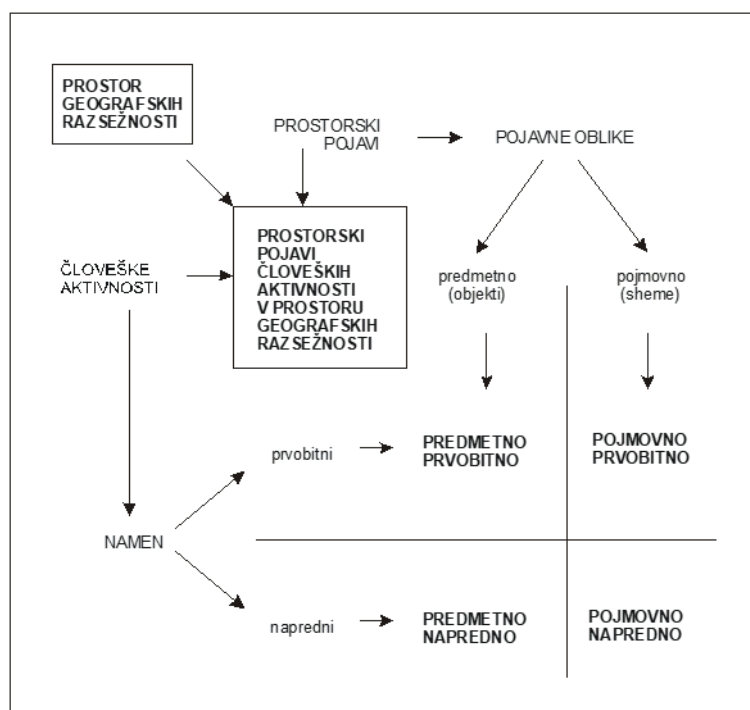
V ontologiji znanja (Sowa, 2000a) trditve povezujejo namen z aktivnostmi in povzročitelji, agenti. Trditve za namen in aktivnosti lahko vključujejo parametre, kot so časovno zaporedje, naključnost in uspeh ali neuspeh. Nameni človeških aktivnosti v prostoru geografskih razsežnosti se posredujejo v prostor geografskih razsežnosti preko dveh skupin dojetanja udeležbe agentov, in sicer kot prvobitne in napredne aktivnosti. Prvobitne človeške aktivnosti imajo svoj temeljni vzrok v namenu zadovoljevanja posameznikovih, individualnih hotenj in potreb. Napredne človeške aktivnosti imajo svoj vzrok v namenu zadovoljevanja kolektivnih hotenj skupnosti.

Pojavne oblike v prostoru geografskih razsežnosti je mogoče opredeliti s kategorijama objekt in shema. Posamezna kategorija (objekt, shema) je konjunkcija enostavnejših monadičnih predikatov (Sowa, 2000a):

$$\exists \text{ Objekt} = \text{predmetno} \wedge \text{samostojno} \wedge \text{nepretrgano} \quad [2]$$

$$\exists \text{ Shema} = \text{pojmovno} \wedge \text{samostojno} \wedge \text{nepretrgano} \quad [3]$$

Trditve [1], [2] in [3] so izpeljane iz splošne ontologije predstavitve znanja (Sowa, 2000).



Slika 6 Konceptualno omrežje prostora geografskih razsežnosti: izhodiščna in inducirana raven

Nadalje so opredeljene kombinacije kategorij prvobitne in napredne aktivnosti ter kategorij pojavnih oblik objekt in shema. Kombinacije so izražene v matriki kategorij, razsežnosti 2 krat 2, ki predstavlja inducirano raven izdelane ontologije

(Slika 6 in Preglednica 11). Kategorijam inducirane ravni (1) ontologije prostora so določene oznake vsebovanih kategorij. Formalno so zapisane kategorije inducirane ravni kot naslednje trditvene formule:

$\exists$ predmetno-prvobitno = predmetni objekti $\wedge$ prvobitnih človeških aktivnosti $\wedge$ geografske razsežnosti [4]

$\exists$ predmetno-napredno = predmetni objekti $\wedge$ naprednih človeških aktivnosti $\wedge$ geografske razsežnosti [5]

$\exists$ pojmovno-napredno = pojmovne sheme $\wedge$ naprednih človeških aktivnosti $\wedge$ geografske razsežnosti [6]

$\exists$ pojmovno-prvobitno = pojmovne sheme $\wedge$ prvobitnih človeških aktivnosti $\wedge$ geografske razsežnosti [7]

Tako so na formalni način ustvarjene kategorije inducirane ravni ontologije prostora geografskih razsežnosti.

2.1.3.3 Predpostavke modela

Konceptualno omrežje prostora temelji na človeških aktivnostih, namenih v geografskem prostoru. Kriterij za opredeljevanje aktivnosti je prostorska razsežnost – razpoložljivost prostora geografskih razsežnosti za izvajanje aktivnosti. Pojavna oblika človeške aktivnosti je prostorski objekt ali pojmovna shema prostorskih razsežnosti. Za aktivnosti modeliranja prostora na analogni ali digitalni medij z računalnikom ali drugimi sredstvi, na primer kartografskimi ali geodetskimi, prostor geografskih razsežnosti ni potreben. Za aktivnost modeliranja in napovedovanja je potreben miselni prostor v umu, na papirju ali v računalniškem pomnilniku in procesorju, ne pa prostor geografskih razsežnosti. Za aktivnosti opazovanja prostora je potreben prostor geografskih razsežnosti.

Kriterij za opredelitev novega temeljnega koncepta aktivnosti v ontologiji je razlika v namenu med človeškimi aktivnostmi, s katerih izvajanjem se ustvari prostorski predmet – objekt – oziroma pojmovna

shema prostora. Kriterij za opredelitev nove pod aktivnosti (razreda, podrazreda) je razlika v namenu izvedbe aktivnosti ali razlika v vrsti pojavne oblike.

Zapletene objekte in sheme človeških aktivnosti v geografskem prostoru je mogoče predstaviti s kombiniranjem enostavnih kategorij objektov/shem – navedenih v OPR. Nameni človeških aktivnosti in pojavne oblike se prekrivajo v prostoru in času. Konceptualno omrežje prostora geografskih razsežnosti je zasnovano kot večplastna sestava domen raznovrstnih drugotnih prostorskih konceptov, zgrajenih nad osnovo homogene prostorske nepretrganosti nižje ravni kot domene prostorskih objektov in shem (Brentano, 1988).

Simboli, izrazi, ki se standardno uporabljajo za raznovrstne prostorske objekte, so atributi prostorskih namenov oziroma aktivnosti te ontologije (na primer hiša, stanovanje, lokal, parcela). Terminologija pod-objekt, nad-objekt je uporabljena le v smislu semantične primerjave med objekti. Ustvarjena ontologija ima hierarhično zasnovo. Zanj je značilna neodvisnost od podatkovnih objektov kakor tudi od stvarnih prostorskih objektov, ker je obravnavana ontologija opredeljena z nameni (prvobitni, napredni) oziroma aktivnostmi v prostoru ter značilnostmi njihovih pojavnih oblik, ki so lahko predmetne ali pojmovne. Primeri v hierarhiji modela so simboli, s katerimi se označujejo koncepti stvarnosti. Univerzalni obseg modela obravnava celotno domeno prostora geografskih razsežnosti. Mogoče je dinamično razširjanje modela z opredeljevanjem novih simbolov. Izražanje zapletenih konceptov se izvaja s kombiniranjem osnovnih konceptov.

Pojavne oblike v prostoru geografskih razsežnosti je mogoče opredeliti s kategorijama objekt in shema. Pojavne oblike v prostoru geografskih razsežnosti so stanja v prostoru, ki jih akterji, agenti, dojemajo kot predmetne pojavne oblike – objekte – in pojmovne pojavne oblike – sheme. Obstajajo prostorski predmetni objekti kot pojavne oblike v prostoru geografskih razsežnosti, za katere se štejejo le objekti človeškega izvora – ustvarjeni objekti. Obstajajo tudi prostorske pojmovne sheme kot pojavne oblike v prostoru geografskih razsežnosti, za katere se štejejo dogovorno določene – fiat meje – in opazovane pojavne oblike naravnih – bona-fide – in socioloških pojavov.

V preglednicah so navedeni koncept prve, inducirane ravni (Preglednica 11) in druge ravni (Preglednica 12) Ontologije prostora geografskih razsežnosti (OPGR).

Šifra	Prva raven
1	<i>predmetni objekti prvobitnih človeških aktivnosti</i>
2	<i>predmetni objekti naprednih človeških aktivnosti</i>
3	<i>pojmovne sheme naprednih človeških aktivnosti</i>
4	<i>pojmovne sheme prvobitnih človeških aktivnosti</i>

Preglednica 11 Ontologija prostora geografskih razsežnosti - prva raven

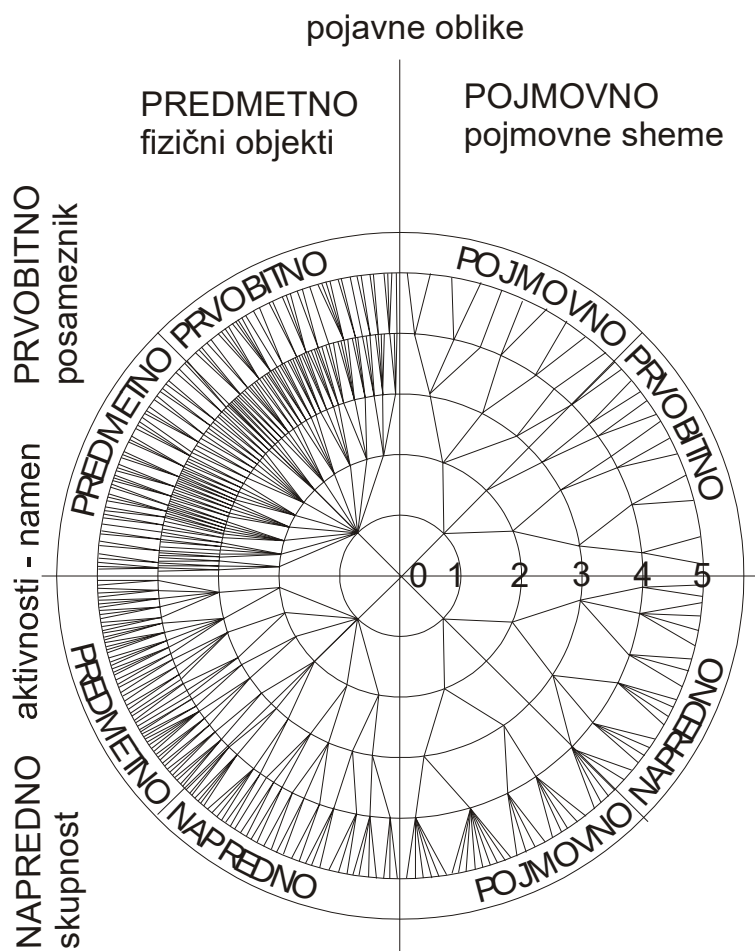
2.1.3.4 Izdelava stvarnih ravni ontologije prostora

V tretji fazi izdelave ontologije so obravnavane stvarne ravni ontologije prostora. Vnaprej določeni štirje inducirani koncepti so osnova za ustvarjanje stvarnih konceptov na hierarhični ravni z oznako 2. Na tej ravni se izvede razvejanje ontologije. Da bi se ohranila preglednost, je razvejanost na drugi ravni omejena na 12 konceptov za vsako od štirih skupin, izhajajočih iz induciranih konceptov, tako je največji teoretični obseg ravni 2 48 konceptov. V razviti metodologiji je treba upoštevati načelo »od manj podrobnega k bolj podrobnemu«, ki bi ga lahko primerjali s strategijo enkapsulacije objektov v razrede v domeni objektnega modeliranja.

Za oblikovanje novih konceptov je treba izvajati miselni postopek naravnega učenja, in sicer učenja z razumevanjem in vpogledom (Bratko et al., 1998), podrobneje s strategijo iskanja v širino ter raziskovanja znanja kot strategijo iskanja v globino (Kononenko, 1997). Pri strategiji učenja iskanje v

globino človek zaporedno predstavlja hipoteze in jih preskuša in spreminja, dokler je to možno, potem pa se vrne na izhodišče in postavi novo hipotezo. Tako učenje je za ljudi lahko, vendar ne vodi nujno k optimalni hipotezi in običajno zahteva veliko učnih primerov.

S strategijo iskanja v širino in globino so bili na drugi ravni ustvarjeni splošni, stvarni in razumljivi koncepti, ki so nadalje razvejani na nižjih ravneh ontologije (Slika 7). Osnovno vodilo za ustvarjanje novih konceptov so aktivnosti ljudi v prostoru geografskih razsežnosti, katerih posledica je stvaritev predmetnih objektov ali pojmovnih shem geografske razsežnosti.



Slika 7 Semantično referenčno omrežje, ontologija prostora geografskih razsežnosti (OPGR)

Koncepti druge ravni dedujejo lastnosti koncepta, ki predstavlja izbrano inducirano kategorijo (na ravni 1). Dodatno lastnost vsakemu konceptu druge ravni določa obravnavana aktivnost. Koncepti ravni 2 so formalno izraženi s formulo [8]:

$$\text{koncept (superrazred)} = \text{pojavnna oblika} \wedge \text{namen aktivnosti} \wedge \text{aktivnost} \wedge \text{razsežnost} \wedge \text{pomenska podrobnost} \quad [8]$$

Primer formalnega zapisa koncepta :

$$\text{koncept (POSTAVLJANJE mejnih predmetov)} = \text{predmetni objekti} \wedge \text{prvobitne aktivnosti} \wedge \text{postavljanje mejnih predmetov} \wedge \text{geografske razsežnosti} \wedge \text{superazred}$$

Tako izraženi koncept predstavlja odločevalno pravilo za uvrščanje konceptov na tretjo raven ontologije prostora. Sestavljanje govorečih šifer na ravni 2: enomestna šifra pripadajočega koncepta inducirane ravni (1) in k njej dodana dvomestna šifra koncepta druge ravni (10). Primer govoreče šifre za predhodno predstavljeni koncept (POSTAVLJANJE mejnih predmetov): 1_10 = 110.

Preglednica 12 - Ontologija prostora geografskih razsežnosti – prva in druga raven

Šifra	Prva raven	Druga raven
	predmetni objekti	
1	<i>prvobitnih</i> človeških aktivnosti	
101		objekti aktivnosti BIVANJA
102		objekti aktivnosti PRIDOBIVANJA surovin in energije
103		objekti aktivnosti PROIZVODNJE izdelkov (industrijsko/obratno)
104		objekti aktivnosti PRENOSA (TRANSPORTA)
105		objekti aktivnosti IZMENJAVE proizvedenih izdelkov in urbanih storitev
106		objekti aktivnosti ODLAGANJA ostankov
107		objekti aktivnosti SPROSTITVE na prostem
108		objekti aktivnosti MANIPULACIJE na prostem
109		objekti aktivnosti PARKIRANJA prevoznih sredstev
110		objekti aktivnosti POSTAVLJANJA mejnih predmetov
111		objekti aktivnosti OSVETLJEVANJA površin in spomenikov
112		objekti aktivnosti VERSKIH OBREDOV
2	<i>naprednih</i> človeških aktivnosti	
201		objekti aktivnosti JAVNEGA UPRAVLJANJA (državno/lokalno) socialne storitve, združenja
202		objekti aktivnosti IZOBRAŽEVANJA
203		objekti aktivnosti ZDRAVSTVENE OSKRBE
204		objekti aktivnosti OBRAMBE države in ZAŠČITE ljudi in sredstev
205		objekti aktivnosti NADZORA PREHAJANJA DRŽAVNE MEJE
206		objekti aktivnosti SHRANJEVANJA narodnih rezerv
207		objekti aktivnosti IZMER državnih mrež
208		objekti aktivnosti REGULACIJE naravnih pojavov
209		objekti aktivnosti ODDAJANJA informacij skupnosti
	pojmovne sheme	
3	<i>naprednih</i> človeških aktivnosti	
301		sheme aktivnosti UPRAVLJANJA PRAVIC v prostoru
302		sheme aktivnosti VAROVANJA naravnih virov in spomenikov
303		sheme aktivnosti PROSTORSKEGA PLANIRANJA
304		sheme aktivnosti POPISOVANJA PREBIVALSTVA
305		sheme aktivnosti OBDAVČEVANJA
306		sheme aktivnosti RAZDELJEVANJA DENARNIH PODPOR
307		sheme aktivnosti DOLOČANJA MEJA ADMINISTRATIVNIH STATUSOV V PROSTORU
4	<i>prvobitnih</i> človeških aktivnosti	
401		sheme aktivnosti OPAZOVANJA NARAVNIH PREDMETNIH POJAVOV geografskih razsežnosti

Šifra	Prva raven	Druga raven
402		sheme aktivnosti VZPOSTAVLJANJA lastnih imen (shem) in geometričnih shem
403		sheme aktivnosti OPAZOVANJA POJAVOV V ČLOVEŠKI DRUŽBI pojavov geografskih razsežnosti

Simboli konceptov četrte in pete ravni so semantični slovar izrazov prostora geografske razsežnosti. Za koncepte, za katere obstajajo istopomenski izrazi, se dodaja povezava na njihove sinonimne simbole. Na enak način se izvaja tudi označevanje konceptov z izrazi v tujih jezikih. To omogoča uporabo izdelane ontologije za povezovanje konceptov zbirk podatkov, izraženih v različnih nacionalnih jezikih. Osnovna ontologija prostora geografskih razsežnosti obsega 588 konceptov na petih ravneh in je razširljiva. Konceptualno ogrodje je osnova za organizacijo semantičnega slovarja. Formalni zapis OPGR je bil izveden le za najvišje ravni ontologije, in sicer v jeziku OIL.

Opomba: Trenutni operativni model UOPGR (glej poglavje 3) se deli na 4 kategorije, ki se skupno delijo na 34 razredov, ki se delijo skupno na 105 podrazredov in obsegajo 446 entitet (objektov, shem ali pojavov) na četrty ravni. Skupno 590 konceptov.

2.2 Rezultati naloge 2.2: Izdelava konceptualne zasnove za ocenjevanje stopnje podobnosti podatkovnih prostorov

Zagotavljanje semantične kakovosti podatkov je tesno vezano na *ohranjanje in sporočanje* pomenskih razmerij, sklicevanj in opredelitve *okoliščin* (konteksta), ki dajejo *podatkom pomen*. Ocenjevanje podatkov za doseganje semantične kakovosti vodijo naslednje usmeritve:

- Določiti je treba cilje in *merila ocenjevanja*.
- Izbrati je treba ustrezna *vnosna besedila* (opise) in sistem za *primerjavo* (mere, referenčni sistem).
- Uporabiti je treba več *dopolnjujočih metod* (hibridno) ocenjevanja.
- V ocenjevalni proces je treba vključiti *povratne informacije uporabnikov*.
- Podrobno je treba *dokumentirati predpostavke, omejitve* ocenjevanja in končne rezultate.

Ko poskušamo *uporabiti podatke iz drugega sistema*, mora biti na voljo dovolj bogatih informacij o kontekstu podatkov (na primer metapodatkovnih opisov, taksonomije, ontologije), sicer ni mogoče doseči semantične medopravilnosti.

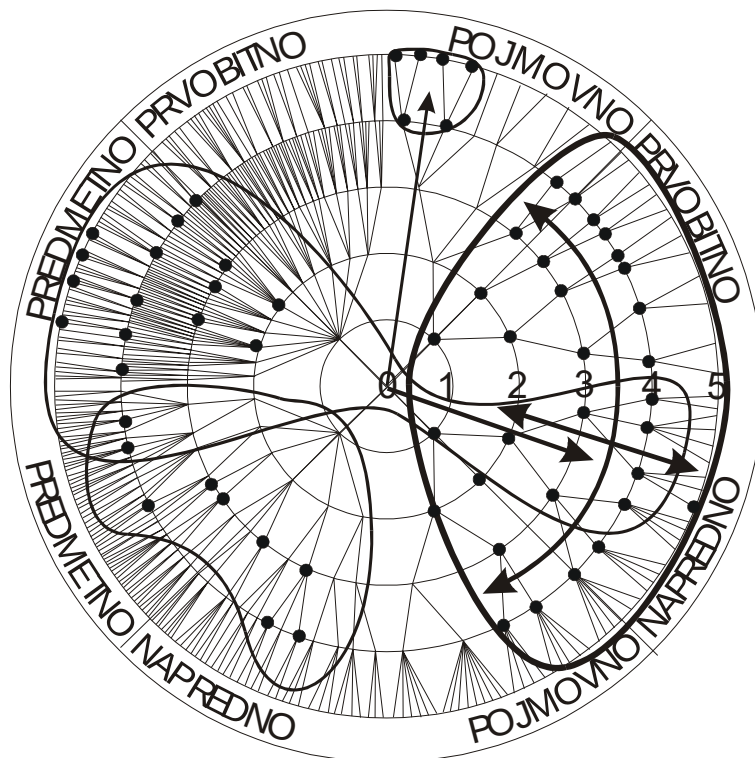
Ocenjevanje semantične kakovosti vključuje mere za ocenjevanje, *kako dobro* podatkovni modeli, opisi (metapodatki) ali *sistemi predstavljajo in prenašajo pomen*. Osnovne mere za ocenjevanje semantične kakovosti zbirk in delov zbirk podatkov so:

- Pravilnost: *točnost* posredovanih informacij.
- Doslednost: *kako enotno* se ista pravila ali standardi uporabljajo v različnih primerih uporabe.
- Popolnost: *obseg, delež pomembnih* informacij, ki so predstavljene.
- Podrobnost: *spособnost razlikovanja* med različnimi ravnmi ali kategorijami pomena podatkov.
- Veljavnost: *skladnost* z logičnimi, semantičnimi ali pragmatičnimi standardi.
- Uporabnost: Kako dobro sistem v pomenskem smislu izpolnjuje *potrebe uporabnikov*.

Glavne metode za merjenje semantične kakovosti so:

1. Metoda *humanega* (ročnega) ocenjevanja, ki vključuje strokovnjake, ki pregledujejo metapodatke in jih ocenjujejo na osnovi vnaprej določenih mer semantične kakovosti in potencialnih semantičnih referenčnih sistemov (taksonomija, ontologija, graf znanja). Ta metoda lahko zajame tudi zapletena semantična razmerja, vendar je zamudna in lahko povzroči pristranskost.
2. Metode *samodejnega* ocenjevanja; uporabljajo se mere kakovosti, za primerjavo ocenjenih semantičnih vrednosti z referenčno mero - standardom ali osnovnim sistemom. Samodejna metoda je učinkovita in objektivna, vendar lahko spregleda pomenske nianse in kontekst (okoliščine).
3. Metode *hibridnega* ocenjevanja; združuje samodejne in ročne metode, pri čemer izkorišča prednosti vsake za zagotavljanje celovitejše ocene. Ta pristop je lahko zapleten zaradi potrebe po koordinaciji med avtomatiziranimi sistemi in človeškimi ocenjevalci.

Za namene ocenjevanja semantike zbirk prostorskih podatkov je opredeljen nabor mer za ocenjevanje semantičnih značilnosti zbirk podatkov kot zaključenih celot in njihovih delov (Slika 8). Semantična skladnost zbirk prostorskih podatkov je utemeljena kot stopnja podobnosti zbirk prostorskih podatkov glede na dve merili: *semantično globino zbirke* podatkov in *semantično razpršenost* zbirke podatkov glede na ontologijo domene prostora. Večja, ko je stopnja podobnosti med semantičnimi parametri primerjanih zbirk, večja je semantična skladnost primerjanih zbirk podatkov in s tem semantična skladnost posameznih miselnih zasnov (Čeh, 2001).



Slika 8: Grafični prikaz semantičnih parametrov zbirk prostorskih podatkov

Podatek o razpršenosti zbirke podatkov je za poizvedovanje v namene integracije zbirk podatkov v okolju porazdeljenih geografskih informacijskih sistemov uporaben meta podatek. Semantična ustreznost ontologije je ocena stopnje primernosti določene ontologije za izražanje konceptov zbirk podatkov. To je relativna mera, s katero se izraža kakovost ontologije glede na stvarne zbirke podatkov. Zato se za izračun uporabijo rezultati izražanja konceptov zbirk podatkov s koncepti ontologije. Vsote ekvivalentnih in agregatnih razredov se povprečijo in izrazijo v odstotkih.

Semantična pristranost koncepta je posledica podrejanja metodam zajemanja podatkov pri opredeljevanju konceptov prostora. Še posebej je značilna za koncepte v katalogih prostorskih podatkov, ki so zajeti z metodami daljinskega zaznavanja (satelitskimi in letalskimi). V opredelitvah pristranskih konceptov je značilna uporaba izrazov za opisovanje snovi (pesek, kamen, beton, asfalt, vegetacija), za katere je enostavno interpretirati valovno dolžino zaznanega valovanja. Druge v tem delu opredeljene mere za ocenjevanje semantike zbirk prostorskih podatkov so še:

- semantična globina prostorskega koncepta,
- semantična globina zbirke prostorskih podatkov,
- semantična razdalja dveh konceptov,
- semantična razpršenost zbirke prostorskih podatkov,
- kakovost predstavitve semantike zbirke prostorskih podatkov.

V prihodnosti bodo eksplicitno izražena razmerja med shranjenimi podatki v zbirki podatkov (semantičnih podatkov) postala sestavni del vsakega informacijskega sistema. Zbirke prostorskih podatkov se bodo kakovostno razlikovale tudi po opisu vsebovanih semantičnih podatkov, ker bo semantiko zbirke prostorskih podatkov mogoče predstaviti z različnimi semantičnimi merami. Če bo zbirka prostorskih podatkov, med svojimi meta podatki, vključevala mere, kot so absolutna semantična globina zbirke prostorskih podatkov, relativna prostranost zbirke podatkov, relativno razsipanje zbirke podatkov, semantična razpršenost zbirke podatkov, bo kakovost semantične predstavitve ocenjena z drugo stopnjo. Če bo zbirka prostorskih podatkov poleg naštetih mer med meta podatki vključevala tudi prevod kataloga v semantično referenčno omrežje, ogrodje (OPGR), bo kakovost semantične predstavitve, ocenjena s prvo stopnjo.

V metodi posrednega povezovanja zbirk prostorskih podatkov je treba določiti semantična razmerja med semantičnim referenčnim sistemom, ki služi kot posredovalni sistem, ter katalogi obravnavanih zbirk podatkov. Izdelana je metoda ocenjevanja semantike zbirk prostorskih podatkov. Metoda temelji na semantičnem referenčnem omrežju – ontologiji prostora geografskih razsežnosti (OPGR). Metodologija ocenjevanja semantike zbirk prostorskih podatkov je sestavljena iz zaporednih faz, ki so podrobneje opisane v nadaljevanju:

1. prevedba konceptov kataloga zbirke podatkov v semantično referenčno omrežje,
2. določitev semantičnih razmerij med koncepti zbirke in koncepti omrežja,
3. analiza semantične ustreznosti OPGR za povezovanje zbirk,
4. določitev semantične globine zbirke prostorskih podatkov,
5. določitev semantične razpršenosti zbirke prostorskih podatkov,
6. ocena kakovosti predstavitve semantike zbirke prostorskih podatkov.

S postopkom prevedbe konceptov iz kataloga zbirke podatkov v koncepte semantičnega referenčnega sistema se istočasno izvajata procesa uvrščanja konceptov kakor tudi pomensko bogatenje obravnavane zbirke podatkov.

Po opisanem postopku v preglednici kataloga ustvarimo atribut »pomen v OPGR«. S tem pridobljenim atributom je vsak od izvirnih konceptov kataloga zbirke podatkov pomensko obogaten. Pravimo, da je koncept pomensko obogaten, ker je z oznako (etiketo) ustreznega koncepta/konceptov iz OPGR pridobil vse pomenske attribute, ki so v OPGR formalno opredeljeni v okviru semantičnega referenčnega sistema, ki smo ga metodološko izvedli v razdelku o metodologiji izdelave ontologije prostora geografskih razsežnosti.

Za določitev semantičnih razmerij med koncepti zbirke in koncepti semantičnega referenčnega sistema smo privzeli in nadgradili opredelitev semantičnih razmerij med razredi ontologije domene in razredi ontologije aplikacije (Sheth in Kashyap, 1993). Obstajata dve osnovni semantični razmerji:

opredelitev razmerja *enakovrednih* razredov ontologije domene in razredov ontologije aplikacije:
`se_nanaša_na_enakovredni_razred [RazredDomene, RazredAplikacije];`

opredelitev razmerja *povezave* (dveh ali več) razredov ontologije domene tako, da tvorijo en sestavljen (agregatni) razred v ontologiji aplikacije:

`se_nanaša_na_sestavljene_razred [RazredDomene, RazredAplikacije]`

Ker je potrebna tudi pomenska analiza konceptov, za katere v ontologiji domene (OPGR) ni mogoče odkriti ustreznega koncepta, smo navedenima razmerjema dodali dodatno razmerje [11]:

neobstoječe razmerje med razredi ontologije domene in razredi ontologije aplikacije

`se_nanaša_na_nezdružljiv_razred [RazredDomene, RazredAplikacije]`

Za operativno uporabo smo navedena razmerja izrazili kot formule:

ekvivalent = $\text{se_nanaša_na_enakovredni_razred} [\text{RazredDomene}, \text{RazredAplikacije}];$ [9]

agregat = $\text{se_nanaša_na_sestavljeni_razred} [\text{RazredDomene}, \text{RazredAplikacije}];$ [10]

nezdružljiv = $\text{se_nanaša_na_nezdružljivi_razred} [\text{RazredDomene}, \text{RazredAplikacije}];$ [11]

V postopku določitve semantičnih razmerij med koncepti zbirke podatkov in koncepti semantičnega referenčnega sistema uporabimo predhodno v OPRGR prevedeni in bogateni katalog obravnavane zbirke podatkov. V katalog uvedemo tri dodatne attribute (ekvivalent, agregat, nezdružljiv), s katerimi označujemo razmerja med koncepti zbirke podatkov in semantičnim referenčnim sistemom.

2.2.1.1 Analiza semantične ustreznosti OPRGR za povezovanje

Ker želimo poznati stopnje primernosti razvite ontologije za izražanje konceptov zbirk podatkov, ki jih obravnavamo, smo razvili metodo za izvedbo ocene semantične ustreznosti ontologije. Oceno semantične ustreznosti ontologije smo izvedli kot razmerje med številom uspešno izraženih konceptov zbirk podatkov in številom vseh konceptov zbirke podatkov (normalizacija):

$$\text{semantična ustreznost ontologije} = \frac{\text{število uspešno izraženih konceptov}}{\text{število vseh konceptov}} \quad [12]$$

Za izvedbo postopka določitve semantične ustreznosti ontologije uporabimo rezultate druge faze metode, in sicer prevedene kataloge obravnavanih zbirk podatkov z določenimi semantičnimi razmerji med koncepti zbirke podatkov in koncepti semantičnega referenčnega sistema (OPRGR).

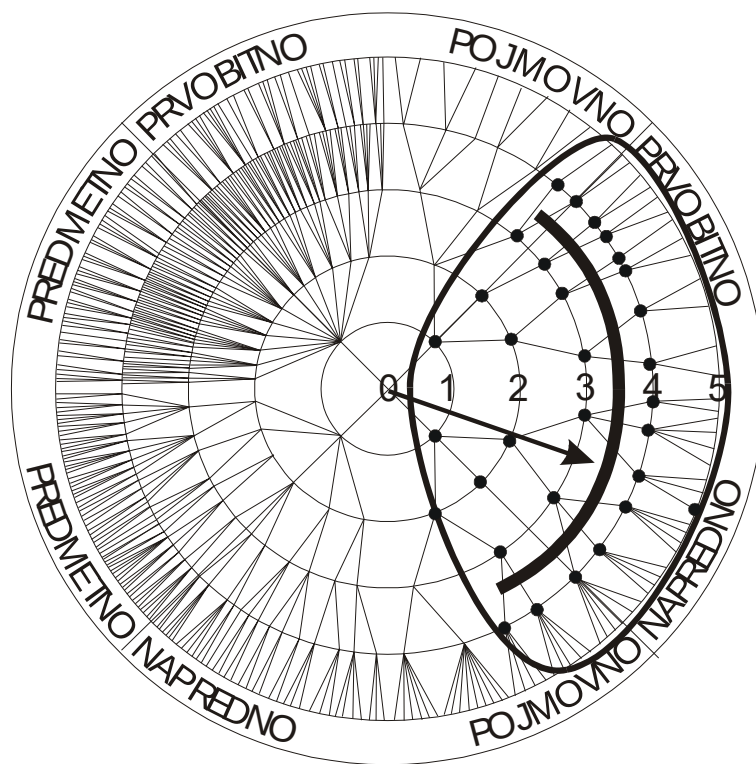
2.2.2 Semantična globina

Za vsako zbirko prostorskih podatkov, ki bi jo radi povezali z drugimi zbirkami prostorskih podatkov, želimo poznati stopnjo tematske, konceptualne podrobnosti oziroma splošnosti, specializacije/generalizacije. Znanje o stopnji tematske podrobnosti zbirk podatkov omogoča primerjavo zbirk podatkov na meta podatkovni ravni, zato smo razvili metodo opisano v nadaljevanju.

2.2.2.1 Določitev absolutne semantične globine

Absolutna semantična globina (Slika 9) zbirke prostorskih podatkov (*asg pz*) je mera, opredeljena kot povprečna globina vseh konceptov zbirke podatkov glede na izhodiščno raven oziroma osrednji koncept ontologije prostora geografskih razsežnosti. Utemeljena je na osnovi izračunov semantičnih globin konceptov (*sgk*) zbirke podatkov glede na izhodiščno raven 0, ontologije domene prostora geografskih razsežnosti; (Slika 10),

(n – število konceptov zbirke podatkov).



Slika 9: Grafični prikaz semantične globine zbirke prostorskih podatkov

$$asg_{pz} = (\sum sgk)/n \quad [13]$$

Za ročno odčitavanje – izračun semantičnih globin konceptov na osnovi govorečih šifer konceptov si pomagamo s pripomočkom (Preglednica 13)

Oznake	Izhodišče ontologije	Inducirana raven	Superrazred	Razred	Objekt/shema	Atribut
Oznaka ravni	0	1	2	3	4	5
Razsežnost šifre	–	1	2	2	2	2
Globina	0	1	2	3	4	5

Preglednica 13 Pripomoček za izračun semantičnih globin konceptov na osnovi govorečih šifer

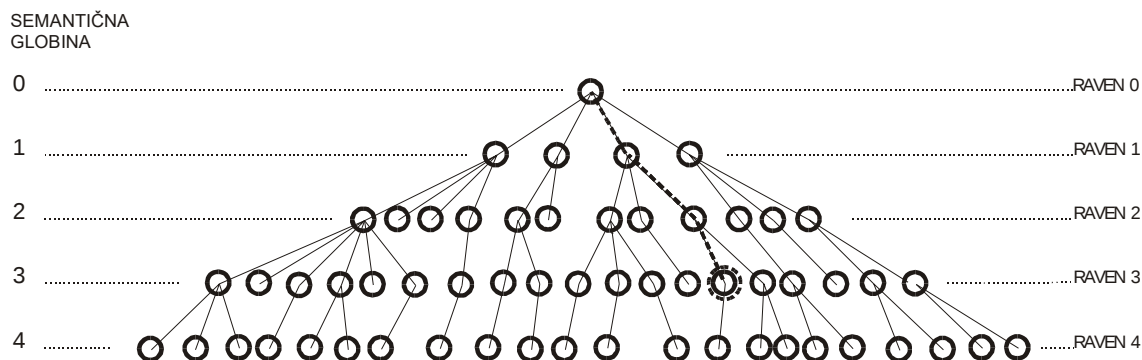
Primeri določanja semantične globine koncepta s pripomočkom (Preglednica 13), so navedeni v Preglednica 14:

Koncept	Šifra					Globina
hmeljišče	1	02	03	03		4
zemljišča/objekti za pridobivanje rastlinskih surovin	1	02	03			3
ruševje	4	01	07	01	09	5
Globina	1	2	3	4	5	

Preglednica 14: Določanje semantične globine koncepta na osnovi govorečih šifer

2.2.2.2 Absolutna semantična globina posameznega prostorskega koncepta

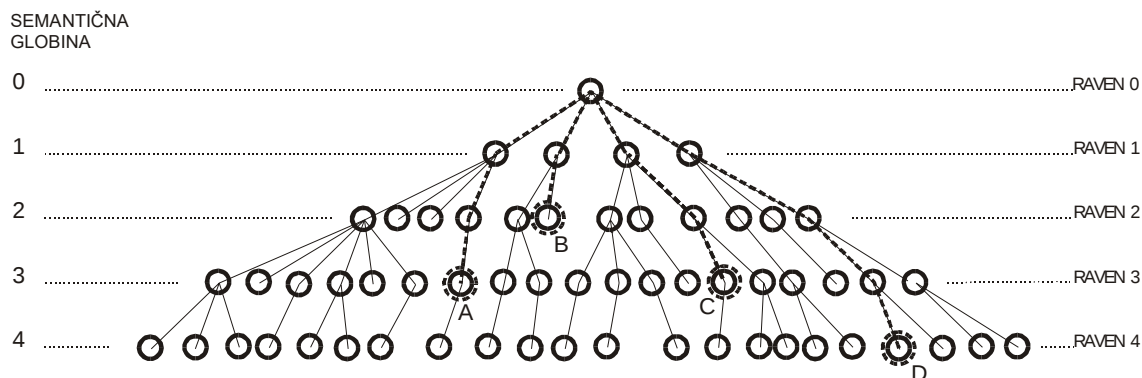
Semantična globina prostorskega koncepta (*sgk*) je mera, opredeljena glede na izhodiščno raven ontologije prostora geografskih razsežnosti (Slika 10). Utemeljena je na osnovi izračuna semantične razdalje med izhodiščno ravno ontologije domene prostora ter obravnavanim prostorskim konceptom. Za semantične razdalje med posameznimi ravni ontologije je normirana vrednost 1 (ena enota).



Slika 10: Prikaz semantične globine koncepta (*sgk*) v konceptualnem drevesu (grafu). Semantična globina izbranega koncepta je 3.

2.2.2.3 Absolutna semantična globina zbirke prostorskih podatkov

Absolutna semantična globina zbirke prostorskih podatkov (*asgzp*) je mera, opredeljena kot povprečna globina vseh konceptov zbirke podatkov glede na izhodiščno raven oziroma osrednji koncept ontologije prostora geografskih razsežnosti (Slika 11). Utemeljena je na osnovi izračuna povprečne semantične globine konceptov (*sgk*) zbirke podatkov glede na izhodiščno raven ontologije domene prostora.



Slika 11 Prikaz absolutne semantične globine (*asgzp*) zbirke prostorskih podatkov $P \sqcup A, B, C, D$ v konceptualnem drevesu (grafu). Semantična globina zbirke podatkov je $(3 + 2 + 3 + 4)/4 = 3$.

Koncept, izražen z OPGR		Semantična globina koncepta	Semantična globina zbirke
1020303	hmeljišče	4	
101	bivanje	2	
10405	tirni	3	
		9	3

Preglednica 15 Primer izračuna absolutne semantične globine (*asgzp*) hipotetične zbirke prostorskih podatkov

Primer izračuna absolutne semantične globine zbirke prostorskih podatkov (Preglednica 15): za hipotetično zbirko podatkov je izbran podniz treh konceptov iz kataloga zbirke podatkov rabe kmetijskih zemljišč P1 {hmeljišče, bivanje, tirni}. Za to zbirko podatkov treh elementov je izračunana semantična globina 3.

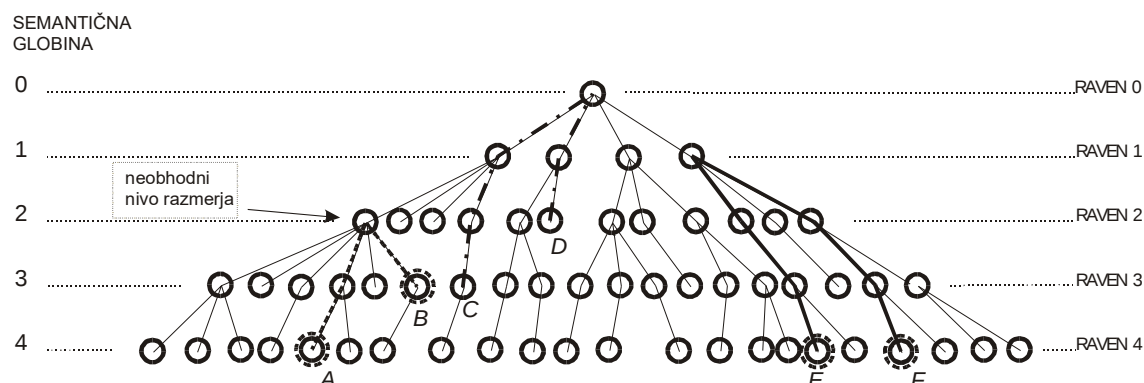
2.2.3 Semantične razdalje

Za vsako zbirko prostorskih podatkov, ki bi jo radi povezali z drugimi zbirkami prostorskih podatkov, želimo poznati stopnjo tematske, konceptualne *osredotočenosti*. Znanje o stopnji tematske osredotočenosti zbirk podatkov omogoča primerjavo zbirk podatkov na meta podatkovni ravni, zato smo razvili metodo, opisano v nadaljevanju. Semantična razpršenost zbirke prostorskih podatkov je mera splošnosti oziroma osredotočenosti, ki jo v razviti metodologiji izračunamo na osnovi semantičnih razdalj med vsemi koncepti obravnavane zbirke podatkov in uteži. Semantične razdalje se morajo izračunati za vse permutacije razmerij med koncepti.

Semantična razdalja razmerja

Semantična razdalja semantičnega razmerja med konceptoma je vsota odsekov poti v semantičnem referenčnem sistemu, ki so potrebni za vzpostavitev razmerja med konceptoma. V ta namen je treba poznati potrebno raven v shemi, ki jo je za vzpostavitev razmerja med konceptoma treba doseči v semantičnem referenčnem sistemu.

Neobhodno raven je smiselno izraziti s semantično globino. Semantične razdalje razmerij med koncepti ($d(A, B) = 3$, $d(C, D) = 5$, $d(E, F) = 6$); (Slika 12).



Slika 12 Semantične razdalje razmerij med koncepti ($d(A, B) = 3$, $d(C, D) = 5$, $d(E, F) = 6$)

Preglednica 16 Določitev globin neobhodnih ravni (*gnr*) semantičnih razmerij (za vsak par primerjanih konceptov hipotetične zbirke podatkov P1 {hmeljišče, bivanje, tirni})

Globina neobhodne ravni razmerja (<i>gnr</i>)		hmeljišče	bivanje	tirni
		1020303	101	10405
hmeljišče	1020303	x		
bivanje	101	1	x	
tirni	10405	1	1	x

Določitev globlin neobhodnih ravni semantičnih razmerij (za vsak par primerjanih konceptov hipotetične zbirke podatkov P1 {hmeljišče, bivanje, tirni}) je izveden v Preglednica 16.

Vsota semantičnih razdalj zbirke podatkov P1 {hmeljišče, bivanje, tirni} znaša 12 enot (Preglednica 17).

Semantična razdalja (r)		hmeljišče	bivanje	tirni
		1020303	101	10405
hmeljišče	1020303	0		
bivanje	101	1 + 3 = 4	0	
tirni	10405	2 + 3 = 5	2 + 1 = 3	0

Preglednica 17 Izračun semantičnih razdalj (r) semantičnih razmerij (za vsak par primerjanih konceptov hipotetične zbirke podatkov P1 {hmeljišče, bivanje, tirni})

2.2.3.1 Uteži semantičnih razdalj

Vsako izračunano semantično razdaljo med dvema konceptoma je treba razviti po metodi utežiti z dvema raznovrstnima utežema.

Utež neobhodne ravni (unr); globlja, ko je neobhodna raven za vzpostavitev povezave med dvema konceptoma, manjša je razsutost zbirke in manjša naj bo utež. Utež naj bo obratno sorazmerna z neobhodno globino, povečano za vrednost 1 (da se izognemo deljenju z 0 na ničelni ravni). Utež izraža *prostranost zbirke* v okviru ontologije domene (Slika 13):

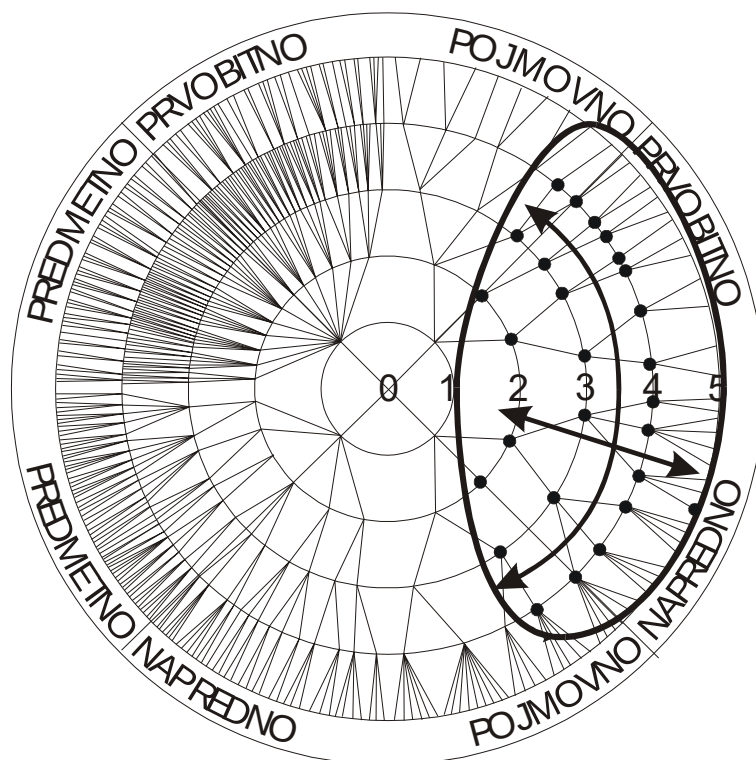
$$unr = 1/(gnr + 1) \quad [14]$$

Absolutna prostranost zbirke podatkov ($apzp$) je vsota semantičnih razdalj konceptov (r) uteženih z utežjo neobhodne ravni (unr):

$$apzp = \sum((r) \cdot (unr)) \quad [15]$$

Relativna prostranost zbirke podatkov ($rpzp$) je razmerje med absolutno prostranostjo zbirke podatkov [15] in absolutno prostranostjo semantičnega referenčnega sistema – ogrodja ($apso$):

$$rpzp = apzp/apso \quad [16]$$



Slika 13 Relativna prostranost in relativna globina zbirke podatkov

Utež relativne globine (urg); manjša kot je povprečna relativna globina (globinska razlika) od obeh konceptov do neobhodne ravni za vzpostavitev povezave med konceptoma, manjša naj bo utež. Utež odraža stopnjo *relativnega razsipanja* zbirke podatkov *po semantični globini*.

Za vsako izračunano semantično razdaljo med primerjanima konceptoma je treba po razviti metodi na osnovi razlik semantičnih globin konceptov (*sgk*) in neobhodnih ravni razmerij konceptov, ki se izražajo kot globina neobhodne ravni razmerja (*gnr*), izračunati *utež relativne globine (urg)*:

$$urg = sgk - gnr \quad [17]$$

Primer izračuna absolutne semantične prostranosti zbirke podatkov

Absolutna prostranost zbirke podatkov (*apzp*) [15] je vsota semantičnih razdalj zbirk podatkov, uteženih z utežmi neobhodne ravni (*unr*). Zato najprej izračunamo uteži neobhodne ravni (Preglednica 18).

Utež neobhodne ravni (<i>unr</i>) $1 / (gnr + 1)$		hmeljišče	bivanje	tirni
		1020303	101,00	10405
hmeljišče	1020303	x		
bivanje	101	0,50	x	
tirni	10405	0,50	0,50	x

Preglednica 18 Izračun uteži neobhodne ravni (*unr*) semantičnih razmerij (za vsak par primerjanih konceptov hipotetične zbirke podatkov P1 {hmeljišče, bivanje, tirni})

Utež neobhodne ravni (unr); ta utež odraža razsipanje zbirke podatkov po semantični širini (prostranosti). Za izračun absolutne semantične prostranosti zbirke podatkov ($apzp$) semantične razdalje razmerij (r) pomnožimo z utežmi neobhodne ravni (unr) in tako utežene razdalje seštejemo.

Utežene semantične razdalje (r) · (unr)		hmeljišče	bivanje	tirni
		1020303	101	10405
hmeljišče	1020303	x		
bivanje	101	$4 \cdot 0,5 = 2$	x	
tirni	10405	$5 \cdot 0,5 = 2,5$	$3 \times 0,5 = 1,5$	x

Preglednica 19 Izračun uteženih semantičnih razdalj (r)x(unr)

Izračun vsote uteženih semantičnih razdalj oziroma izračun absolutne prostranosti zbirke podatkov P1 {hmeljišče, bivanje, tirni}: $\Sigma(r) \cdot (unr) = 6$ (Preglednica 19).

2.2.4 Semantična razpršenost

Semantična razpršenost zbirke prostorskih podatkov je mera splošnosti oziroma osredotočenosti, generalizacije oziroma specializacije zbirke prostorskih podatkov. Izračun temelji na uteženih semantičnih razdaljah razmerij konceptov zbirke podatkov. V razviti metodi ločujemo dva pristopa k izračunu semantične razpršenosti zbirke podatkov.

Izračun semantične razpršenosti zbirke prostorskih podatkov – osnovni pristop

Pri osnovnem pristopu izračuna semantične razpršenosti zbirke podatkov izračunamo vsoto *relativne prostranosti zbirke podatkov* ($rpzp$), [16], in *relativne razsutosti zbirke podatkov po semantični globini* ($rrzpg$), [19].

Absolutna razsutost zbirke podatkov *po semantični globini* ($arzpg$) je vsota semantičnih razdalj (r), uteženih z utežmi relativne globine (urg), [17]:

$$arzpg = \sum r \cdot urg \quad [18]$$

Relativna razsutost zbirke podatkov *po semantični globini* ($rrzpg$) je razmerje med absolutno razsutostjo zbirke podatkov *po semantični globini* ($arzpg$), [20], in absolutno razsutostjo semantičnega referenčnega sistema – ogrodja ($arso$):

$$rrzpg = arzpg/arso \quad [19]$$

Semantična razpršenost zbirke podatkov ($srzp$) je vsota *relativne prostranosti zbirke podatkov* ($rpzp$), [16], in *relativne razsutosti zbirke podatkov po semantični globini* ($rrzpg$), [19]:

$$(srzp) = (rpzp) + (rrzpg) \quad [20]$$

2.2.4.1 *Izračun semantične razpršenosti zbirke prostorskih podatkov – poenostavljeni pristop*

Pri poenostavljenem pristopu izračuna semantične razpršenosti zbirke podatkov izračunamo vsoto (dvakratno) uteženih semantičnih razdalj. Semantične razdalje se morajo izračunati za vse permutacije razmerij med koncepti. Vsako izračunano semantično razdaljo med dvema konceptoma istočasno utežimo z dvema raznovrstnima utežema:

- utež neobhodne ravni; utež izraža prostranost zbirke v okviru ontologije domene,
- utež relativne globine; utež odraža stopnjo relativega razsipanja zbirke podatkov po semantični globini.

Semantična razpršenost zbirke podatkov se izračuna kot vsota uteženih semantičnih razdalj razmerij konceptov zbirke podatkov.

Primer izračuna uteži relativne globine

Za razmerja konceptov obravnavane zbirke podatkov na osnovi razlik semantičnih globin konceptov in neobhodnih ravni razmerij konceptov je najprej treba izračunati uteži relativne globine (*urg*), [17] (Preglednica 20).

Utež relativne globine (<i>urg</i>)		hmeljšče	bivanje	tirni	Semantična globina koncepta
		1020303	101	10405	
hmeljšče	1020303	x			4
bivanje	101	(2-1, 4-1) 0,5 = 2	x		2
tirni	10405	(3-1, 4-1) 0,5 = 2,5	(3-1, 2-1) 0,5 = 1,5	x	3
Semantična globina koncepta		4	2	3	

Preglednica 20 Izračun uteži relativne globine (*urg*) semantičnih razmerij (za vsak par primerjanih konceptov hipotetične zbirke podatkov P1 {hmeljšče, bivanje, tirni})

Utež relativne globine odraža relativno razsipanje zbirke podatkov po semantični globini.

Primer izračuna semantične razpršenosti zbirke podatkov s poenostavljenim pristopom

Uteži semantične razpršenosti $usr = (unr) \cdot (urg)$		hmeljšče	bivanje	tirni
		1020303	10100	10405
hmeljšče	1020303	x		
bivanje	101	4	x	
tirni	10405	6,25	2,25	x

Preglednica 21 Izračun uteži semantične razpršenosti (*usr*) zbirke podatkov kot produkt uteži neobhodne ravni (*unn*) in uteži relativne globine (*urg*)

Izračun semantične razpršenosti hipotetične zbirke podatkov P1 znaša 54 enot (Preglednica 21).

Utežene semantične razdalje		hmeljšče	bivanje	tirni
		1020303	10100	10405
1020303	hmeljšče	x		
101	bivanje	$4 \cdot 4 = 16$	x	
10405	tirni	$5 \cdot 6,25 = 31,25$	$3 \cdot 2,25 = 6,75$	x

Preglednica 22 Izračun uteženih semantičnih razdalj. Semantična razpršenost hipotetične zbirke prostorskih podatkov P1 {hmeljšče, bivanje, tirni}, $srzpp\ 1 = (16 + 31,25 + 6,75) = 54$

Dobljeno vrednost (Preglednica 22) je treba normalizirati, postaviti v razmerje do absolutne vrednosti (vsote uteženih razdalj vseh možnih razmerij med koncepti v ontologiji domene).

2.2.5 Analiza semantičnih razmerij aplikativnih konceptov v okviru semantičnega referenčnega omrežja

V tem razdelku je izvedena analiza semantičnih razmerij aplikativnih konceptov v okviru univerzalne ontologije prostora (OPGR).

2.2.5.1 Analiza semantične ustreznosti OPGR

Izvedena je bila analiza povezljivosti posameznih katalogov prostorskih podatkov s semantičnim referenčnim omrežjem, ontologijo prostora geografskih razsežnosti. Določitev razmerij med simboli katalogov, ontologij aplikacij in koncepti ontologije domene OPGR je izvedena v ustreznih preglednicah (Preglednica 25). Razmerja so ocenjena z dvema osnovnima semantičnima razmerjema, predikatoma in z negacijo razmerja:

razmerje *enakovrednih razredov* (ekvivalent) ontologije domene in razredi ontologije aplikacije;
razmerje *povezave razredov* (dveh ali več konceptov) ontologije domene, tako da v ontologiji aplikacije tvorijo *en sestavljeni razred* (agregat);

če v OPGR ni mogoče najti ustreznega koncepta oziroma ustreznega agregata konceptov, je ocenjeno, da razmerje ne obstaja (nezdružljiv).

Katalog konceptov	Semantična skladnost				Semantična ustreznost OPGR			
	ekvivalent	agregat	nezdružljiv	skupaj	e + a	e + a/s (%)	n (%)	S (%)
Raba zemljišč – kataster	31	5	6	42	36	86	14	100
Kmetijske rabe zemljišč	34	15	9	58	49	84	16	100
TNN VM 500	155	56	3	214	211	99	1	100
Topografske baze VN	10	1	5	16	11	69	31	100
CORINE Land Cover	27	27	5	59	54	92	8	100
Prostorske oblike	95	60	7	162	155	96	4	100
Skupaj	352	164	35	551	516	94	6	100
	64 %	30 %	6 %	100 %				

Preglednica 23 : Izračun semantične ustreznosti OPGR

Povprečno je mogoče štiriindevetdeset odstotkov (94 %) od skupno 551 konceptov iz obravnavanih katalogov prostorskih podatkov izraziti s katalogom ontologije prostora geografskih razsežnosti (Preglednica 23). Rezultat analize potrjuje visoko stopnjo semantične ustreznosti izdelane ontologije, še posebej, ker ontologija ni bila ustvarjena kot spoj obravnavanih zbirk podatkov – lokalnih ontologij, ampak kot neodvisna zasnova. Dokazano visoka stopnja semantične ustreznosti izdelane ontologije v razmerju do aplikativnih ontologij potrjuje visoko vrednost neodvisne zasnove OPGR in s tem njene univerzalnosti za domeno prostora geografskih razsežnosti. Zaradi visoke stopnje semantične ustreznosti je OPGR primerno orodje za ugotavljanje semantične skladnosti konceptov različnih zbirk prostorskih podatkov.

2.2.5.2 Analiza izrazov semantično nezdružljivih konceptov

Določenih konceptov posameznih katalogov zbirk prostorskih podatkov ni mogoče izraziti s simboli konceptov OPGR. Zato je bila izvedena analiza izrazov v nezdružljivih konceptih obravnavanih zbirk podatkov.

Katalog	Izraz					
	stavba	zgradba	objekt	zemljišče	površina	ostalo/drugo
Katalog rabe zemljišč zemljiškega katastra	2	1	1	2	1	
Katalog kmetijske rabe zemljišč (projekt Zajem)				7		5
Katalog konceptov temeljnih topografskih načrtov velikih meril (TNN VM 500)	1	1				1
Katalog konceptov topografske baze večje natančnosti		1	1	1		1
Katalog konceptov kataloga CORINE land cover					3	
Katalog konceptov prostorskih oblik (GURS)	4		1	2		
Skupaj	7	3	3	12	4	7

Preglednica 24 Analiza izrazov semantično nezdružljivih konceptov

Rezultati analize razkrivajo, da se v nevrstjenih konceptih najpogosteje pojavlja prostorski izraz zemljišče (12-krat). Sledijo mu prostorski izrazi stavba (7-krat), površina (4-krat), zgradba (3-krat) in objekt (3-krat). Torej sta najbolj dominantna izraza zemljišče in stavba. Izraza ostalo in drugo ne bosta predmet obravnave, ker se ne nanašata na prostor (Preglednica 24 in Preglednica 25).

Izraz zemljišče se tako pomensko izkazuje kot sinonim izraza prostor, ki v kreirani ontologiji OPGR predstavlja izhodišče. Kot izhodišče ontologije je izraz prostor izjemno splošen in zato neprimeren za praktično označevanje konceptov. Očitno je pri oblikovanju kataloga kmetijske rabe zemljišč prevladala uporaba besede zemljišče kot sinonima za izraz prostor. V tem katalogu se je izraz zemljišče pojavil kar 7-krat. Tudi za uporabo izraza površina bi bilo mogoče sklepati, da je uporabljen za izražanje splošnega koncepta prostora, četudi preko svojega enakozvočnega, homonimnega pomena za koncept ploskev. Tako je mogoče izpeljati enakovredno razmerje med koncepti prostor – ploskev – zemljišče oziroma združeno prostorska ploskev – zemljišče.

Analizo nadaljujemo z drugim najpogosteje zastopanim izrazom v nevrstljivih konceptih, to je z izrazom stavba (7), ki se mu sopomensko, kot sinonima, pridružujeta s skoraj enakim skupnim deležem v nevrstljivih konceptih izraz zgradba (3) in izraz objekt (3). Tudi za to skupino izrazov je mogoče izpeljati enakovredno razmerje med koncepti stavba – zgradba – objekt, ki jih ti izrazi predstavljajo. Izraz stavba ima v tem razmerju glede na numerične rezultate analize rahlo dominantno vlogo.

Če analizirane izraze in koncepte strnemo z vidika kognitivnega zaznavanja je mogoče pritrčiti, da je koncept prostora na najbolj splošni ravni sestavljen iz dveh elementov, ki sta:

- ploskev – zemljišče; element, ki ga zaznavamo kot odprt prostor, in
- stavba – zgradba – objekt; element, ki ga v človeški percepciji zaznavamo kot tridimenzionalne kvadre, kubuse bolj ali manj pravilnih oblik in zato samodejno sklepamo na njihov umetni izvor.

Oba elementa sta predstavitev koncepta prostora, ki ga je mogoče predstaviti kot unijo:

$$\text{prostor} = \text{ploskev} - \text{zemljišče} \cup \text{stavba} - \text{zgradba} - \text{objekt} \quad [21]$$

Pri opredeljevanju konceptov prostora za praktično uporabo (na primer v zbirkah podatkov) je treba presegati splošnost navedenih izrazov. Če se njihovi uporabi ni mogoče izogniti, jim je treba bogatiti pomen z dodajanjem simbolov/izrazov, ki imajo manj splošno semantiko oziroma večjo semantično globino. Takšni izrazi so povezani s koncepti človeških aktivnosti v prostoru.

Določenih konceptov posameznih katalogov zbirk prostorskih podatkov ni mogoče izraziti s koncepti OPGR. Zato je bila izvedena analiza izrazov v nezdružljivih konceptih obravnavanih zbirk podatkov.

Šifra	Koncept	Izraz					
		stavba	zgradba	objekt	zemljišče	površina	ostalo/drugo
Katalog rabe zemljišč zemljiškega katastra							
207	porušeni objekt			1			
217	stavbišče	1					
218	stavba ali zgradba	1	1				
299	nerazčiščeno grad. zemljišče				1		
399	nerazčiščene zelene površine					1	
499	nerazčiščeno nerodovitno zemljišče				1		
	skupaj	2	1	1	2	1	
Katalog kmetijske rabe zemljišč (zajem)							
14	vsa ostala kmetijska zemljišča				1		1
141	zemljišča v zaraščanju				1		
33	poslovna zemljišča				1		
35	zemljišča z mešano rabo				1		
366	druga zemljišča pod transportnimi in komunikacijskimi napravami				1		
38	sorodna odprta zemljišča				1		
422	ostala zamočvirjena zemljišča				1		1
63	ostalo (nikjer drugje op.)						1
714	ostale celinske vode						1
73	ostale vode (nikjer drugje op.)						1
	skupaj				7		5
Katalog konceptov temeljnih topografskih načrtov velikih meril TNN VM 500							
332300	drugo značilno znamenje						1
334400	manjša zgradba nedef. Namena		1				
334400	stavba nedifiniranega pomena	1					
	skupaj	1	1				1
Katalog konceptov topografske baze večje natančnosti							
102	opis zgradbe		1				
105	visoki objekt			1			
204	druga prometna površina						1
301	vegetacija						
302	zemljišče v posebni rabi RTP				1		
	skupaj		1	1	1		1
Katalog konceptov CORINE Land Cover							
11	urbane površine					1	
111	sklenjene urbane površine					1	
112	nesklenjene urbane površine					1	
133	gradbišča						

Šifra	Koncept	Izraz					
		stavba	zgradba	objekt	zemljišče	površina	ostalo/drugo
51	celinske vode						
	skupaj					3	
<i>Katalog konceptov prostorskih oblik (GURS)</i>							
3060002	komunalni objekt			1			
7030002	vrsta rabe zemljišča				1		
7040001	degradirano zemljišče				1		
8020006	točkovni elementi stavb	1					
8060001	manjša stavba	1					
8060002	stavba	1					
8060003	večja stavba	1					
	skupaj	4		1	2		
	skupaj	7	3	3	12	4	7

Preglednica 25 Analiza izrazov semantično nezdružljivih konceptov

2.3 Rezultati naloge 2.3: Izdelava idejne zasnove platforme za podporo procesu pomenskega usklajevanja

Za dolgoročno doseganje pomenske medopravilnosti in s tem povezljivosti različnih virov geoprostorskih podatkov, je ključno usmerjeno in vodeno sodelovanje med vsemi deležniki v našem podatkovnem prostoru. Za operativno podporo takšnega procesa v geodetski in sorodnih strokah v Sloveniji, predlagamo zasnovo digitalne spletne platforme, katere namen je podpora usklajevanju, izobraževanju ter upravljanju z elementi pomenske medopravilnosti, kot so slovarji, ontologije in usklajeni podatki ter metapodatki.

Usklajevanje je zasnovano po tematskih skupinah deležnikov geografskega prostora. Predvidena je izmenjava opredelitev (definicij) pomenov (semantike) besed (terminov) in simbolov drugih vrst, ki so vključeni v podatkovne sheme in same zbirke podatkov geodetske in strok, ki se veže na geoprostorske podatke.

2.3.1 Postopek usklajevanja pomenske medopravilnosti

Namen zasnovane platforme je podpora procesu usklajevanja pomenov aplikativnih podatkov za doseganje tako imenovane semantične medopravilnosti. Za lažje razumevanje in opis funkcionalnosti smo v tem poglavju opredelili korake postopka pomenskega usklajevanja (Slika 14). Na podlagi korakov smo opredelili potrebne funkcionalnosti platforme in podporne sisteme.



Slika 14 - Diagram poteka postopka pomenskega usklajevanja prostorskih podatkov

Usklajevanje se začne z ***nalaganjem pripravljenih aplikativnih meta podatkov v platformo***. V tej fazi se bodo analizirale podatkovne sheme kar bo omogočalo poznejšo primerjavo naloženih meta podatkov z že naloženimi shemami in jedrnimi ontologijami. Po končani izvedbi nalaganja meta podatkov v skupno platformo se lahko izvede postopek primerjanja pomenov metapodatkov oziroma ugotavljanja razlik v pomenih posameznih oznak, simbolov. Ker so trenutno podatki večinoma v različnih sistemih, z različnimi podatkovnimi modeli in pogosto v različnih formatih pravimo, da so ti podatki heterogeni.

Nato sledi faza ***harmonizacije*** oziroma pomenskega usklajevanja meta podatkov. Platforma vsebuje t.i. modul za harmonizacijo, katerega naloga je, da primerja naložene meta podatke iz heterogenih virov s standardiziranimi opredelitvami v besednjakih predhodno dosegljivih v semantičnem repozitoriju. Z utemeljevanjem pomenske podobnosti, (reasoning) ali raznolikosti, se identificirajo odstopanja in predlaga smiselne dopolnitve oziroma pomenske bogatitve aplikativnih meta podatkov. Za doseg harmonizacije se podatki lahko še dodatno pomensko nadgradijo in obogatijo.

Zadnji korak v modulu za harmonizacijo je preverjanje kakovosti oziroma ***validacija pomenske usklajenosti metapodatkov glede na standarde v domeni semantične medopravilnosti***. V postopku validacije, se preveri, kako dobro je bil izveden postopek harmonizacije glede izbrane pomenske in tehnološke standarde in se izpostavijo možne napake ali anomalije, ki se bodo obravnavale v postopku usklajevanja.

Z zaključkom validacije postopka harmonizacije, so podatki pripravljeni na medorganizacijsko ***usklajevanje***. Za podporo tega postopka je predviden portal za usklajevanje in sodelovanje, ki je bolj podrobno opisan v nadaljevanju. Vmesnik omogoča, da se povabijo vsi deležniki, k sodelovanju v

razpravi in predlaganju uskladitvenih sprememb. Ob zaključku vsakega cikla procesa lahko deležniki o uskladitvah glasujejo ali pa sprejmejo predlagano ontološko zavezo za spoštovanje dogovorjenih uskladitev pomenov metapodatkov v nadaljnjem poslovanju. V postopku usklajevanja, se dopolnjujejo in spreminjajo opredelitve besed in simbolov za posamezne miselne zasnove oziroma koncepte, kar vodi v iterativen proces izboljšave pomenskih definicij in ontologij in s tem do usklajene pomenske medopravilnosti.

Podporni proces usklajevanja je **obveščanje**. Preko modula za obveščanje na portalu za usklajevanje in sodelovanje so vsi deležniki obveščeni o pomembnih novostih, spremembah in pomembnih razpravah.

Ko se relativno širok krog deležnikov strinja z obravnavanimi pomenskimi vrzelmi in novimi, obogatenimi opredelitvami prej neskladnih pomenov besed oziroma simbolov, se lahko celoten postopek pomenskega usklajevanja zaključi s pomensko **integracijo**. V tem postopku se harmonizirani in validirani prostorski meta podatki, javno objavijo preko API-jev na tej platformi ali na izbranem sistemu za distribucijo.

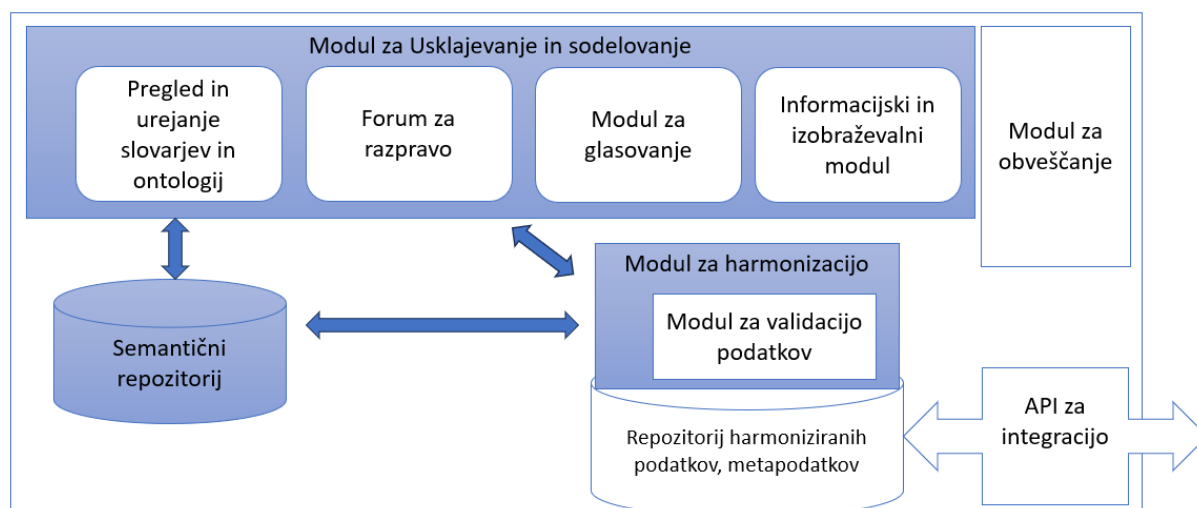
2.3.2 Elementi platforme za usklajevanje pomenske medopravilnosti

Za izvedbo opisanega procesa, mora biti izpolnjeno več pogojev, ki jih mora zagotavljati platforma v obliki orodij, podsistemov in organizacijske podpore. V idejni zasnovi platforme (Slika 15) smo opredelili naslednje elemente:

Glavni elementi z njihovimi podsistemi:

- Semantični repozitorij
- Modul za harmonizacijo
- Portal za usklajevanje in sodelovanje

Ti podporni elementi, so glede na končno izvedbo lahko implementirani centralizirano ali razpršeno.



Slika 15 - Shema idejne zasnove platforme za podporo procesu pomenskega usklajevanja

2.3.2.1 Semantični repozitorij

Semantični repozitorij predstavlja centralni del te platforme, saj vsebuje ključne elemente za semantično medopravilnost, ter njihovo urejanje, hrambo ter integracijo z ostalimi podsistemi. Predstavlja centralizirano vozlišče za standardne opredelitve pomenov besed, slovarje, ontologije, sheme in podobno. To vozlišče je osnova za harmonizacijo ter omogoča pomensko usklajeno razumevanje in interpretacijo podatkov iz različnih virov.

Zagotavljati bi moral vsaj naslednje funkcionalnosti:

- Orodje za pregledovanje, urejanje ter hrambo standardnih opredelitev in ontologij osnovnih geoporostorskih metapodatkov. To zagotavljajo orodja za upravljanje z ontologijami.
- Sistem za sledenje verzijam, zgodovini in spremembam (kot recimo GIT ali Github).
- Sistem za iskanje in poizvedovanje po RDF parih.

2.3.2.2 Modul za usklajevanje, harmonizacijo podatkov

Pomenska raznolikost (heterogenost) meta podatkov je največja ovira pri razvoju homogenih geoinformacijskih sistemov ¹, na ravni sodelovanja med javnimi sektorji v okviru posamezne države in usklajenega delovanja med več državami (na primer članicami EU), med več strokami ali vsebinskimi domenami. Proces harmonizacije se lahko izvaja na različnih ravneh, bodisi na ravni podatkov samih (vrsta geometrije, podatkovni model, zaloge vrednosti atributov) ali na ravni metapodatkov.

Namen modula za harmonizacijo je, da ob prevzemu podatkov, z avtomatiziranimi postopki usklajuje heterogene vire s standardiziranimi opredelitvami iz pomensko usklajenega semantičnega repozitorija. Na tem projektu smo se osredotočili na potrebne funkcionalnosti za harmonizacijo metapodatkov.

Osnovne funkcionalnosti tega modula so:

- pred obdelavo podatkov (to obsega pripravo podatkov na harmonizacijo, pregled skladnosti s standardi, morebitno čiščenje podatkov, pomensko bogatenje, normalizacijo in transformacijo)
- povezovanje (uparjanje) podatkov z različnih vsebinskih področij s standardnimi definicijami z uporabo algoritmov in predlogi najboljših pomenskih ujemanj z obstoječimi opredelitvami besed/simbolov. Hkrati izpostavlja pomenska neskladja, ali neskladne definicije za humani (ročni) pregled,
- bogatenje pomena simbolov, dopolnjevanje podatkov z različnimi informacijami iz semantičnega repozitorija, kar obogati podatke in poveča njihovo uporabnost za medopravilno povezljivost,
- transformiranje meta podatkov v standardizirano obliko ali format semantičnega spleta
- potrjevanje oziroma validacija rezultatov uskladitve, glede na standarde.

Za izgradnjo tega dela platforme bi lahko uporabili naslednja orodja:

- ETL Orodja kot so FME, Hale, org2ogr, Talend ali Apache Nifi,
- napredne algoritme za semantično usklajevanje (Semantic Matching),
- orodja za transformacijo kot XSLT
- namenske validatorje za izbrane formate.

Modul za pomensko harmonizacijo predstavlja most med heterogenimi viri podatkov in našim ciljem, homogenimi standardiziranimi odprtimi pomensko usklajenimi metapodatki.

2.3.2.3 Portal za humano usklajevanje in sodelovanje

Uspeh Semantične koordinacijske platforme za geoprostorske podatke temelji ne le na tehnični spretnosti, ampak tudi na sodelovanju in vključevanju deležnikov.

¹ BOOSTEE-CE D.T1.1.2 – Methodology to harmonize and integrate heterogeneous data, 2018

Glavni namen portala je, da vsem deležnikom predstavlja enoto vstopno točko za sodelovanje v procesu semantičnega usklajevanja geoprostorskih metapodatkov, preko katere lahko pregledujejo trenutno stanje definicij znotraj svojih vsebinskih sklopov, ter podajajo predloge in pregledujejo stanje vsebine. Poleg tega portal nudi tudi vse potrebne izobraževalne in aktualne vsebine.

Usklajevanje je zasnovano po tematskih skupinah deležnikov geografskega prostora in izmenjava opredelitev pomenov terminov, ki so vključeni v zbirke podatkov geodetske in drugih strok, ki ustvarjajo, ali uporabljajo geoprostorske podatke.

Portal sestavljajo naslednji moduli:

Uporabniški profili

Kot vsak standardni portal, ta modul omogoča deležnikom ustvarjanje in upravljanje lastnih profilov, nastavljanje lastnih nastavitvev, sledenje njihovim prispevkom in prilagajanje interakciji s platformo.

Vmesnik za oddajo povratnih informacij

Uporabniku prijazen vmesnik, kjer lahko deležniki podajo povratne informacije, predlagajo nove izraze ali predlagajo spremembe obstoječih opredelitev besed, izrazov.

Forumi za razprave

Forumi, specifični za posamezno temo semantične medopravilnosti, kjer uporabniki lahko začnejo ali se pridružijo razpravam, delijo vpoglede in sodelujejo pri pomensko določenih geoprostorskih temah.

Modul za glasovanje

Modul za glasovanje je integriran sistem, kjer so predlagane spremembe ali dodatki postavljeni na glasovanje, omogoča deležnikom, da imajo besedo pri oblikovanju enotnega semantičnega sistema.

Izobraževalno središče

Izobraževalno središče vsebuje pisna in video navodila za uporabo portala, tečaje za spoznavanje osnovnih pojmov in podobno.

Center za obvestila

Središče za obvestila obvešča uporabnike o posodobitvah, tekočih razpravah ali rokih za glasovanje, primernih glede na njihove vloge in interese, ter o novostih na področju obravnave.

Za izgradnjo tega dela platforme bi lahko uporabili naslednja orodja:

- Drupal ali WordPress, omogočajo ustvarjanje, upravljanje in spreminjanje digitalne vsebine portala.
- Platforme, kot so Discourse ali phpBB, zagotavljajo osnovo za diskusijske forume, zagotavljajoč nemotene in organizirane pogovore.
- Integracije z orodji, kot so Slack ali Microsoft Teams, lahko spodbujajo razprave v realnem času in sodelovanje pri odločanju. MS Teams je že prisotno orodje v državni upravi.
- Orodja, kot so UserVoice ali Zendesk, omogočajo učinkovito sledenje, upravljanje in reševanje povratnih informacij deležnikov
- VocBench.

Portal za sodelovanje ni le uporabniški vmesnik, ampak jedro »Semantične koordinacijske platforme«. S spodbujanjem aktivnega sodelovanja deležnikov, spodbujanjem bogatih razprav in dajanjem prednosti povratnim informacijam portal zagotavlja, da platforma ostane odzivna, dinamična in v skladu z evolucijo potreb geoprostorske skupnosti.

3 Delovni sveženj (DS3): Testno pregledovanje izbranih podatkovnih shem in šifrantov GURS

V tem delu raziskave smo se osredotočili na analizo ustreznosti podatkovnih shem in šifrantov za pomensko povezovanje (semantično povezovanje). Testno smo pregledali omejen nabor podatkov iz različnih podatkovnih zbirk, izbor pa bo temeljil na možnosti tematske povezljivosti slojev in šifrantov, ki jih bo možno na koncu medsebojno primerjati.

Semantična povezljivost geoprostorskih slojev temelji na obstoju metapodatkov. Glede na različne definicije metapodatkov in različne ravni opisa evidenc, glede na globino obstoj smo analizirali semantično opredeljenost testnih zbirk podatkov na treh različnih ravneh:

- **Raven Evidence** (metapodatki pogosto opisujejo le raven evidence, ki jo sestavlja več različnih slojev z različnimi pomeni, zgodovino izdelav itd. Primer je evidenca Topografskih podatkov, ki jo sestavljajo ceste, stavbe, vegetacija, itd. Obstoj metapodatkov na tej ravni je pomemben, a predstavlja nižjo semantično povezljivost)
- **Raven Sloja** (ti metapodatki, opisujejo raven sloja, enotne tematske evidence, ki opisuje enak tip objektov. Najpogosteje so bili vsi elementi sloja zajeti na enak način)
- **Raven atributa** (to, čemur se v GIS imenuje »atribut sloja«, se na področju podatkovne znanosti imenuje, v angleškem jeziku "feature metadata" ali »metapodatek virtualnega/informacijskega objekta«. Metapodatkom na tej ravni v GIS običajno imenujemo *podatkovni slovar* ali opis vrednosti in pomena atributov.

Za vire podatkov smo izbrali določene podatkovne zbirke Geodetske uprave RS (topografija, katastri, vrednotenje) in vire zunanjih deležnikov (na primer kmetijsko/gozdarska domena).

Večina podatkov poslovnih informacijskih sistemov je shranjena v relacijskih podatkovnih bazah. Za uspešno uporabo metod semantičnega spleta je treba povezati obe domeni. Pri relacijskih podatkovnih bazah smo, zaradi specifičnosti podatkov, ki se v njih hranijo, precej omejeni pri poskusih izboljšanja njihove kvalitete. Podatke semantično težko izboljšamo samo s čiščenjem in povezovanjem obstoječih podatkov, temveč predvsem z njihovim *bogatenjem* (Brooks et al.). Podatke obogatimo tako, da jim izboljšamo *semantično izraznost*. Šele obogatene podatke lahko povezujemo z viri iz oblaka strukturiranih podatkov OLD (prosto dostopni viri, združeni v oblak množic povezanih podatkov). bogatenje podatkov izvajamo s tehnikami *usklajevanja ontologij*. Izvedemo preslikavo in povezovanje obravnavane ontologije in konkretnih primerkov vhodne podatkovne baze z drugimi ontologijami. Po končanem procesu bogatitve izbrane relacijske podatkovne baze, začetna shema pridobi dodatne trditve v obliki trojic, in povezave z obstoječimi shemami oziroma ontologijami (Bališa, Lavbič, 2015).

Z analitičnim delom je bil dosežen cilj prepoznavanja ovir za semantično povezovanje geoprostorskih zbirk podatkov interno in na zunaj za GURS, in sicer na dveh ravneh (občinski in državni). S sintezo ugotovitev za posamezno tematiko je bil postavljen temelj za začetek razvoja procesov integracije na operativni ravni v okviru prihodnjih razvojnih projektov za bolj učinkovito upravljanje in bolj kakovostne javne storitve.

3.1 Metodologija

Predstavljena metodologija omogoča sistematično oceno možnosti povezovanja podatkovnih virov, identifikacijo ovir pri semantičnem povezovanju, oceno kakovosti dokumentacije in metapodatkov ter določitev stopnje semantične povezljivosti med različnimi viri.

Rezultati analize po tej metodologiji nam služijo kot osnova za izboljšanje semantične medopravilnosti podatkovnih zbirk, optimizacijo procesov povezovanja podatkov ter identifikacijo področij, kjer so potrebne izboljšave v dokumentaciji ali strukturi podatkov

Ta metodološki okvir predstavlja strukturiran pristop k analizi semantične povezljivosti in omogoča konsistentno ocenjevanje različnih podatkovnih virov ter njihove primernosti za semantično povezovanje.

3.1.1 Postopek analize

Analiza semantične povezljivosti poteka v naslednjih korakih:

- **Opredelitev podatkov za analizo**
 - Identifikacija relevantnih podatkovnih slojev
 - Zbiranje pripadajoče dokumentacije
 - Pregled prosto dostopnih virov in pridobivanje dodatne dokumentacije od upravljavcev
- **Ocena pogojev za semantično povezovanje**
 - Pregled sintaktičnih lastnosti (Analiza tehničnih in vsebinskih omejitev)
 - Format zapisa geometrije in dokumentacije
 - Ocena razčlenjenosti zbirke
 - Obstoje enoličnih identifikatorjev
 - Ocena strojne berljivosti po 5-zvezdični lestvici (Berner's Lee)
 - Ocena obstoja in obsega elementov kakovosti
 - Izvor podatkov (lineage)
 - Metodologija zajema
 - Časovna umeščenost podatkov
 - Konsistentnost objave in atributov
 - Pregled obstoja semantičnih elementov
- **Ocena dostopnosti semantičnih elementov; analiza z vidika:**
 - Tehnične (sintaktične) dostopnosti
 - Organizacijske dostopnosti
 - Stopnje uradnosti
 - Semantične bogatosti vira
- **Umestitev šifrantov in njihovih vrednosti v UOPGR taksonomijo**
 - Uparjanje atributnih tipov/vrednosti z koncepti UOPGR
 - Grafični prikaz umeščenosti v UOPGR
 - Analiza semantične globine ter razpršenosti konceptov.

3.1.2 Opredelitev podatkov za analizo

Pred vsako analizo je potrebno opredeliti natančno katere elemente smo uporabili. Količina virov (podatki in dokumentacija), v obravnavanih evidencah je zelo obsežna, zato smo izbrali tiste reprezentativne primere, na podlagi katerih smo dobili celotno sliko za celotno evidenco. Drugi kriterij za izbor podatkovnega vira, je bil kako smo jih lahko primerjali z tematsko podobnimi viri iz ostalih evidenc. S tem smo pokazali možnost uporabe referenčnega semantičnega sistema UOPGR.

3.1.3 Ocena pogojev za semantično povezovanje

Namen ocene pogojev za semantično povezovanje pri analizi semantične povezljivosti je zagotavljanje sistematične in temeljite osnove za razumevanje, kako in ali je možno med seboj povezati različne podatkovne zbirke. Ter katere nesemantične pogoje moramo izpolniti, da bo semantično povezovanje sploh mogoče. Skozi analizo obstoja izbranih kriterijev smo preučili tako tehnične kot vsebinske omejitve, ki vplivajo na povezovanje podatkov. To vključuje razumevanje obstoječih podatkovnih struktur, formatov in standardov.

3.1.3.1 Sintaktična povezljivost:

Analizirane so tehnične lastnosti podatkov, kot so formati zapisa geometrije in šifrantov, formati zapisa in dostopnost dodatne dokumentacije, Enostavnost oziroma Kompleksnost povezovanja šifrantov v evidenci, obstoj enoličnih identifikatorjev ter strojna berljivost in dostopnost vira glede na 5- zvezdično Berner's Lee lestvico. Poznavanje teh pogojev, nam pove kakšna je tehnična možnost povezovanja semantičnih elementov. Z oceno pogojev se identificirajo potrebni koraki za pripravo podatkov na semantično povezovanje. To vključuje morebitne izboljšave v podatkovnih zbirkah ali dokumentaciji.

3.1.3.2 Opis elementov kakovosti:

Analiza obstoja elementov kakovosti pripomore k zanesljivosti, uporabnosti in učinkoviti obdelavi prostorskih podatkov, kar je ključno za semantično povezovanje in analizo. V tej analizi smo ocenjevali obstoj in kakovost metapodatkov, kar vključuje informacije o izvoru, metodologiji zajema, časovni natančnosti in konsistentnosti. Dobra kakovost podatkov je osnovni pogoj za uspešno semantično povezovanje. Na osnovi te informacije lažje sprejmemo ukrepe, ki zagotovijo višjo povezljivost, oziroma se lažje odločamo ali je povezovanje sploh smiselno.

3.1.3.3 Obstoj semantičnih elementov:

Za semantično povezovanje morajo najprej obstajati osnovni ali celo bolj kompleksni semantični elementi. V tej analizi smo, za različne ravni (evidenca, podatkovni sloj oziroma Kategorija ter šifrante) ugotavljali obstoj semantičnih elementov. Iskali smo:

- obstoj metapodatkov v standardizirani strojni obliki,
- obstoj ključnih besed za evidenco,
- dokumentacije z opisi za opredelitev konceptov, entitet ter njihovih lastnosti in relacij,
- povezanost v širše klasifikacije, kar že samo po sebi povečuje povezljivost,
- dokumentacijo o kontekstu nastanka, uporabe ter vzdrževanja evidence, ter njenih podatkov in strukture
- ter obstoj naprednih, strojno berljivih semantičnih elementov, kot so ontologije, taksonomije, vokabularji, enolični identifikatorji, podatkovni modeli, semantične anotacije, itd.

Na osnovi teh informacije lažje sprejmemo ukrepe, ki zagotovijo višjo povezljivost, oziroma se lažje odločamo, ali je povezovanje sploh smiselno.

S tem pristopom se povečuje transparentnost in kakovost povezovanja prostorskih podatkov, kar omogoča boljše analize in odločanje na podlagi teh podatkov. Ocenjevanje pogojev predstavlja ključen korak v procesu zagotavljanja, da so podatki ne le tehnično povezljivi, ampak tudi smiselni in zanesljivi za uporabnike.

3.1.4 Mere dostopnost semantičnih elementov

Na podlagi analize pogojev, smo razvili kvantitativne mere za dostopnost semantičnih elementov, ki ocenjujejo tehnično oziroma Sintaktično in organizacijsko dostopnost, stopnjo uradnosti vira oziroma Evidence, semantično bogatost vira ter časovna umeščenost – ali koliko je podatek oziroma Dokumentacija stara.

Skupni največji možni seštevek točk vseh kategorij za dostopnost semantičnih elementov je 29. Mera dostopnosti semantičnih elementov se na koncu izrazi kot delež, izražen v odstotkih. Za primer glej Preglednica 35.

3.1.4.1 Tehnična dostopnost vira pomena (Sintaktična dostopnost)

Na podlagi predhodne analize sintaktične povezljivosti razvili lestvico tehnične dostopnosti, ki temelji na prej omenjeni 5-zvezdični klasifikaciji in jo nadgrajuje z dodatnimi elementi za vire, ki niso v digitalni obliki.

Preglednica 26 - Razredi tehnične dostopnosti vira pomena (sintaktična dostopnost)

Sintaktična dostopnost:	Točke
Ni informacij o dodatnih opisih	0
Ekspertno znanje - V glavah ljudi (potreben intervju)	1
Analogni viri in digitalizirani viri (pisni skenirani, video, avdio, diagrami)	2
Pisni viri v digitalni obliki	3
Strukturirani viri (lastni formati)	4
Lokalni Standardizirani strukturirani viri v digitalni obliki (spletne strani)	5
Spletno dostopni standardizirani servisi	6

3.1.4.2 Organizacijska dostopnost vira

Organizacijska dostopnost vira nam daje informacijo, kako široko je lahko vsebina sprejeta, kar daje dodatno informacijo o odprtosti in skladnosti podatkov ter dokumentacije. Podatki so manj povezljivi, če moramo za dokumentacijo prositi upravitelja evidence, če sploh vemo, da obstajajo.

Preglednica 27 - Razredi organizacijske dostopnosti vira

Organizacijska dostopnost vira:	Točke
Opisi so del dokumentacije izvajalca ali nepovezani ostali viri	1
Opisi so del dokumentacije evidence Organizacije	2
Opisi so del dokumentacije evidence v vsebinski domeni (strokovna javnost)	3
Zunanji vokabularji (splošni, specialni) Nacionalni in nacionalno priznana vsebina	4
Splošno mednarodno dostopna in sprejeta vsebina	5

3.1.4.3 Stopnja uradnosti vira

Stopnja uradnosti vira nam daje informacijo o splošni sprejetosti in skladnosti vsebine. Najvišjo skladnost predstavljajo *pravne podlage*.

Preglednica 28 – Razredi, stopnje uradnosti semantičnih virov

Stopnja uradnosti:	Točke
Znanstveni in strokovni članki	1
Splošni vokabularji (Nacionalni in mednarodni besednjaki)	2
opisi so del besedil znanstvenih in strokovnih knjig, učbenikov, priročnikov ter tehničnih načrtov, risb in skic	3
Opisi so del strokovnih specifikacij in posebnih vokabularjev	4
opisi so del standardov	5
Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti, Pravna praksa (precedenčne sodbe))	6

3.1.4.4 Semantična bogatost vira

Bolj kot je vir *podrobno opisan*, bolj je razumljiv in nedvoumno opredeljen (Preglednica 29).

Preglednica 29 – Razredi semantične bogatosti vira

Semantična bogatost vira:	Točke
enobesedni opis	1
stavek, izraz	2
odstavek	3
podroben opis s primeri	4
podroben opis s primeri, skicami ali slikovnim gradivom	5
podroben opis s primeri, skicami s kotiranjem ali slikovnim gradivom	6

3.1.4.5 Starost vira

Starejši, kot je vir, manjša je možnost, da je še vedno veljaven ali relevanten (Preglednica 30). Trenutno veljavni viri imajo lahko najvišjo stopnjo zaupanja, čeprav so bili napisani pred več leti.

Preglednica 30 - Razredi starosti virov za semantično povezovanje

Starost vira:	Točke
Neznana starost	0
starejše od 10 let	1
med 5 in 10 let	2
med 3 in 5 let	3
do 3 leta	4
Trenutno veljavno	5

3.1.5 Umestitev šifrantov in njihovih vrednosti v UOPGR taksonomijo

V tej analizi, vrednosti atributov in tematiko posamezne evidence, podatkovnega sloja ali kategorije, ter določene vrednosti v šifrantih umestimo v razvit semantični referenčni okvir UOPGR (Univerzalna Ontologija Prostora Geografskih razsežnosti), kot je bil opisan v 2. Poglavju.

Postopek analize je obsegal kartiranje vrednosti šifrantov posameznega sloja oziroma kategorije evidence s koncepti UOPGR. V naslednjem koraku smo izdelali grafični prikaz mapiranja na polarnem grafu, ter na koncu ovrednotili število vseh konceptov, njihovo semantično globino in razpršenost po kategorijah.

3.2 Rezultati naloge 3.1: Testna analiza možnosti semantične povezljivosti Topografskih zbirk

Namen tega razdelka je opis izvedbe testne analize možnosti semantične povezljivosti topografskih zbirk podatkov (občinska in državna raven). Med raziskavo možnih virov podatkov za analizo, z več občin smo dobili informacijo, da na občinski ravni ne vodijo več svojih topografskih baz podatkov, temveč uporabljajo podlage GU.

Zato smo se odločili za izvedbo interne analize izbranih topografskih slojev in evidentiranje njihovih možnosti za semantično povezovanje. Možnost povezovanja smo izvedli naknadno z ostalimi analiziranimi podatki v ostalih poglavjih.

3.2.1 Opredelitev podatkov za analizo topografskih zbirk

Glede na možnosti kasnejše primerjave semantičnega povezovanja, smo v skupini podatkov zbirke topografskih podatkov, za analizo semantične povezljivosti izbrali naslednje podatkovne sloje, navedene v Preglednica 31:

Preglednica 31 - Opredelitev podatkov Topografske zbirke za analizo

Ime sloja	Opis	Vir
BU_STAVBE_P.shp	Podatkovni sloj Stavbe (Poligon), objektnega področja ZGRADBE	https://ipi.eprostor.gov.si/jgp/data
TN_CESTE_L.shp	Podatkovni sloj Cesta (Linija), objektnega področja PROMETNA OMREŽJA	https://ipi.eprostor.gov.si/jgp/data
LU_PokritostTal.shp	Podatkovni sloj Pokritost tal (Linija), objektnega področja POKRITOST TAL	https://ipi.eprostor.gov.si/jgp/data

Te podatkovne sloje smo izbrali zato, da smo jih lahko povezovali z ostalimi sloji, ki se nanašajo na *stavbe, ceste in pokritost tal*, ki so bila eni od ciljnih konceptov analize.

Nabor virov za analizo obsega vse gradivo, ki smo ga pridobili s prosto dostopnih virov na spletu, kot tudi dodatno dokumentacijo, ki smo jo pridobili v komunikaciji z upravljavcem zbirke.

V naslednjem poglavju smo prikazali analizo obstoja elementov, ki predstavljajo pogoj za semantično povezovanje. To so informacije o tehnični, ozirom sintaktični povezljivosti, o obstoju izbranih elementov kakovosti in obstoju semantičnih elementov.

3.2.2 Ocena pogojev za semantično povezovanje topografskih zbirk

Da bi podatke med seboj lahko semantično povezovali, je treba poznati tudi ostale tehnične in vsebinske omejitve, ki se ne nanašajo samo na razumevanje okoliščin, konteksta. Zato smo pred samo analizo povezljivosti semantičnih elementov, opravili pregled tehnično - sintaktičnih lastnosti, ki omogočajo povezovanje, obstoj informacij o izbranih elementih kakovosti, ter pregled obstoja semantičnih elementov.

3.2.2.1 Sintaktična povezljivost topografskih zbirk

V spodnji tabeli smo opredelili tehnične oziroma sintaktične elemente za izbrane sloje zbirke topografskih podatkov, ki posledično vplivajo na semantično povezljivost zbirke. Določili smo:

- formate zapisa geometrije in pripadajoče dokumentacije,
- ocenili razčlenjenost zbirke na eno ali več tabel, kar predstavlja enostavnost oziroma dodatno kompleksnost pri semantičnem povezovanju,
- ali za posamezni zapis objekta obstaja enolični identifikator in na kakšni ravni (lokalno, državno, globalno),
- na kakšni ravni je vsebina, glede na 5 zvezdični kriterij (Berners) strojne berljivosti podatkov.

S tem smo opredelili, kako dobro so objavljeni podatki pripravljeni za strojno razumevanje in povezovanje. V primeru podatkov v topografski zbirki, za vse sloje veljajo enake lastnosti, zato vrednosti v tabeli veljajo za vse tri obravnavane sloje.

Preglednica 32 - Sintaktična povezljivost DTK Evidence, slojev in šifrantov (lastne prikaz)

Sintaktična povezljivost: DTK			
KRITERIJ	Opis Evidence DTK	Podatkovni sloj/kategorija	Šifranti, Atributi
Format zapisa geometričnega sloja in atributov	<i>Ne nanaša se na to kategorijo</i>	Esri shp ATOM	Json, csv
Dodatna dokumentacija	Inspire Metapodatki Docx	Pdf	Pdf
Enostavnost/ Kompleksnost	<i>Ne nanaša se na to kategorijo</i>	Kompleksno (1:n)	<i>Ne nanaša se na to kategorijo</i>
Obstoj enoličnega identifikatorja:	Inspire Metapodatki	SID – Buildings (državno) ST_CES – Ceste (državno) LC_DTM_ID (Organizacija)	<i>Ne nanaša se na to kategorijo</i>
Strojna berljivost (5 Star)	4 – RDF 1 – Docx	1 – Shp, Pdf	3 – Json, csv 1 - Pdf

Iz tabele je razvidno, da je večina obravnavanih podatkov in dokumentacije sicer na voljo v strojno berljivi obliki, a pretežno na osnovni ravni. Izstopa le metapodatkovni opis, ki ga je možno izvoziti v obliki RDF in šifranti, ki so zapisani v standardizirani odprtokodni obliki (Json in csv). Prav tako je podrobnejša dokumentacija o opisu zaloge vrednosti atributov in snemalnih pravih dostopna šele na zahtevo. Sama evidenca je sestavljena kompleksno, kar pomeni, da je treba geometrični sloj povezati z zunanjimi šifranti, a je ključ za povezavo dovolj podrobno opisan v dodatni dokumentaciji. Vsi obravnavani objekti imajo unikatne ID-je, pri čemer sloja Stavbe in Ceste vsebujeta ID-je, ki jih povezujejo z ostalimi tematsko skladnimi evidencami v državi. S tehničnega vidika je semantično povezovanje topografskih podatkov možno ob predhodni pripravi podatkov.

V nadaljevanju smo analizirali obstoj elementov kakovosti, ki so ključnega pomena za semantično povezovanje.

3.2.2.2 Dokumentacija elementov kakovosti topografskih podatkov

V tem delu smo opredelili opis izbranih elementov kakovosti podatkovnega vira za topografske podatke, ki posledično vplivajo na njeno uporabnost pri semantičnem povezovanju predvsem z vidika opredelitve konteksta. Določili smo:

- obstoj informacij o izvoru podatkovnega vira (angl. lineage) na ravni geometrije podatkov o lastnostih objekta (atributov), o zgodovini transformacij podatkov in metodologijah zajema,
- kako je opisana časovna umeščenost veljavnosti informacij o geometriji in lastnostih objekta,
- ter o konsistentnosti objave evidence, ter atributov v sloju.

Preglednica 33 - Pregled dokumentiranosti elementov kakovosti za semantično povezovanje podatkov DTK

Elementi kakovosti: DTK		
KRITERIJA	Elementi kakovosti	Pomanjkljivosti
Izvor, metodologija	Opis metodologije v DTM_Navodila_1.6.pdf; Dodatno v M_Pokritost_tal_1_0.docx Metoda v atributu [MET_ZAJ]	/
Natančnost geometrije	1:5000	/
Časovna umeščenost	Atribut [D_VIR]	Arhivskih podatkov ni na voljo
Konsistentnost podatkov v tabeli	Vsi podatki atributov so izpolnjeni	Večina vrednosti RABA_STAV; VRSTA_STAV; ST_CES =0

Z analizo smo ugotovili, da so z vidika izvora podatkov in metodologije zajema, v podatkovnem modelu ti elementi prisotni. Hkrati je metodologija zajema podrobno opisana v dodatni dokumentaciji, ki jo je dostavil upravljavec. Z vidika deklarirane natančnosti je v metapodatkih le ta navedena kot podatek o merilu, in sicer 1:5000. Časovna umeščenost je določena za vsak objekt v tabeli z atributom D_VIR, hkrati imamo še atribut ST_POS (ni jasno opisan pomen kratice), ki nas verjetno usmeri na dokumentacijo o izvedbi postopka spremembe, če bi to potrebovali. Arhivskih vrednosti za nazaj javno, neposredno ni dostopnih. Z vidika konsistentnosti podatkov, so vsi atributi izpolnjeni, le opažamo, da imamo pri nekaterih atributih pri večini objektov vrednost 0.

Iskani elementi kakovosti so opisani in z vidika semantične povezljivosti podajajo kontekst, v katerem podatke lahko uporabljamo. Arhivski viri niso na voljo na spletu, a atributi nakazujejo, da bi jih, če bi bilo treba, lahko pridobil.

V naslednjem poglavju smo analizirali obstoj semantičnih elementov topografskih.

3.2.2.3 Obstoj semantičnih elementov topografskih podatkov

V tem delu smo opredelili obstoj semantičnih elementov za izbrane sloje zbirke topografskih podatkov, ki omogočajo razumevanje konteksta in so ključni za semantično povezljivost podatkov. Preverili smo:

- obstoj metapodatkov,
- obstoj ključnih besed, opisov in dokumentacije na ravni evidence, posameznih slojev,
- opredelitev entitet oziroma objektov, ter pomenov besed, ki opisujejo posamezne entitete,
- povezanost v širše sprejete klasifikacije (ontologije, taksonomije, šifranti, vokabularji na nacionalnem oziroma mednarodni ravni),
- dokumentiranost konteksta uporabe, kot so uporabniški primeri, okoliščine nastanka, pravne podlage, vrste deležnikov,

- obstoj naprednih semantičnih elementov, kot so ontologije, taksonomije, grafi znanja na ravni organizacije, evidence, sloja ali tipa objektov.

S tem smo opredelili, kako dobro so objavljeni podatki pripravljeni za strojno razumevanje.

Preglednica 34 - Pregled obstoja semantičnih elementov v dokumentaciji evidence DTK

Semantični elementi DTK Slojev			
OPIS KRITERIJA	Evidenca	Podatkovni sloj	Šifranti
Obstoj metapodatkov	Izpolnjeni INSPIRE metapodatki	Razvidno iz navodil za zajem	Podatkovni katalog
Opisi , ključne besede, dokumentacija	Izpolnjeni INSPIRE metapodatki	Podatkovni katalog; navodila za zajem	<i>Ne nanaša se na to kategorijo</i>
Opredelitev entitet	<i>Ne nanaša se na to kategorijo</i>	Opredeljene v obliki stavkov oziroma izrazov.	Opredeljene v obliki stavkov oziroma izrazov.
Povezanost v širše klasifikacije	Del Inspire Tematike	/	Stavbe.SID in Stavbe.KN_STANJE > povezava na Kat. Stavb. Ceste.KAT_CES - delno navezava na Uradni List
Kontekst	Del Inspire Tematike Ni uporabnih primerov Deležniki so delno opisani v navodilih; Zakonodaja je dostopna, a več klikov stran	/	Zgolj referenca na pravilnik o Kategorizaciji cest
Napredni sem. elementi	/	/	/

Z analizo semantičnih elementov topografskih podatkov smo ugotovili, naslednje:

- *Metapodatki* za topografske podatke na ravni evidence obstajajo in so zapisani in objavljeni skladno z zahtevami INSPIRE (objavljeno v strojno berljivi in strukturirani obliki). To omogoča hitrejše strojno odkrivanje evidence in njeno uporabo. Na ravni posameznih slojev strojno opisanih metapodatkov nismo našli, obstaja pa gradivo v digitalni obliki, iz katerega je te informacije možno pridobiti. Na ravni šifrantov, prav tako obstajajo podrobni opisi v dodatni dokumentaciji, ki nam jo je na zahtevo dostavil upravljavec, a je na prosto dostopnih spletnih straneh ni.
- *Opisi, ključne besede in dokumentacija ter opredeljenost entitet*. Na ravni evidence, so v metapodatkih INSPIRE, ključne besede navedene in strojno dostopne preko portala. Na ravni posameznih slojev in opisov, ključnih besed ni eksplicitno navedenih, so pa podrobni opisi objektnih tipov in vrednosti atributov jasno in podrobno opisani v navodilih za zajem in vzdrževanje.
- *Povezanost v širše klasifikacije*, je na ravni evidence razvidna iz umeščenost v tematske sklope INSPIRE. Na ravni posameznih analiziranih slojev in njihovih atributov, so določene atributne prevzete iz zunanjih zbirk: pri Stavbah iz Katastra nepremičnin in pri cestah iz Zbirnega katastra

gospodarske javne infrastrukture. Podrobnih informacij o povezanosti Pokrovnost tal v dostopni literaturi nismo našli, a je navedeno, da obstaja Priloga 2, ki naj bi to opisovala. Sloji zbirke topografskih podatkov so tudi del kart, ki jih izdeluje EuroGeographics.

- *Kontekst:* Na ravni evidence je kontekst v obliki povzetka vira opisan v Inspire metapodatkih, pravna podlaga pa je na voljo na strani o zakonodaji, ki ureja delo Geodetske uprave (Pravilnik o katalogu topografskih podatkov in topografskem ključu²). Podrobnih informacij o načinu uporabe in posebnem namenu posameznih slojev in atributov v tej obliki nismo zaznali, so pa v navodilih opisani deležniki, ki prispevajo podatke iz svojih evidenc.
- *Napredni semantični elementi:* O okviru dokumentacije evidence zbirke topografskih podatkov, za evidenco, njene sloje ali šifrante nismo našli naprednih semantičnih elementov, kot so ontologije, taksonomije, posebni strojno berljivi vokabularji in podobno.

Če povzamemo, smo ugotovili, da na ravni opisa evidence obstajajo zelo dobro strojno berljivi semantični elementi, ki so skladni z direktivo INSPIRE. Na ravni podatkov in šifrantov, pa na javno dostopnih straneh nimamo podrobnih informacij, a predvidevamo, da z njimi razpolaga upravljavec zbirke, čeprav le v najbolj osnovni strojni obliki. Naprednih oblik za enkrat še ni v odprti objavi. Z vidika vsebine, je za to zbirko na voljo veliko besedilnega gradiva, ki predstavljajo dobro osnovo za razumevanje zbirke, posameznih slojev in entitet.

V naslednjem razdelku smo, glede na razvite mere opredelili dostopnost do semantičnih elementov, s čimer smo dobili informacijo, kako dobro topografska zbirka podatkov izpolnjuje pogoje za semantično povezovanje z vidika skladnosti in dostopnosti.

² Pravilnik o katalogu topografskih podatkov in topografskem ključu - <https://pisrs.si/pregledPredpisa?id=PRAV12141>

3.2.3 Mere dostopnost semantičnih elementov topografskih podatkov

V tem delu smo ocenili dostopnost in verodostojnost semantičnih elementov za izbrane sloje zbirke topografskih podatkov, ki z vidika dostopnosti in semantične celovitosti in skladnosti, vplivajo na semantično povezljivost zbirk. Ocenili jih smo z izbranimi merami, kot so *tehnična (sintaktična)* dostopnost, *organizacijska* dostopnost, *stopnja uradnosti* in *semantična bogatost* vira. V sklopu posamezne zbirke smo primarno upoštevali vse vire, ki so bili, z visoko stopnjo odprtosti, pridobljeni na javnih spletnih straneh upravljavca evidence; sekundarno smo pozvali skrbnike evidence, da nam priskrbijo interno dokumentacijo.

Preglednica 35 - Analiza dostopnosti gradiva v evidenci in šifrantih DTK

DTK	Ocena za evidenco	Točke	Ocena za šifrante evidence	Točke
Tehnična dostopnost vira pomena (Sintaktična dostopnost)	Spletno dostopni standardizirani servisi	6	Pisni viri v digitalni obliki	3
Organizacijska dostopnost vira	Splošno mednarodno dostopna in sprejeta vsebina	5	Opisi so del dokumentacije evidence Organizacije	2
Stopnja uradnosti	Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti..., Pravna praksa (precedenčne sodbe))	6	Opisi so del strokovnih specifikacij in posebnih vokabularjev	4
Semantična bogatost vira	odstavek	3	Podroben opis s primeri, skicami ali slikovnim gradivom	5
Starost vira	Trenutno veljavno	5	do 3 leta	4
TOČK		25		18
DELEŽ(%)		86		62

Na ravni zbirke topografskih podatkov imamo strojno razumljive vire, ki so objavljeni tako na slovenskem portalu INSPIRE kot tudi na mednarodnih portalih. Omogočen je strojni dostop do opisov RDF, kar je že strojno zelo razumljiva raven. Zato je ocena teh virov dosegla 86 točk od 100, kar že predstavlja visoko oceno za semantično povezljivost.

Na ravni posamičnih slojev in šifrantov, so podatki manj dostopni, saj smo morali za dodatne opise naknadno pozvati upravljavca zbirke, a so kljub tehnično in organizacijsko slabši dostopnosti semantično zelo bogati.

V naslednjem poglavju smo obravnavane podatke umestili v razviti semantični referenčni okvir, kar nam bo v nadaljevanju pomagalo pri oceni možnosti povezovanja tematsko podobnih evidenc glede na njihovo tematsko prekrivanje.

3.2.4 Umestitev šifrantov in njihovih vrednosti DTK v UOPGR taksonomijo

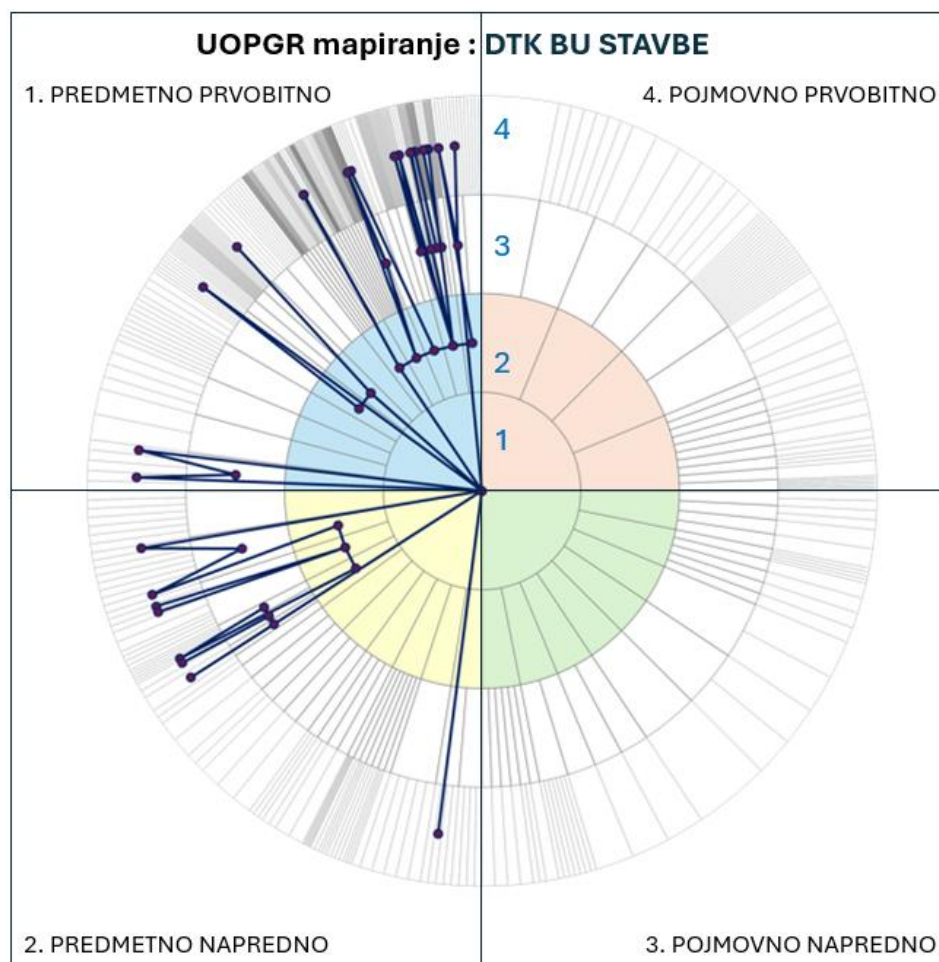
V tem delu smo za izbrane podatke (DTK Stavbe, DTK Ceste in DTK_PokritostTal) in vrednosti v njihovih šifrantov umestili v razviti semantični referenčni sistem UOPGR, glede na njihovo tematiko in semantično globino.

3.2.4.1 DTK Stavbe

Preglednica 36 - Umeščenosti vrednosti šifrantov sloja DTK Stavbe v UOPGR s semantično globino

Šifra UOPGR	Opis kategorije UOPGR	Sem. globina	RABA_ID	RABA_OPIS
10102	Nevsakodnevno in sezonsko bivanje	3	3231	gozdarska koča
10102	Nevsakodnevno in sezonsko bivanje	3	3452	planinski bivač
10102	Nevsakodnevno in sezonsko bivanje	3	3454	lovski bivač
1010202	uta	4	17	uta
1010207	prenočišče	4	3451	planinska koča
1010207	prenočišče	4	3453	lovska koča
1020201	zajetje	4	3821	vodno zajetje
1020202	črpališče	4	3822	vodno črpališče
1020203	zbiralnik	4	3823	vodohran
1020314	rastlinjak	4	5	rastlinjak
1020405	silos	4	6	silos
1020406	kozolec	4	11	kozolec
1020601	elektrarna	4	3811	termoelektrarna
1020601	elektrarna	4	3812	jedrska elektrarna
1020605	transformator	4	9	transformator
103	Objekti aktivnosti PROIZVODNJE izdelkov (industrijsko/obratno)	2	16	trajni montažni objekt
103	Objekti aktivnosti PROIZVODNJE izdelkov (industrijsko/obratno)	2	3511	industrijski objekt
103	Objekti aktivnosti PROIZVODNJE izdelkov (industrijsko/obratno)	2	3621	dvorana, hala
1030404	rezervoar	4	7	skladiščni rezervoar
10401	Objekti aktivnosti PREMOSKANJA	3	14	arkade
10401	Objekti aktivnosti PREMOSKANJA	3	15	pasaža
1040106	brezina	4	13	svetilnik
1050201	lokal	4	3771	bencinski servis
1050202	skladišče	4	3441	skladišče
1070101	stadion	4	12	stadion
1080103	nadstrešek	4	10	nadstrešek
1120001	cerkev	4	2	mošeja
1120001	cerkev	4	3	cerkev
1120004	kapelica	4	8	kapelica
2020001	šola	4	3351	šola
2020008	vrtec	4	3352	vrtec
2030102	ambulanta	4	3312	zdravstveni dom
2030103	lekarna	4	3313	lekarna

2030201	bolnica	4	3311	bolnica
2040105	Bunker	4	1	bunker
2040201	zaklonisce	4	3921	3921 gasilski dom
2040301	Policijska Postaja	4	3341	policijska postaja
2100102	Grad	4	4	grad



Slika 16 - Graf umeščenosti vrednosti šifrantov sloja DTK Stavbe v semantični referenčni sistem UOPGR

Semantični referenčni sistem UOPGR se deli na 4 kategorije, ki se skupno delijo na 34 razredov, ki se delijo skupno na 105 podrazredov in obsegajo 446 entitet (objektov, shem ali pojavov) v četrti globini. Referenčni sistem obsega skupno 590 konceptov.

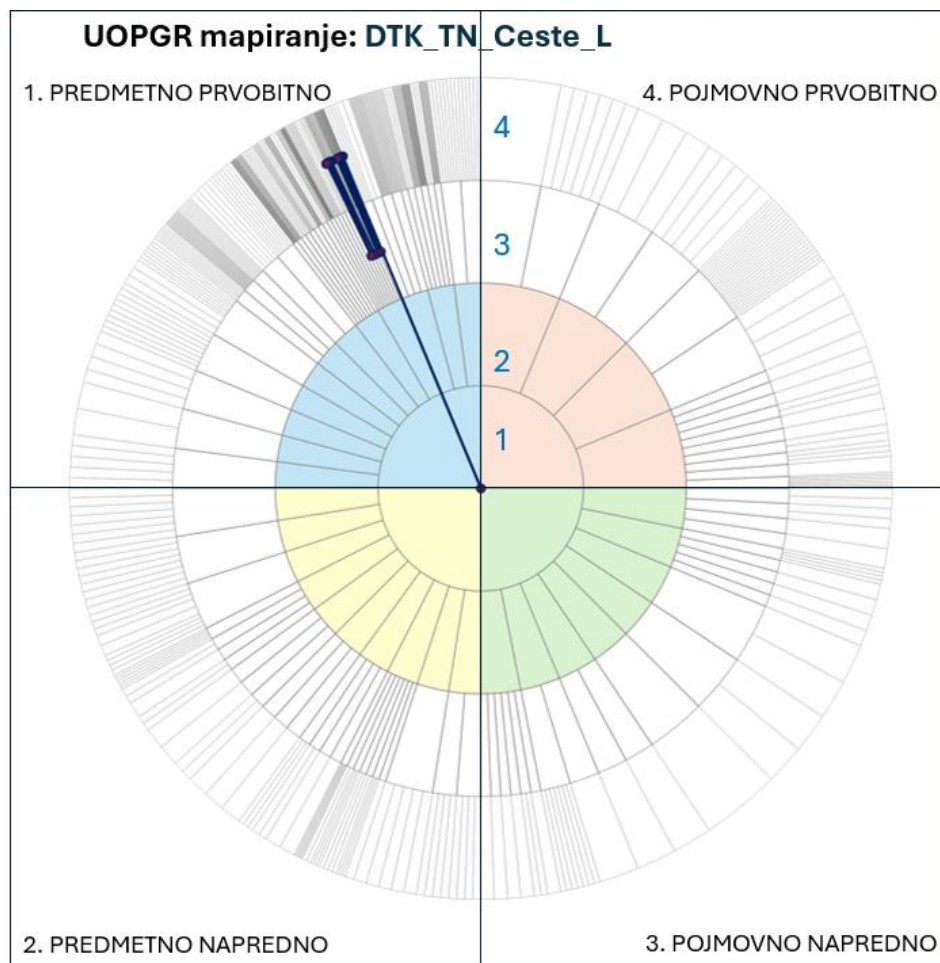
V podatkovnem sloju DTK Stavbe, smo identificirali 20 opisov iz šifrantov, ki ustrezajo entitetam na 4. semantični globini UOPGR. Zaradi presplošnega opisa smo uvrstili 5 opisov v šifrantih v podrazrede na 3. globini in 2 opisa v razred na 2. globini (glej Slika 16 - Graf umeščenosti vrednosti šifrantov sloja DTK Stavbe v semantični referenčni sistem UOPGR). Umeščanje smo izvedli glede na trenutno raven razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov, bodo te povezave lahko bolj številčne in bolj bogate in s tem doprinesle k več možnostim povezovanja.

Na grafu je razvidno, da se večina opisi iz šifrantov umeščajo med Predmetne objekte Prvobitnih človeških aktivnosti (1. Kategorija) in Predmetne objekte Naprednih človeških aktivnosti (2. Kategorija). V 3. in 4. kategoriji ni nobenega opisa.

3.2.4.2 Umeščenost sloja DTK Ceste (linije) v semantični referenčni sistem

Preglednica 37 - Umeščenosti vrednosti šifrantov sloja DTK Ceste v UOPGR s semantično globino

Šifra UOPGR	Opis kategorije UOPGR	Sem. globina	RABA_ID	RABA_OPIS
1040101	most	4	1	brv
1040101	most	4	3	most
1040102	predor	4	8	predor
1040107	nadvoz	4	5	nadvoz
1040108	podvoz	4	7	podvoz
1040110	viadukt	4	9	viadukt
1040111	galerija	4	2	galerija
1040205	svetilnik	4	PP	planinska pot
1040205	svetilnik	4	11	povezava skozi peščono
1040205	svetilnik	4	12	pešpot
1040305	podhod	4	6	podhod
1040306	nadhod	4	4	nadhod
1040401	cesta	4	13	servisna cesta
1040401	cesta	4	AC	avtocesta
1040401	cesta	4	HC	hitra cesta
1040401	cesta	4	G1	glavna cesta I. reda
1040401	cesta	4	G2	glavna cesta II. reda
1040401	cesta	4	R1	regionalna cesta I. reda
1040401	cesta	4	R2	regionalna cesta II. reda
1040401	cesta	4	R3	regionalna cesta III. reda
1040401	cesta	4	RT	turistična cesta
1040401	cesta	4	LC	lokalna cesta
1040401	cesta	4	LG	glavna mestna cesta
1040401	cesta	4	LZ	zbirna mestna ali krajevna cesta
1040401	cesta	4	LK	mestna ali krajevna cesta
1040401	cesta	4	GZ	gozdna cesta
1040401	cesta	4	NK	nekategorizirana cesta
1040401	cesta	4	1	avtocesta
1040401	cesta	4	6	hitra cesta
1040402	vozišče	4	5	enovito cestišče
1040402	vozišče	4	10	ločeno cestišče
1040403	prikljucek	4	2	cestni prikljucek
1040403	prikljucek	4	3	dostop do parkirišča
1040403	prikljucek	4	4	dostop do počivališča
1040406	krozišce	4	9	krožišče
1040407	pot	4	14	povezava čez dvorišče
1040407	pot	4	JP	javna pot
1040407	pot	4	KD	daljinska kolesarska pot
1040407	pot	4	KG	glavna kolesarska pot
1040407	pot	4	KJ	javna pot za kolesarje
1040407	pot	4	7	kolesarska pot
1040407	pot	4	8	kolovoz



Slika 17 - Graf umeščenosti vrednosti šifrantov sloja DTK Ceste v UOPGR

V podatkovnem sloju DTK Ceste, smo identificirali 42 opisov iz šifrantov, ki vsi ustrezajo entitetam na 4. semantični globini UOPGR (Slika 17).

Umeščanje smo izvedli glede na trenutno raven razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov bodo te povezave lahko bolj številčne in bolj bogate in s tem prispevale k več možnostim povezovanja.

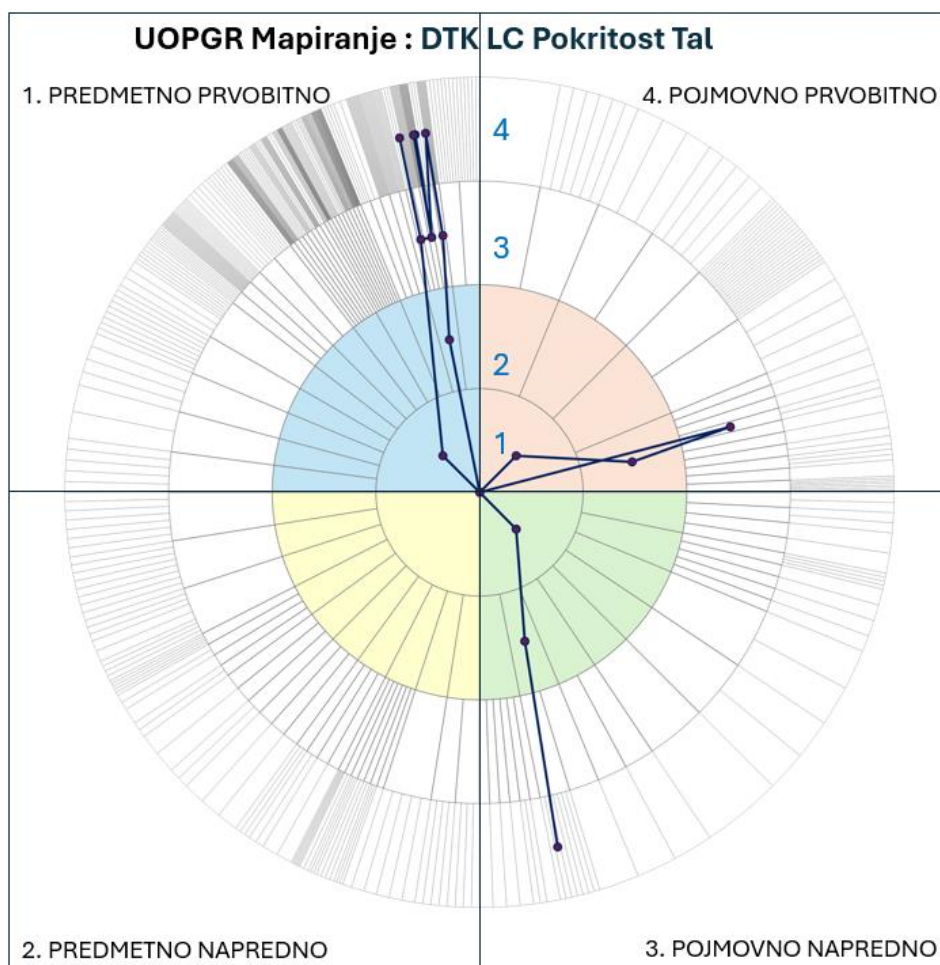
Na grafu je razvidno, da se vsi opisi iz šifrantov umeščajo med Predmetne objekte Prvobitnih človeških aktivnosti (1. Kategorija). V 2., 3., in 4. kategoriji ni nobenega opisa.

To nakazuje, da so zaloge vrednosti zelo dobro opisane v tem referenčnem modelu, pod drugi strani pa so zelo specifične.

3.2.4.3 Umeščenost sloja DTK PokritostTal v semantični referenčni sistem

Preglednica 38 - Umeščenosti vrednosti šifrantov sloja DTK Pokritost tal v UOPGR s semantično globino

Šifra UOPGR	Opis kategorije UOPGR	Sem. globina	RABA_ID	RABA_OPIS
1020109	soline	4	14	soline
1020303	hmeljisce	4	25	hmeljišče
1020304	olcnjak	4	24	oljčnik
1020305	sadovnjak	4	22	sadovnjak
1020306	vinograd	4	23	vinograd
1020501	gozd	4	21	gozd
1020501	gozd	4	26	grmovje
3020102	VODA	4	31	celinske vode
3020102	VODA	4	32	morje
40107	Opazovanje naravnih okolij - habitatov	3	11	skale
40107	Opazovanje naravnih okolij - habitatov	3	12	kamnita tla
40107	Opazovanje naravnih okolij - habitatov	3	13	prodišča in sipine
40107	Opazovanje naravnih okolij - habitatov	3	27	rušje
40107	Opazovanje naravnih okolij - habitatov	3	33	ledenik



Slika 18 - Graf umeščenosti vrednosti šifrantov sloja DTK Pokritost Tal v UOPGR

V podatkovnem sloju DTK Stavbe, smo identificirali 9 opisov iz šifrantov, ki ustrezajo entitetam na 4. semantični globini UOPGR. Zaradi presplošnega opisa smo uvrstili 5 opisov v šifrantih v podrazred na 3. globini (Slika 18).

Umeščanje smo izvedli glede na trenutno raven razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov bodo te povezave lahko bolj številčne in bolj bogate in s tem prispevale k več možnostim povezovanja.

Na grafu je razvidno, da se opisi iz šifrantov umeščajo med Predmetne objekte Prvobitnih človeških aktivnosti (1. kategorija), med Pojemovne sheme Naprednih človeških aktivnosti (3. kategorija) in med Pojemovne sheme Prvobitnih človeških aktivnosti (4. kategorija). V 2. kategoriji ni nobenega opisa.

3.3 Rezultati naloge 3.2: Testna analiza možnosti interne semantične povezljivosti nepremičninskih zbirk podatkov GURS

V tem razdelku analiziramo semantiko dveh testnih zbirk podatkov oziroma njunih šifrantov ter opisov, in sicer *katastrske evidence*, ter *evidence množičnega vrednotenja*.

V tem razdelku analiziramo semantiko dveh testnih zbirk podatkov oziroma njunih šifrantov ter opisov, in sicer *katastrske evidence*, ter *evidence množičnega vrednotenja*. Analizirali smo zadnje različice tehnične dokumentacije IS Kataster, ki smo jo pridobili s strani naročnika, in sicer Tehnična dokumentacija, (18.1.2024) ver. 1.22 (Izmenjevalni formati v 1.22 – šifranti OK).

3.3.1 Opredelitev podatkov za analizo

Za namen primerjave semantičnega povezovanja, ki bo izvedena na koncu analize, smo v skupini podatkov zbirke *nepremičninskih* podatkov, za analizo semantične povezljivosti izbrali naslednje podatkovne sloje:

Preglednica 39 - Opredelitev izbranih slojev/kategorij Katastra nepremičnin za analizo semantične povezljivosti

Ime sloja/kategorija	Opis	Vir
Kataster nepremičnin - <i>Parcele</i>	Podatki o ZEMLJIŠČIH	Portal prostor, Javni vpogled v nepremičnine
Kataster nepremičnin - <i>Tloris stavbe na parceli</i>	Podatki o ZEMLJIŠČU POD STAVBO	Portal prostor, Javni vpogled v nepremičnine
Kataster nepremičnin - <i>Stavbe</i>	Podatki o STAVBAH	Portal prostor, Javni vpogled v nepremičnine

Te podatkovne sloje smo izbrali zato, ker predstavljajo osnovne nepremičninske koncepte in tudi zato, ker v testnem vzorcu drugih tematik tudi obravnavamo stavbe.

Nabor virov za analizo obsega vse gradivo, ki smo ga pridobili s prosto dostopnih virov na spletu, kot tudi dodatno dokumentacijo, ki smo jo pridobili v komunikaciji z upravljavcem zbirke.

V naslednjem poglavju, smo analizirali obstoj elementov, ki predstavljajo pogoj za semantično povezovanje. To so informacije o tehnični oziroma sintaktični povezljivosti, obstoj izbranih elementov kakovosti in obstoj semantičnih elementov.

3.3.2 Ocena pogojev za semantično povezovanje nepremičninskih podatkov

Da bi podatke med seboj lahko semantično povezovali, je treba poznati tudi ostale tehnične in vsebinske omejitve, ki se ne nanašajo samo na razumevanje konteksta. Zato smo pred samo analizo povezljivosti semantičnih elementov *nepremičninskih zbirk podatkov*, opravili pregled tehničnih oziroma sintaktičnih lastnosti, ki omogočajo povezovanje, obstoj informacij o izbranih elementih kakovosti, ter pregled obstoja semantičnih elementov.

3.3.2.1 Sintaktična povezljivost nepremičninskih podatkov

V spodnji tabeli smo opredelili tehnične oziroma sintaktične elemente za izbrane sloje zbirke nepremičninskih podatkov, ki posledično vplivajo na semantično povezljivost. Analizo smo izvedli za raven celotne evidence in vključuje tri omenjene kategorije oziroma sloje. Morebitne razlike so bile opisane v povzetku

Določili smo:

- formate zapisa geometrije in pripadajoče dokumentacije,

- ocenili razčlenjenost zbirke na eno ali več tabel, kar predstavlja enostavnost oziroma dodatno kompleksnost pri semantičnem povezovanju,
- ali za posamezni zapis objekta obstaja enolični identifikator in na kakšni ravni (lokalno, državno, globalno),
- na kakšni ravni je vsebina, glede na »5 zvezdni« merilo strojne berljivosti podatkov.

S tem smo opredelili, kako dobro so objavljeni podatki pripravljeni za strojno razumevanje. V primeru nepremičninske zbirke podatkov so opredeljene kategorije informacijskih objektov, ki niso zasnovane v obliki slojev (lahko pa se kategorije sicer predstavijo kot sloje). Kot sloj so obravnavani gostujoči podatki (na primer dejanska raba).

Vir: Izmenjevalni formati za vpis podatkov v kataster nepremičnin. Pravilnik o vodenju podatkov katastra nepremičnin (Uradni list RS, št. 41/22)

Preglednica 40 - Sintaktična povezljivost evidence katastrov nepremičnin

Sintaktična povezljivost: IS Kataster			
KRITERIJ	Opis Evidence	Podatkovni sloj/kategorija	Šifranti, Atributi
Format zapisa geometričnega sloja in atributov	<i>Ne nanaša se na to kategorijo</i>	ESRI Shapefile WMS, WFS, GML, XML ovojnice	JSONComplexType ŠifrantiKataster_prilogalzmenjevalnega v0.72.xlsx.
Dodatna dokumentacija	Metapodatki Spletna stran GU, Docx; xlsx	mimeType	PDF/A, Base64Binary JPEG (fotografije stavb)
Enostavnost/ Kompleksnost	<i>Ne nanaša se na to kategorijo</i>	zelo kompleksno (m:n)	ŠifrantiKataster_priloga
Obstoj enoličnega identifikatorja:	v INSPIRE Metapodatkih	EID parcele EID stavbe EID dela stavbe	<i>Json vgnезden v XML</i>
Strojna berljivost (5 Star)	4 – RDF 1– Docx	3– WFS, WMS 2- xlsx 1 – Docx, pdf, html, shp	1 - Pdf 2 - UML 3 – Json, csv

3.3.2.1.1 Izmenjava podatkov in formati

Iz tabele je razvidno, da se za *izmenjavo* podatkov uporablja *princip XML ovojnic* in *format json*, s katerim je omogočeno *razumevanje* podatkovnih struktur in *shem*.

Samodejna izmenjava podatkov med sistemi evidentiranja nepremični temelji na predpisanih obveznostih posameznih elementov za zagotavljanje *kontroliranih sprememb* in je zasnovana na podlagi *identifikatorjev informacijskih objektov*.

Interni identifikatorji (začasni, fiktivni) nastajajo ob vzpostavitvi novega informacijskega objekta (začasno, neuveljavljeno stanje), potrebni za *vzpostavitev relacij* do drugih info objektov.

Pravi (stalni) identifikatorji »Eid« ob uveljavitvi postopka sprememb nadomestijo interne, začasne identifikatorje. Torej je povezovanje med informacijskimi objekti izvedeno z enotnimi identifikatorji *Eid veljavnega stanja* nepremičninskih evidenc.

Govoreči identifikatorji so semantično bogatejši od internih in pravih identifikatorjev, saj se vežejo na pojme, ki so uveljavljeni v evidencah nepremičnin (Preglednica 41).

Preglednica 41 - Predlogi za Govoreče identifikatorje

Govoreči identifikator	Semantično obogaten opis (naš predlog)
številke točk (KO + številka točke)	številka <i>mejnika</i> v okviru katastrske občine
številke parcel - nove in poddelilke (KO + številka parcele)	številka posameznega <i>zemljišča</i> v okviru katastrske občine
številke stavb (KO + številka stavbe)	oštevilčenost <i>stavb</i> v okviru katastrske občine
številke delov stavb (KO + številka stavbe + številka dela stavbe)	oštevilčenost <i>etažne lastnine</i> (stanovanje, lokal) v okviru stavbe in katastrske občine

3.3.2.2 Dokumentacija elementov kakovosti ISK

V tem delu smo opredelili *opis izbranih elementov kakovosti podatkovnih virov* za nepremičninsko zbirko podatkov GURS, ki posledično vplivajo na njeno uporabnost pri semantičnem povezovanju, predvsem z vidika opredelitve okoliščin (konteksta). Določili smo (Preglednica 42):

- dokumentiranost informacij o *izvoru podatkovnega vira* (angl. lineage) na ravni geometrije podatkov o lastnostih objekta (atributov), o *zgodovini transformacij* podatkov in *metodologijah* *zajema*,
- kako je opisana *časovna umeščenost veljavnosti informacij* o geometriji in lastnostih objekta,
- ter o *konsistentnosti objave* evidence, ter atributov v sloju.

Preglednica 42 - Dokumentiranost elementov kakovosti katastrov nepremičnin

Elementi kakovosti: IS Kataster		
KRITERIJA	Elementi kakovosti	Pomanjkljivosti
Izvor, metodologija	Kataster_Tehnicna_dokumentacija_V14_aplikacije_Kataster 37 prilog (doc, pdf, excel) Priporočila uvajanja pravil ZKN in uvedbe IS Kataster (v. 12)	/
Natančnost geometrije	točnostDolovitveEn koordinat šifrant TOCNOSTI_DOLOCITVE_POLOZAJA. SIFRA	/
Časovna umeščenost	datumPostopka, uraPostopka, trajanjePostopka	/
Konsistentnost podatkov v tabeli	243 kontrol_IS_Kataster_seznam z_opisi	/

Vir: Priporocila_z_dne_26_11_2024-verzija_12_za_objavo.pdf

Z analizo razpoložljive dokumentacije smo ugotovili, da so z vidika izvora podatkovnega vira (lineage) in metodologiji, odlično določeni v 38 dokumentih, vključno s podatki o geometriji in lastnostih objekta. Ti dokumenti sicer niso objavljeni na spletu, kar pomeni, da niso odprti za spletno rabo.

Za novejša opazovanja pa so elementi točnosti določeni s parametrom »točnostDolocitveEn koordinat«. Točnost/natančnost geometrije je opredeljena v dokumentu

»šifrant TOCNOSTI_DOLOCITVE_POLOZAJA_SIFRA«.

Časovna umeščenost katastrskih postopkov je določena s tremi parametri, in sicer »datumPostopka«, »uraPostopka«, »trajanjePostopka«. Sicer je v bazi uporabljeno mnogo časovnih značk.

Konsistentnost podatkov v zbirki se zagotavlja z 243 različnimi kontrolnimi elementi, ki so navedeni v Seznam_kontrol_IS_Kataster_seznam_opisi, v preglednici Excel formata. Seznam kontrol v opisnem delu delno predstavlja poslovna pravila katastrov nepremičnin, delno pa zahteve informacijskega sistema za zagotavljanje konsistence podatkov v tabelah baze. Nekatera pravila so semantično bogato opisana, druga pa skromno in bi bila potrebna nadgradnje.

Iskani elementi kakovosti so opisani in z vidika semantične povezljivosti podajajo kontekst, v katerem podatke lahko uporabljamo.

V naslednjem razdelku smo analizirali obstoj semantičnih elementov ISK.

3.3.2.3 *Obstoj semantičnih elementov ISK*

V tem delu smo opredelili obstoj semantičnih elementov za izbrane sloje zbirke nepremičninskih podatkov GURS, ki omogočajo razumevanje konteksta in so ključni za semantično povezljivost podatkov. Preverili smo:

- obstoj metapodatkov,
- obstoj ključnih besed, opisov in dokumentacije na ravni evidence, posameznih slojev,
- opredelitev entitet oziroma objektov, ter pomenov besed, ki opisujejo posamezne entitete,
- povezanost v širše sprejete klasifikacije (ontologije, taksonomije, šifranti, vokabularji na nacionalnem oziroma mednarodni ravni),
- dokumentiranost konteksta uporabe, kot so uporabniški primeri, okoliščine nastanka, pravne podlage, vrste deležnikov,
- obstoj naprednih semantičnih elementov, kot so ontologije, taksonomije, grafi znanja na ravni organizacije, evidence, sloja ali tipa objektov.

S tem smo opredelili, kako dobro so objavljeni podatki pripravljeni za strojno razumevanje (

Preglednica 43).

VIR: Informacijska prenova nepremičninskih evidenc - Uporabniška navodila V 1.14 (referent) za aplikacijo Kataster

Preglednica 43 - Obstoje semantičnih elementov: Kataster nepremičnin

Obstoje semantičnih elementov: Kataster nepremičnin			
OPIS KRITERIJA	Evidenca	Podatkovni sloj/kategorija	Šifranti
Obstoje metapodatkov	Izpolnjeni INSPIRE metapodatki	/	glej servis za šifrante
Opisi, ključne besede, dokumentacija	Izpolnjeni metapodatki (INSPIRE), državni metapodatkovni sistem, Pravilnik	Podatkovni katalog; navodila za zajem	glej servis za šifrante
Opredelitev entitet	<i>Ne nanaša se na to kategorijo</i>	Opredeljene v obliki stavkov oziroma izrazov.	Opredeljene v obliki stavkov oziroma izrazov.
Povezanost v širše klasifikacije	Da, v INSPIRE Tematike	CC-SI klasifikacija vrst objektov	CC-SI klasifikacija vrst objektov
Kontekst	Del Inspire Tematike Deležniki so opisani v navodilih; Zakonodaja je dostopna	Da, podrobni opisi v dokumentaciji	
Napredni sem. Elementi	UML	UML	EID – Koda unikatno določa objekt na ravni več evidenc

Z analizo smo ugotovili, naslednje:

- **Metapodatki** za na ravni evidence obstajajo in so zapisani in objavljeni skladno z INSPIRE zahtevami (objavljeno v strojno berljivi in strukturirani obliki). To omogoča hitrejše strojno odkrivanje evidence in njeno uporabo. Na ravni posameznih slojev/kategorij in šifrantov, metapodatkov v standardizirani obliki nismo našli, obstaja pa prosto dostopno gradivo v digitalni obliki (nestrukturirana oblika, ki ni v izvorno odprtokodnem formatu) katerega je te informacije možno pridobiti oziroma ustvariti. Na ravni šifrantov, prav tako obstajajo podrobni opisi v dodatni dokumentaciji, ki je delno dostopna preko spleta, delno nam jo je na zahtevo dostavil upravljalec.
- **Opisi, ključnih besed in dokumentacija ter opredeljenost entitet**, Ključne besede so na ravni evidence, navedene in strojno dostopne preko portala INSPIRE, v metapodatkih. Na ravni slojev in šifrantov je obstoj elementov podrobno opisan v podpoglavju »Šifranti za izmenjavo podatkov in semantični opisi šifer«.
- **Povezanost v širše klasifikacije**, je na ravni evidence razviden iz umeščenost v tematske sklope INSPIRE. Kategorije oziroma sloji cest in kanalizacije, so dodatno povezane v CC_SI klasifikacijo, ki je določena z Uredbo o razvrščanju objektov³. Le ta pa temelji na mednarodni klasifikaciji Classification of types of construction⁴.
- **Kontekst**: Na ravni evidence je kontekst v obliki povzetka vira opisan v INSPIRE metapodatkih, pravna podlaga pa je na voljo na strani o zakonodaji, ki ureja delo geodetske uprave. Dokumentacija o tej evidenci zelo dobro dokumentira, kako poteka nastajanje in vzdrževanje podatkov, kar nam daje širši kontekst o deležnikih v procesu. Manj je informacij o tem kdo in na kakšne načine te podatke uporablja. Tudi zakonodaja je dostopna s povezavo na

³ Uredba o klasifikaciji objektov; (<https://pisrs.si/pregledPredpisa?id=URED8497>)

⁴ Classification of types of construction (CC); [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Classification_of_types_of_construction_\(CC\)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Classification_of_types_of_construction_(CC))

spletno stran, ki opisuje zakone in podzakonske predpise z delovnega področja Geodetske uprave. Informacija o zakonodaji je sicer na voljo, vendar dostop zahteva kar nekaj klikov stran ter iskanje v celotni pravni podlagi za GU. Na boljši način je zakonodaja opisana na straneh OPSI⁵, kjer so specificirane vse podlage, ki se nanašajo na to evidenco.

- **Napredni semantični elementi:** V okviru dokumentacije evidence zbirke Kataster nepremičnin, za evidenco, njene sloje ali šifrante, je stanje glede naprednih semantičnih elementov naslednje:
 - **Ontologije, taksonomije** –objekti IS Kataster so umeščeni v CC_SI klasifikacijo, ki je neke vrste taksonomija. Ontologij za posamezne kategorije znotraj GJI nismo našli.
 - **Podatkovni modeli** – podatkovni modeli javno niso objavljeni, jih je pa možno pridobiti s strani upravljavca zbirke, a to ni nikjer eksplicitno navedeno. Ti modeli so na voljo v obliki diagramov UML, v digitalni obliki, v ne odprtokodnem formatu, so strukturirani in jih je možno javno objaviti.
 - **Principi Linked data** – vsi objekti imajo določeno enotno ID številko na ravni države, podatki pa se prevzemajo v Json in GeoJson obliki. To so dobri pogoji za povezovanje po principu Linked data.
 - **Semantične anotacija** – Podatki nimajo semantičnih anotacij na strojni ravni, so pa v dostopni dokumentaciji dovolj dobro opisano, da bi lahko z obdelavo podatkov to spremenili v semantične anotacije.

Če povzamemo, so podatki na ravni evidence dobro standardizirani in strojno berljivi, vendar na ravni posameznih slojev in šifrantov obstajajo vrzeli, predvsem glede uporabe naprednih semantičnih elementov. Gradivo sicer omogoča nadaljnji razvoj in nadgradnjo v avtomatizirano semantično povezljivost. Dodatno ga je možno obogatiti z opisi iz strokovne literature, navodil za izvajalce vzdrževanja na lokalni ravni in dokumentacije ostalih deležnikov.

3.3.2.3.1 Šifranti za izmenjavo podatkov in semantični opisi šifer

Za izmenjavo podatkov v ISK je oblikovanih 66 šifrantov. V semantični analizi tega projekta smo jih uvrstili v 17 pomensko raznolikih skupin,

- o *deležnikih* (stranke, lastniki, vlagatelji, upravniki, upravljavci, pooblaščen podjetja),
- o *stvarnih objektih* (zemljiščih, stavbah in delih stavb-stanovanjskih in poslovnih, prostorih),
- o *stvarnih nepremičninskih identifikatorjih* (označitve mejnih točk – mejnikov; stavb - hišne identifikacijske tablice naslova),
- o *pravicah* (stvarnopravnih, planerskih-namenskih rab, gradbenih),
- o *postopkih* in pod postopki (sestavine, zahteve, predlogi, vpisi),
- o *dokumentaciji* (priloge),
- o *formatih* (MIME),

⁵ OPSI Opis GJI (<https://podatki.gov.si/dataset/zbirni-kataster-gospodarske-javne-infrastrukture>)

- o številčenju / *indeksaciji* (EID objekti),
- o *prostorskih entitetah* na različnih hierarhičnih ravneh (državna, občinska, davčno-katastrska občina),
- o geometričnih entitetah na več ravneh (točka, daljica, poligon, poligon dela stavbe),
- o *statusih* (upravni, status geometričnih entitet – npr. mejnih točk, status vpisa, status deležnika, meritve),
- o *geometričnih* lastnostih (geodetski datum, površina, tloris, točka povezave, lastnosti reliefa),
- o načinih *izmere in izračuna* (metode določitev višine, položaja, površin, model transformacije),
- o *kakovosti* podatkov (točnost določitve položaja, višin, površine),
- o *okolščinah* (časovnih, zakonskih, dejanske rabe prostora),
- o *fizikalnih* lastnostih (geografskih regij – klima; zemljišč – pedoloških – globina tal; geoloških sistematskih enotah, podlagah; stavb – nosilne konstrukcije),
- o *virih* podatkov (GURS, CRP ZKNJ, MOP, Občine, ...),

Šifranti so organizirani v tabele, ki vsebujejo šifre in nazive, v nekaterih redkih primerih pa tudi pomenske opise nazivov. Pri zasnovi sistema šifriranja v ISK so sodelovali strokovnjaki s področja evidentiranja nepremičnin in običajno je, da so njim poznani pomeni eno- ali dvo- besednih terminov, uporabljenih za označitev konceptov v sistemu ISK. To ne pomeni, da so vsi izrazi razumljivi tudi strokovnjakom, ki niso sodelovali pri izgradnji sistema ali strojem.

Za nehumano, strojno samodejno semantično povezovanje podatkov pa ocenjujemo, da bi bilo treba izboljšati stopnjo doslednosti pri izvedbi opisov (z obogatitvijo, navedbo pozitivnih in negativnih primerov, opisanih z drugimi, v posameznem nazivu šifre še neuporabljenimi besedami tam, kjer so prekratki ali neobstoječi).

3.3.2.3.2 Analiza semantične povednosti in bogatosti šifrantov

Semantično bogatejši opisi nazivov vsebujejo ustrezno število besed v naslednjih šifrantih:

Bogati opisi

Lastnosti podnebja (klime)
 UPRAVNI_STATUSI_TOCK
 LASTNOSTI_RELIEFA
 LASTNOSTI_KLIME

Srednje bogati

KP_VRSTE_PRILOG
 STATUSI_UPRAVNIKOV
 VISINSKE_TOCNOSTI_STAVBE
 VISINSKI_DATUMI_STAVBE
 POLOZAJNE_TOCNOSTI_STAVB
 VRSTE_POSTOPKOV_BONITET
 GLOBINA_TAL
 GEOLOSKA_PODLAGA
 TEKSTURA
 SISTEMATSKA_ENOTA

TOCNOSTI_DOLOCITVE_VISINE

METODE_DOLOCITVE_POVRSINE

Skromni opisi

KP_KAT_POS_X_PODPOSTOPKI

KP_KAT_POS_X_PRILOGE (atribut obveznosti)

KP_KAT_POS_X_SESTAVINE

MODEL_TR (manjkata dva ključa opisa šifer, delni opis v nazivu)

METODE_DOLOCITVE_VISINE (delni opis v nazivu)

METODE_DOLOCITVE_POLOZAJA (manjka opis za 9 šifer)

GEO_DATUMI_EN

DATUMI_H

Slabo opisano/ni opisa

Tipi vlagateljev

Tipi strank

Dodatne zakonske podlage

TIPI_TOCK (delni opis v nazivu)

TIPI_DALJIC (delni opis v nazivu)

MIME_TYPE

EID_OBJEKTI (v polju Opis navedba imena tabele)

NACINI_DOLOCITVE_POVRSIN_DELA_S

VRSTE_SESTAVIN_DELOV_STAVB

TIPI_LASTNIKOV

VRSTE_UPRAVLJAVCEV

VRSTE_STVARNO_PRAVNIH_PRAVIC

VRSTE_STAVBA_PARCELA

VRSTE_PROSTOROV (delni opis v nazivu)

VRSTE_NAMENSKE_RABE (delni opis v nazivu)

VRSTE_GRADBENIH_PARCEL

VRSTE_DEJANSKIH_RAB_DEL_ST (delni opis v nazivu)

VRSTE_DEJANSKE_RABE (delni opis v nazivu)

VRSTE_STANOVANJ

VIRI_RK (V opisu ni niti razlag kratic v nazivu)

VIRI_PRIDOBITVE_PODATKOV (V opisu ni niti razlag kratic v nazivu)

VIRI_OZ (V opisu ni niti razlag kratic v nazivu)

VIRI_NAMENSKIH_RAB (V opisu ni niti razlag kratic v nazivu)

VIRI_DEJANSKIH_RAB (V opisu ni niti razlag kratic v nazivu)

UPRAVNI_STATUSI

TIPI_TABLIC (ni pojasnila za splošno besedo »tablica«; delni opis v nazivu)

TIPI_STAVB

TIPI_POOBLAŠČENEGA_PODJETA

TIPI_POLIGONOV_DELOV_STAVB

TIPI_LASTNISTVA

STATUSI_VPISA_STAVBE (V opisu ni niti razlag kratic v nazivu)

STATUSI_VPISA_DELOV_STAVBE (V opisu ni niti razlag kratic v nazivu)

STATUSI_UPRAVLJAVCEV

STATUS_MERITVE (delni opis v nazivu)

NOSILNE_KONSTRUKCIJE (delni opis v nazivu)

TOCNOSTI_DOLOCITVE_POLOZAJA (opis v nazivu)

NACINI_OZNACITVE_TOCK (trije opisi od 12 šifer, ostalo delni opis v nazivu)

KATASTRSKE_OBCINE (lahko bi v opis dodali vsaj politično občino kateri pripada)

DA_NE

Opis v nazivu

KP vrste kat postopkov

KP vrste kat pod postopkov

VRSTE_UPRAVLJAVCEV

Iz zgornjega pregleda je razvidno, da iz semantičnega vidika, v ISK prevladuje kategorija »Slabo opisano/ni opisano«. Ugotavljamo, da so v kategorijo *Bogati opisi* uvrščeni mnogi šifranti gostujočih podatkov.

Semantično bogatejši so tudi opisi nazivov samodejnih, logičnih kontrol v preglednici »Seznam_kontrol_IS_Kataster_z_opisi«, v kateri so imena in opisi 243 kontrolnih mehanizmov ISKataster.

S tehničnega vidika je samodejno semantično povezovanje možno, a v večini primerov ob predhodni pripravi podatkov.

Če povzamemo, so podatki na ravni evidence dobro standardizirani in strojno berljivi, vendar na ravni posameznih slojev in šifrantov obstajajo vrzeli, predvsem glede uporabe naprednih semantičnih elementov. Gradivo sicer omogoča nadaljnji razvoj in nadgradnjo v avtomatizirano semantično povezljivost. Dodatno ga je možno obogatiti iz strokovnih navodil za izvajalce vzdrževanja in ostale dokumentacije, ki je tudi dostopna na portalu eProstor.

V naslednjem odstavku smo, glede na razvite mere opredelili dostopnost do semantičnih elementov, s čimer smo dobili informacijo, kako dobre pogoje imajo podatki za semantično povezovanje z vidika skladnosti in dostopnosti.

3.3.3 Mere dostopnost semantičnih elementov

V tem delu smo ocenili dostopnost in verodostojnost semantičnih elementov za izbrane sloje zbirke nepremičninskih podatkov GURS, ki z vidika dostopnosti in semantične celovitosti in skladnosti, vplivajo na semantično povezljivost zbirk. Ocenili jih smo z izbranimi merami, kot so tehnična (sintaktična) dostopnost, organizacijska dostopnost, stopnja uradnosti in semantična bogatost vira. V sklopu posamezne zbirke smo primarno upoštevali vse vire, ki so bili z visoko stopnjo odprtosti pridobljeni na javnih spletnih straneh upravljavca evidence; drugotno smo pozvali skrbnike evidence, da nam priskrbijo interno dokumentacijo (Preglednica 44).

Preglednica 44 - Analiza dostopnosti semantičnih elementov za ISKataster

IS Kataster	Evidenca	Točke	Šifranti	Točke
Tehnična dostopnost vira pomena (Sintaktična dostopnost)	Spletno dostopni standardizirani servisi	6	Lokalni Standardizirani strukturirani viri v digitalni obliki (spletne strani)	5
Organizacijska dostopnost vira	Splošno mednarodno dostopna in sprejeta vsebina	5	Opisi so del dokumentacije evidence v vsebinski domeni (strokovna javnost)	3
Stopnja uradnosti	Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti..., Pravna praksa (precedenčne sodbe))	6	Opisi so del strokovnih specifikacij in posebnih vokabularjev	4
Semantična bogatost vira	odstavek	3	podroben opis s primeri, skicami ali slikovnim gradivom	5
Starost vira	Trenutno veljavno	5	Trenutno veljavno	5
TOČK		25		22
DELEŽ(%)		86		76

Na ravni zbirke nepremičninskih podatkov imamo strojno razumljive vire, ki so objavljeni tako na slovenskem portalu INSPIRE kot tudi na mednarodnih portalih. Omogočen je strojni dostop do metapodatkov in vokabularjev v obliki RDF, kar je že javno objavljen vir, v strukturirani obliki, v odprtokodnem formatu.

Vsebina je javno dostopna in usklajena z mednarodnimi standardi zapisa (INSPIRE), temelji pa na pravnih podlagah in strokovnih specifikacijah, kar nakazuje na visoko mero skladnosti zakonodajo in stroko, kar omogoča povezovanje. Semantično je vsebina na ravni evidence opisana skladno s standardi INSPIRE, na ravni šifrantov, pa vsi izrazi in entitete niso opisani v strojno razumljivem načinu, vsebinsko pa so dovolj bogati, da omogočajo širše razumevanje konteksta.

Zato je ocena opisa evidence dosegla 86 točk od 100, kar že predstavlja visoko osnovo za semantično povezljivost. Na ravni posamičnih slojev in šifrantov so podatki semantično bolj opisani, a v strojno manj razumljivi obliki in so dosegli 76 točk od 100.

V naslednjem poglavju smo obravnavane podatke umestili v tej nalogi razviti semantični referenčni okvir UOPGR, kar nam bo v nadaljevanju omogočalo oceniti možnosti povezovanja tematsko podobnih evidenc glede na njihovo tematsko prekrivanje oziroma semantično skladnost.

3.3.4 Umestitev šifrantov in njihovih vrednosti v UOPGR taksonomijo

V tem delu smo izbrane podatke iz Katastra nepremičnin - STAVBE in vrednosti v njihovih šifrantov (VRSTE_STANOVANJ, VRSTE_SESTAVIN_DELOV_STAVB, VRSTE_DEJANSKIH_RAB_DEL_ST) umestili v razviti semantični referenčni sistem UOPGR, glede na njihovo tematiko in semantično globino (Preglednica 45).

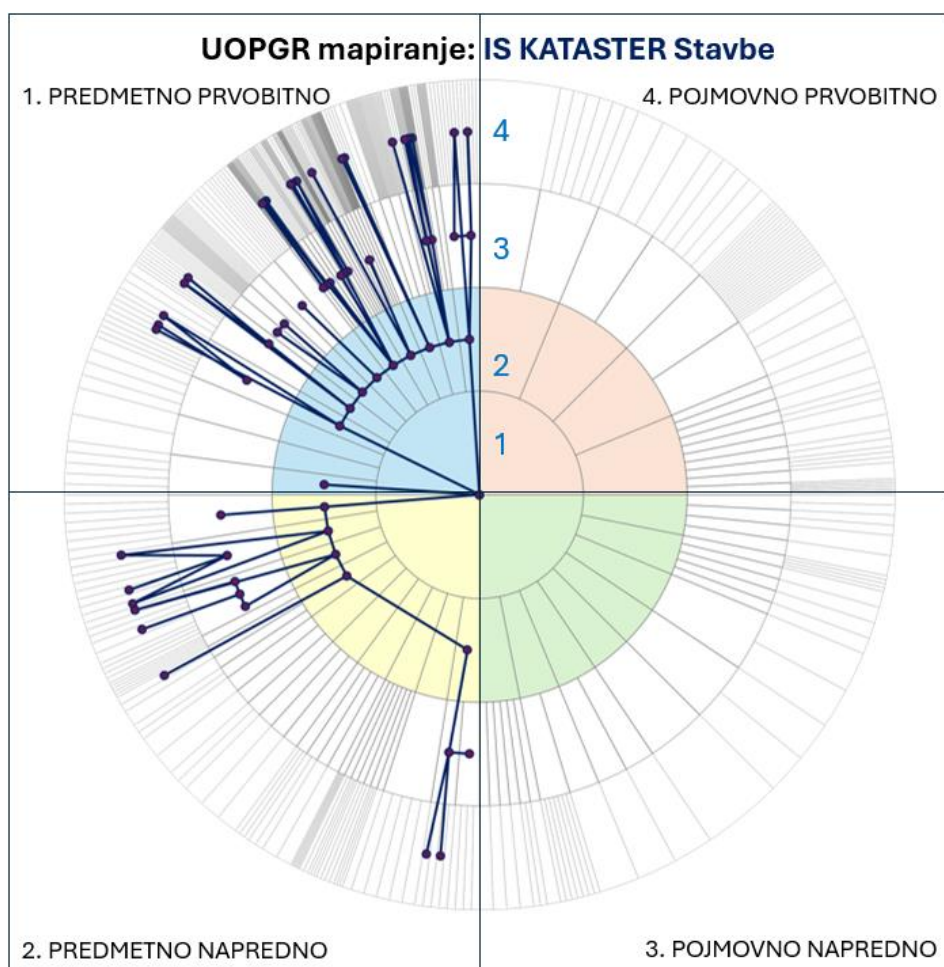
Preglednica 45 - Umestitev šifrantov katastra nepremičnin (Stavbe) v UOPGR in semantična globina

UOPGR	UOPGR OPIS KATEGORIJE	Sem. Globina	RABA_ID	RABA_OPIS
101	Objekti aktivnosti BIVANJA	2	33	klet
101	Objekti aktivnosti BIVANJA	2	34	shramba, sušilnica, pralnica
101	Objekti aktivnosti BIVANJA	2	35	terasa, balkon, loža
101	Objekti aktivnosti BIVANJA	2	36	tehnični prostor
101	Objekti aktivnosti BIVANJA	2	38	skupne sanitarije
101	Objekti aktivnosti BIVANJA	2	39	ruševina
10101	Vsakodnevno bivanje	3	4	bivalna enota
1010104	stanovanje	4	1	Stanovanje
1010104	stanovanje	4	1	stanovanje v enostanovanjski stavbi
1010104	stanovanje	4	2	stanovanje
1010104	stanovanje	4	3	oskrbovano stanovanje
1010104	stanovanje	4	47	stanovanje v dvostanovanjski stavbi
10102	Nevsakodnevno in sezonsko bivanje	3	5	koča, dom
10102	Nevsakodnevno in sezonsko bivanje	3	55	zavetišče in hotel za živali
10102	Nevsakodnevno in sezonsko bivanje	3	6	nastanitveni gostinski obrat
1010201	vikend	4	60	zidanica
1020105	silos	4	20	del stavbe z rezervoarjem, silos
10203	Zemljišca/objekti za pridobivanje rastlinskih surovin	3	60	zidanica
1020308	čebelnjak	4	58	čebelnjak
1020314	rastlinjak	4	45	rastlinjak
10204	Zemljišce/objekt za pridobivanje živalskih surovin	3	29	del stavbe za spravilo pridelka
10204	Zemljišce/objekt za pridobivanje živalskih surovin	3	28	pomožni kmetijski del stavbe
1020402	hlev	4	27	hlev
1020405	silos	4	20	del stavbe z rezervoarjem, silos
1020601	elektrarna	4	19	elektrarna
103	Objekti aktivnosti PROIZVODNJE izdelkov (industrijsko/obratno)	2	17	del stavbe za industrijsko rabo
103	Objekti aktivnosti PROIZVODNJE izdelkov (industrijsko/obratno)	2	18	del stavbe za težko industrijo
1030403	skladišče	4	21	skladišče
1030404	rezervoar	4	20	del stavbe z rezervoarjem, silos

10404	kolesni transport	3	11	bencinski servis
1040703	hangar	4	54	hangar, remiza, tovorni terminal
105	Objekti aktivnosti IZMENJAVE proizvedenih IZDELKOV in urbanih STORITEV	2	2	Poslovni prostor
105	Objekti aktivnosti IZMENJAVE proizvedenih IZDELKOV in urbanih STORITEV	2	12	del stavbe za opravljanje storitev
105	Objekti aktivnosti IZMENJAVE proizvedenih IZDELKOV in urbanih STORITEV	2	9	poslovni del stavbe
1050101	skladisca	4	21	skladišče
10502	Trgovina na drobno	3	10	trgovski del stavbe
10502	Trgovina na drobno	3	52	trgovski del stavbe na bencinskih servisih
1050201	lokal	4	11	bencinski servis
1050202	skladisce	4	21	skladišče
1050204	gostisce	4	7	prehrambni gostinski obrat
10503	monetarne storitve	3	49	banka, pošta, zavarovalnica
10506	storitve BREŽIČNE (mobilne) komunikacije	3	13	del stavbe za izvajanje elektronskih komunikacij
1050704	posta	4	49	banka, pošta, zavarovalnica
1050706	dvorana	4	50	konferenčna ali kongresna dvorana
1050709	sejmisce	4	42	sejemska dvorana, razstavišče
10508	sprostitev pod streho	3	22	del stavbe za kulturo in razvedrilo
10508	sprostitev pod streho	3	26	športna dvorana
10508	sprostitev pod streho	3	53	igralnica, cirkus, plesna dvorana, diskoteka
1050802	dvorana	4	42	sejemska dvorana, razstavišče
10602	pokopalisca	3	31	del stavbe za pokopališko dejavnost
10702	Ozelenjene površine za sprostitev	3	56	del stavbe za živali ali rastline v živalskem ali botaničnem vrtu
10703	Oprema na površinah za sprostitev	3	59	kolesarnica, čolnarna
10801	Objekti aktivnosti kopenske MANIPULACIJE NA PROSTEM	3	54	hangar, remiza, tovorni terminal
1080102	postaja	4	14	del stavbe za potniški promet
1080104	lopa	4	57	drvarnica
1090001	garaza	4	15	garaža
1090003	parkirisce	4	2	Parkirno mesto
1090004	hisa	4	16	garažno parkirno mesto
112	Objekti aktivnosti VERSKIH OBREDOV	2	30	del stavbe za opravljanje verskih obredov
112	Objekti aktivnosti VERSKIH OBREDOV	2	46	del stavbe za pastoralno dejavnost
201	Objekti aktivnosti JAVNEGA UPRAVLJANJA (državno/lokalno) socialne storitve, združenja	2	51	arhiv
20100		3	8	poslovni prostor javne uprave
202	Objekti aktivnosti IZOBRAŽEVANJA	2	24	del stavbe za izobraževanje in znanstveno raziskovalno delo
2020001	šola	4	43	šola, vrtec

2020006	knjiznica	4	23	muzej in knjižnica
2020008	vrtec	4	43	šola, vrtec
203	Objekti aktivnosti ZDRAVSTVENE OSKRBE	2	25	del stavbe za zdravstveno oskrbo
2030101	dom	4	44	bolnica, zdravstveni dom
2030201	bolnica	4	44	bolnica, zdravstveni dom
20303	Objekti PRIMARNE ravni aktivnosti ZDRAVSTVENE OSKRBE	3	48	zdravilišče
2040201	zaklonisce	4	32	del stavbe za zaščito, reševanje in zaklonišče
21001	Kulturna dediščina	3	51	arhiv
2100101	Muzej	4	23	muzej in knjižnica
2100103	Ruševine	4	39	ruševina
21002	Objekti za kulturno udejstvovanje	3	22	del stavbe za kulturo in razvedrilo
x	#N/A		37	skupni komunikacijski prostor
x	#N/A		40	nestavba
x	#N/A		41	del stavbe za posebne namene

V Slika 19 so koncepti za stavbe kartirani v semantični referenčni sistem UOPGR.



Slika 19 - Graf umeščenosti vrednosti šifrantov sloja Stavbe - Kataster nepremičnin v UOPGR

Opomba: Trenutni operativni model UOPGR se deli na 4 kategorije, ki se skupno delijo na 34 razredov, ki se delijo skupno na 105 podrazredov in obsegajo 446 entitet (objektov, shem ali pojavov) v četrti globini. Skupno 590 konceptov.

V podatkovnem sloju DTK Stavbe, smo od 94 opisov identificirali 36 opisov iz šifrantov, ki ustrezajo entitetam na 4. semantični globini UOPGR. Zaradi presplošnega opisa smo uvrstili 23 opisov v šifrantih v podrazrede na 3. globini in 32 opisov v razred na 2. globini. Treh vrednosti atributov, ki določajo vrste rabe objektov, nismo mogli umestiti v UOPGR (glej Slika 19).

Umeščanje smo izvedli glede na trenutno raven razvoja referenčnega sistema in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov bodo te povezave lahko bolj številčne in bolj bogate in s tem prispevale k več možnostim povezovanja.

Na grafu (Slika 19) je razvidno, da se večina opisov iz šifrantov umešča med Predmetne objekte Prvobitnih človeških aktivnosti (1. Kategorija) in Predmetne objekte Naprednih človeških aktivnosti (2. Kategorija). V 3. in 4. kategoriji referenčnega sistema UOPGR nismo umestili nobenega koncepta.

Veliko število konceptov s semantično globino 2 ali 3 nakazuje, da je na teh področjih prostor za nadgradnjo referenčnega modela v globino.

3.4 Rezultati naloga 3.3: Testna analiza možnosti semantične povezljivosti GJI zbirk podatkov

V tem razdelku analiziramo semantiko izbranih evidenc gospodarske javne strukture (občinska in državna raven). Med raziskavo raziskave možnih virov podatkov za analizo, smo z več občin dobili informacijo, da nam podatkov iz njihovih sistemov ne morejo v popolnosti predstaviti. Zato smo uporabili vire v obliki dokumentacije, ki so jih lahko delili s strokovno javnostjo, ter izvedli intervju za pridobitev dodatnega konteksta.

3.4.1 Opredelitev podatkov za analizo zbirk GJI

Glede na možnosti kasnejše primerjave semantičnega povezovanja, smo v skupini podatkov iz zbirnega katastra javne gospodarske infrastrukture, za analizo semantične povezljivosti izbrali podatkovne sloje/kategorije, navedene v Preglednica 46.

Preglednica 46 - Opredelitev izbranih slojev/kategorij Zbirnega katastra Gospodarske javne infrastrukture (GJI) za analizo semantične povezljivosti

Ime dokumenta	Opis	Vir
GJI Geodetska Uprava		
20230724_Izmenjevalni_format_ZKGJI_v1.5.1.pdf	Dokumentacija o izmenjevalnem formatih za vse zbirke v GJI, vključno z opisi šifrantov in	https://www.e-prostor.gov.si/fileadmin/Prodaja/GJI/Postopki/Elaborati_GJI/20230724_Izmenjevalni_format_ZKGJI_v1.5.1.pdf
KGI_SLO_GJI_KANALIZACIJA_linije_20240602.gpkg	Kanalizacija za območje Slovenije - Linijski podatkovni sloj zbirnega katastra javne gospodarske infrastrukture	https://ipi.eprostor.gov.si/jgp/data
KGI_SLO_GJI_KANALIZACIJA_tocke_20240602.gpkg	Kanalizacija za območje Slovenije – Točkovni podatkovni sloj zbirnega katastra javne gospodarske infrastrukture	https://ipi.eprostor.gov.si/jgp/data
KGI_SLO_GJI_CESTE_linije_20241012.gpkg	Ceste za območje Slovenije - Linijski podatkovni sloj zbirnega katastra javne gospodarske infrastrukture	https://ipi.eprostor.gov.si/jgp/data
Komunala Novo Mesto		
Kanalizacija – Kanal	Podatkovni model linijskih objektov kanalizacije v upravljanju komunale Novo Mesto.	Opis podatkovnega modela – Vir Komunala NM
Kanalizacija – Točke	Podatkovni model točkovnih objektov kanalizacije v upravljanju komunale Novo Mesto.	Opis podatkovnega modela – Vir Komunala NM

Iz podjetja Komunala Novo mesto, smo pridobili dokumentacijo za sloj oziroma kategorije vodovodnega in kanalizacijskega omrežja. Za analizo smo izbrali kanalizacijsko omrežje, in v GJI izbrali primerljive koncepte, da smo lahko opravili primerjavo.

Sloj Ceste smo iz GJI izbrali zato, da ga bomo kasneje lahko opravili analizo semantične povezljivosti s podobnimi tematskimi sloji/kategorijami

V naslednjem poglavju smo analizirali obstoj elementov, ki predstavljajo pogoj za semantično povezovanje. To so informacije o tehnični oziroma sintaktični povezljivosti, obstoj izbranih elementov kakovosti in obstoj semantičnih elementov.

3.4.2 Ocena pogojev za semantično povezovanje

Da bi podatke med seboj lahko semantično povezovali, je treba poznati tudi ostale tehnične in vsebinske omejitve, ki se ne nanašajo samo na razumevanje konteksta. Zato smo pred samo analizo povezljivosti semantičnih elementov, opravili pregled tehničnih oziroma sintaktičnih lastnosti, ki omogočajo povezovanje, obstoj informacij o izbranih elementih kakovosti, ter pregled obstoja semantičnih elementov.

3.4.2.1 Sintaktična povezljivost

V spodnji tabeli smo opredelili tehnične oziroma sintaktične elemente za izbrane sloje oz kategorije GJI in evidence kanalizacije Komunale Novo Mesto, ki posledično vplivajo na njihovo semantično povezljivost. Določili smo:

- formate zapisa geometrije in pripadajoče dokumentacije,
- ocenili razčlenjenost zbirke na eno ali več tabel, kar predstavlja enostavnost oziroma dodatno kompleksnost pri semantičnem povezovanju,
- ali za posamezni zapis objekta obstaja enolični identifikator in na kakšni ravni (lokalno, državno, globalno),
- na kakšnem ravni je vsebina, glede na 5 zvezdni kriterij strojne berljivosti podatkov.

S tem smo opredelili, kako dobro so objavljeni podatki pripravljeni za strojno razumevanje. V primeru podatkov v zbirki GJI, za vse sloje veljajo pretežno enake lastnosti, zato vrednosti v tabeli veljajo za vse tri sloje. Če velja samo za posameznega, je to navedeno.

Preglednica 47 - Analiza Sintaktične povezljivosti virov v evidenci GJI

Sintaktična povezljivost virov evidence GJI			
KRITERIJ	Opis Evidence GJI	Podatkovni sloj/kategorija	Šifranti, Atributi
Format zapisa geometričnega sloja in atributov	<i>Ne nanaša se na to kategorijo</i>	Geopackage – gpkg WMS, WFS, Atom	-Json, - csv
Dodatna dokumentacija	INSPIREMetapodatki Spletna stran GU docx	Pdf Docx Geojson	Pdf
Enostavnost/ Kompleksnost	<i>Ne nanaša se na to kategorijo</i>	Kompleksno (n:m)	Kompleksno (n:m)
Obstoj enoličnega identifikatorja:	V INSPIRE Metapodatkih	FEATUREID – Lokalno EID_TOCKA; EID_LINIJA - Nacionalno	<i>Ne nanaša se na to kategorijo</i>
Strojna berljivost (5 Star)	4 – RDF 1– Docx,	3– gpkg, geojson, WFS, WMS 2- xlsx 1 – Docx, pdf, html	3 – Json, csv 1 - Pdf

Pri podatkih evidence GJI so podatki na ravni evidence tehnološko dostopni po standardu za INSPIRE, kar pomeni, da imamo večino informacij na voljo v naprednejši strojno berljivi obliki. Tudi sami podatki so sintaktično na podobni ravni.

Dodatna dokumentacija je na voljo v strojno berljivi obliki, pretežno na osnovnem ravni (word, pdf). Izstopata le metapodatkovni opis evidence, ki ga je možno izvoziti v obliki RDF in struktura in atributnih tabel, ki je na voljo v MS Excel formatu.

Sama evidenca je sestavljena zelo kompleksno, kar pomeni, da je treba geometrični sloj povezati z zunanjimi šifranti. Na spletni strani za prenos ključ za povezovanje zunanjih šifrantov ni jasno opredeljen.

Vsi obravnavani sloji imajo za objekte v tabeli unikatne ID-je na ravni sloja, Evidence GJI, kot tudi na ravni spletnega servisa.

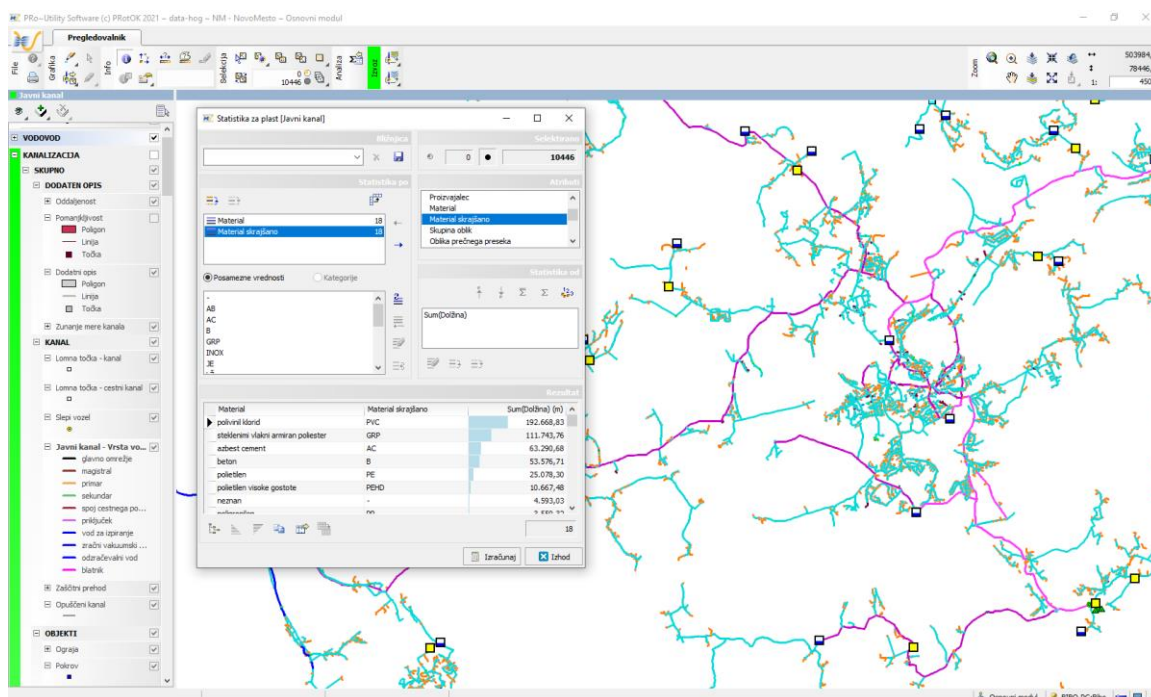
S tehničnega vidika je semantično povezovanje dokumentacije možno ob predhodni pripravi podatkov. Z vsebinskega vidika vidimo izzive, ki smo jih analizirali v in opredelili v naslednjih poglavjih.

Preglednica 48 -Sintaktična povezljivost virov evidence kanalizacije Komunale Novo Mesto

Sintaktična povezljivost virov evidence kanalizacije Komunale Novo Mesto			
KRITERIJ	Opis Evidence	Podatkovni sloj/kategorija	Šifranti, Atributi
Format zapisa geometričnega sloja in atributov	<i>Ne nanaša se na to kategorijo</i>	Lastni format aplikacije Protok	docx
Dodatna dokumentacija	Intervju s strokovnjakom	Intervju s strokovnjakom	Intervju s strokovnjakom
Enostavnost/Kompleksnost	<i>Ne nanaša se na to kategorijo</i>	Kompleksno (1:n)	<i>Ne nanaša se na to kategorijo</i>
Obstoj enoličnega identifikatorja:	/	Na ravni podatkovnega modela aplikacije Protok	<i>Ne nanaša se na to kategorijo</i>
Strojna berljivost (5 Star)	1 – Html spletna stran 0 -Vsebina intervjuja	1 – Lastni podatkovni model	1- docx

Na podlagi pridobljene dokumentacije iz podjetja Komunala Novo Mesto, smo ugotovili, da se podatki obdelujejo v digitalni obliki, a v relativno zaprtem sistemu, kar pomeni, da metapodatkov na ravni Evidence ni ali podatki niso javno dostopni. Na voljo je samo pravna podlaga v okviru pravilnikov občine Novo mesto ter izpis šifrantov za sloje kanalizacije. Dodatne informacije smo pridobili z intervjujem s strokovnjakom za omenjeno področje.

Na ravni podatkovnega sloja in šifrantov znotraj aplikacije obstajajo interni enolični identifikatorji, kar je na spodnji sliki, delo razvidno tudi iz zaslonskih slik sistema Komunale Novo Mesto.



Slika 20 - Zajem zaslona aplikacije Protok s podatki in oknom z atributi za kanalizacijsko omrežje občine Novo Mesto

Podatki podjetja Komunale Novo mesto imajo tehnološko potencialno možnost za strojno povezovanje, a so v zaprtem sistemu, kar pomeni, da bi morali biti dodatno obdelani (Glej Slika 20). S semantičnega vidika, so glede na pričevanje strokovnjaka, delno opisani v zakonodaji (Odlok o odvajanju in čiščenju komunalne in padavinske odpadne vode na območju Mestne občine Novo mesto⁶), ki je javno dostopna, a ne v strukturirani obliki.

V nadaljevanju smo analizirali obstoj elementov kakovosti, ki so ključnega pomena za semantično povezovanje. Glede na pomanjkanje gradiva na ravni podatkov Komunale Novo mesto smo analizirali le gradivo evidence GJI.

3.4.2.2 Dokumentacija elementov kakovosti

V tem delu smo opredelili opis izbranih elementov kakovosti podatkovnih virov za vira za nepremičninsko zbirko podatkov GURS, ki posledično vplivajo na njeno uporabnost pri semantičnem povezovanju, predvsem z vidika opredelitve konteksta. Glede na pomanjkanje podatkov in dokumentacije za podatke podjetja Komunala Novo mesto, smo analizo za zadnje izpustili. Določili smo:

- obstoj informacij o izvoru podatkovnega vira (angl. lineage) na ravni geometrije podatkov o lastnostih objekta (atributov), o zgodovini transformacij podatkov in metodologijah zajema,
- kako je opisana časovna umeščenost veljavnosti informacij o geometriji in lastnostih objekta,
- ter o konsistentnosti objave evidence, ter atributov v sloju.

⁶ https://www.novomesto.si/dolenjski-uradni-list/dul-objave/2017101116001414/odlok_o_odvajanju_in_ciscenju_komunalne_in_padavinske_odpadne_vode_na_obmocju_mestne_obcine_novo_mesto/

Preglednica 49 - Analiza obstoja elementov Kakovosti v GJI Slojih

Elementi kakovosti: GJI		
KRITERIJA	Elementi kakovosti	Pomanjkljivosti
Izvor, metodologija	- Opredeljeni postopki priprave in sprememb - Pravna podlaga za GJI Atributi: - VIR - tipSpremembe - xxx_izvajalci.json	/
Natančnost geometrije	Določeno v atributih: - Točnost določitve položaja - Točnost določitve višine	/
Časovna umeščenost	Evidenca sprememb šifrantov Atribut [DATUM_VIRA] in [LETO_IZGRADNJE]	/
Konsistentnost podatkov	- Standardizirani formati - obvezni atributi, - enotni CRS - Unikatni ID - Sledenje sprememb - kontrole	/

Z analizo smo ugotovili, da so z vidika opisa izvora in metodologije izdelave, na voljo zelo podrobni podatki, ki so na ravni evidence na voljo v metapodatkovnem opisu INSPIRE, na podrobnejši ravni slojev oziroma kategorij, pa so na ravni atributov podatkovnega modela in spremljajoče dokumentacije zelo podrobno opredeljeni ključni elementi, ki določajo izvor. Atribut [VIR] določa metodologijo zajema, atribut [tipSpremembe] pa omogoča sledenje spremembam na objektih. Del dokumentacije na ravni GU, je tudi povezana datoteka XXX_udelezenci.json, ki poleg metodologije, z opisom deležnikov daje širši kontekst nastanka podatka.

Z vidika deklarirane geometrijske natančnosti, ima vsak objekt atribut, [NATANCNOST_DOL_POLOZAJA] in [NATANCNOST_DOLOCITVE_VISIN], ki opredeljujeta natančnost geometrije.

Časovna umeščenost je določena z atributom [DATUM_VIRA], ki določa, kdaj je bil podatek zajet, ter [LETO_GRADNJE], ki opredeljuje leto nastanka objekta. Na ravni elaborata se vodi tudi informacija o spremembah na objektu, ki je v prosto podatkovnem modelu prosto dostopnega sloja ni, kar omogoča pregled zgodovine sprememb na objektu.

Z vidika konsistentnosti in celovitosti podatkov, je Zbirnem katastru GJI to zagotovljeno z več mehanizmi, ki se osredotočajo na podatkovne formate, obvezna polja in postopke predložitve sprememb.

Standardizirani podatkovni formati: Podatki o lokaciji se oddajajo v formatu GeoJSON, atributni podatki pa v formatu JSON, kar zagotavlja enotnost in preprečuje neskladnosti.

Obvezni atributi: Vsak objekt GJI mora vsebovati 20 osnovnih in 5 posebnih atributov. Vsi predpisani atributi so obvezni, manjkajoče vrednosti pa se označijo z "null" (nešifrirana polja) ali "0" (šifrirana polja).

Koordinatni sistem: Lokacije objektov se evidentirajo v sistemu D96/TM (EPSG: 3794) s tremi koordinatami (e, n, H) na dve decimalni mesti, kar zagotavlja prostorsko natančnost.

Edinstveni identifikatorji: Ob vnosu novega objekta ('D') se začasno dodeli negativna vrednost, nato GU določi trajni identifikator s specifično predpono za posamezno plast (npr. 3001 za točke).

Sledenje spremembam: Vsaka sprememba se evidentira z atributom 'tipSpremembe', ki opredeljuje dodajanje ('D'), spremembo ('S') ali brisanje ('B').

Zahteve za oddajo podatkov: Oddaja podatkov zahteva hkratno vključitev lokacijskih in atributnih podatkov za vse vrste sprememb (D, S, B).

Rok za posodobitve: Spremembe morajo biti evidentirane najpozneje v treh mesecih po nastanku (novogradnje, rekonstrukcije, rušitve, spremembe atributov ali namembnosti).

Poimenovanje datotek: Izmenjevalne datoteke morajo biti v arhivu ZIP, brez šumnikov, presledkov in posebnih znakov, s kodiranjem v UTF-8.

Validacija podatkov: Sistem preverja skladnost podatkov, pri čemer je za manjkajoče vrednosti v šifriranih poljih zahtevana vrednost '0', v nešifriranih pa 'null'.

Dodatno preverjanje: Po oddaji podatkov, GU posreduje povratno datoteko z dodeljenimi identifikatorji, ki jih morajo upravitelji podatkov integrirati za zagotovitev usklajenosti prihodnjih posodobitev.

Ti ukrepi skupaj zagotavljajo, da so podatki v Zbirnem katastru GJI dosledni, celoviti in posodobljeni, kar omogoča zanesljiv vir informacij za upravljanje infrastrukture in prostorsko načrtovanje. Opisani elementi kakovosti so v dostopni dokumentaciji zelo obsežno opisani in z vidika semantične povezljivosti podajajo kontekst, v katerem podatke lahko celovito razumemo in nadalje uporabljamo.

V naslednjem poglavju, smo analizirali obstoj semantičnih elementov v dokumentaciji.

3.4.2.3 *Obstoj semantičnih elementov*

V tem delu smo opredelili obstoj semantičnih elementov za izbrane sloje zbirke nepremičninskih podatkov GURS, ki omogočajo razumevanje konteksta in so ključni za semantično povezljivost podatkov. Glede na pomanjkanje podatkov in dokumentacije za podatke Komunale Novo mesto, smo analizo za zadnje izpustili, smo pa opažanja zapisali v zaključku. Preverili smo (Preglednica 50):

- obstoj metapodatkov,
- obstoj ključnih besed, opisov in dokumentacije na ravni evidence, posameznih slojev,
- opredelitev entitet oziroma objektov, ter pomenov besed, ki opisujejo posamezne entitete,
- povezanost v širše sprejete klasifikacije (ontologije, taksonomije, šifranti, vokabularji na nacionalnem oziroma mednarodni ravni),
- dokumentiranost konteksta uporabe, kot so uporabniški primeri, okoliščine nastanka, pravne podlage, vrste deležnikov,
- obstoj naprednih semantičnih elementov kot so ontologije, taksonomije, grafi znanja na ravni organizacije, evidence, sloja ali tipa objektov.

S tem smo opredelili, kako dobro so objavljeni podatki pripravljeni za strojno razumevanje.

Preglednica 50 - Analiza obstoja semantičnih elementov za evidenco GJI

GJI Sloji			
OPIS KRITERIJA	Evidenca	Podatkovni sloj/kategorija	Šifranti
Obstoj metapodatkov	Izpolnjeni INSPIRE metapodatki	/	Podatkovni katalog
Opisi , ključne besede, dokumentacija	Izpolnjeni INSPIRE metapodatki	Podrobno opisane v dokumentaciji, izčrpna dokumentacija s primeri	Dokumentacija o izmenjevalnih formatih
Opredelitev entitet	<i>Ne nanaša se na to kategorijo</i>	Opredeljene v obliki stavkov oziroma izrazov.	Opredeljene v obliki stavkov oziroma izrazov.
Povezanost v širše klasifikacije	Del INSPIRE Tematike	Povezava v CC-SI Klasifikacijo	EID – koda unikatno določa objekt na ravni več evidenc
Kontekst	Del INSPIRE Tematike Deležniki so podrobno opisani v navodilih; Zakonodaja je dostopna, a več klikov stran	/	Zgolj referenca na pravilnik o Kategorizaciji cest
Napredni sem. Elementi	/	/	/

Z analizo smo ugotovili, naslednje:

- **Metapodatki** za na ravni evidence obstajajo in so zapisani in objavljeni skladno z zahtevami INSPIRE (objavljeno v strojno berljivi in strukturirani obliki). To omogoča hitrejše strojno odkrivanje evidence in njeno uporabo. Na ravni posameznih slojev/kategorij in šifrantov, metapodatkov v standardizirani obliki nismo našli, obstaja pa prosto dostopno gradivo v digitalni obliki (nestrukturirana oblika, ki ni v izvorno odprtokodnem formatu) katerega je te informacije možno pridobiti oziroma ustvariti.
- **Opisi, ključnih besed in dokumentacija ter opredeljenost entitet.**

Na ravni evidence, so v INSPIRE metapodatkih, opis evidence in ključne besede navedene in strojno dostopne preko portala v standardizirani, strukturirani, prosto objavljeni in strojno berljivi obliki. Za raven posameznih slojev oziroma kategorij in šifrantov, obstaja izčrpno prosto dostopno gradivo v digitalni obliki na različnih ravneh strojne berljivosti (nestrukturirana oblika, ki ni v izvorno odprtokodnem formatu (docx, pdf,), ter strukturirana oblika v odprtokodnem formatu z vključenimi unikatnimi identifikatorji (Json, GeoJson)), v katerem so podrobno razložene kratice, razlage konceptov, sheme in diagrami. Entitete so prav tako podrobno opisane s konkretnimi primeri zapisa v Json in GeoJson.

- **Povezanost v širše klasifikacije**, je na ravni evidence razviden iz umeščenost v tematske sklope INSPIRE. Kategorije oziroma sloji cest in kanalizacije, so dodatno povezane v CC_SI klasifikacijo, ki je določena z Uredbo o razvrščanju objektov⁷. Le ta pa temelji na mednarodni klasifikaciji Classification of types of construction⁸. Pri podatkih o cestah, se klasifikacija cest nanaša na uredbo o kategorizaciji državnih cest⁹
- **Kontekst**: Na ravni evidence je kontekst v obliki povzetka vira opisan v INSPIRE metapodatkih, pravna podlaga pa je na voljo na strani o zakonodaji, ki ureja delo geodestke uprave. Dokumentacija o tej evidenci zelo dobro dokumentira kako poteka nastajanje in

⁷ Uredba o klasifikaciji objektov; (<https://pisrs.si/pregledPredpisa?id=URED8497>)

⁸ Classification of types of construction (CC); [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Classification_of_types_of_construction_\(CC\)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Classification_of_types_of_construction_(CC))

⁹ Uredba o kategorizaciji državnih cest (<https://pisrs.si/pregledPredpisa?id=URED9046>)

vzdrževanje podatkov, kar nam daje širši kontekst o deležnikih v procesu. Manj je informacij o tem, kdo in na kakšne načine te podatke uporablja. Tudi zakonodaja je dostopna s povezavo na spletno stran, ki opisuje zakone in podzakonske predpise z delovnega področja Geodetske uprave. Informacija o zakonodaji je sicer na voljo, vendar dostop zahteva kar nekaj klikov stran ter iskanje v celotni pravni podlagi za GU. Na boljši način je zakonodaja opisana na straneh OPSI¹⁰, kjer so specificirane vse podlage, ki se nanašajo na to evidenco.

- **Napredni semantični elementi:** V okviru javno dostopne dokumentacije zbirnega katastra javne gospodarske infrastrukture, za evidenco, njene sloje ali šifrate, je stanje glede naprednih semantičnih elementov naslednje:
 - o **Ontologije, taksonomije** – vsi elementi GJI so umeščeni v CC_SI klasifikacijo, ki je neke vrste taksonomija. Ontologij za posamezne kategorije znotraj GJI nismo našli.
 - o **Podatkovni modeli** – podatkovni modeli javno niso objavljeni, jih je pa možno pridobiti s strani upravljavca zbirke, a to ni nikjer eksplicitno navedeno.
 - o **Principi Linked data** – vsi objekti imajo določeno enotno ID številko na ravni države, podatki pa se prevzemajo v Json in GeoJson obliki. To so dobri pogoji za povezovanje po principu Linked data.
 - o **Semantične anotacija** – Podatki nimajo semantičnih anotacij na strojni ravni, so pa v dostopni dokumentaciji dovolj dobro opisano, da bi lahko z obdelavo podatkov to spremenili v semantične anotacije.

Če povzamemo, so podatki na ravni evidence dobro standardizirani in strojno berljivi, vendar na ravni posameznih slojev in šifrantov obstajajo vrzeli, predvsem glede uporabe naprednih semantičnih elementov. Gradivo sicer omogoča nadaljnji razvoj in nadgradnjo v avtomatizirano semantično povezljivost. Dodatno ga je možno obogatiti z opisi iz občinskih zakonskih podlag, navodil za izvajalce vzdrževanja na lokalni ravni in dokumentacije ostalih deležnikov (Specifikacije in promocijsko gradivo dobaviteljev opreme)

V naslednjem odstavku smo, glede na razvite mere opredelili dostopnost do semantičnih elementov, s čimer smo dobili informacijo, kako dobre pogoje imajo podatki za semantično povezovanje z vidika skladnosti in dostopnosti.

¹⁰ OPSI Opis GJI (<https://podatki.gov.si/dataset/zbirni-kataster-gospodarske-javne-infrastrukture>)

3.4.3 Mere dostopnost semantičnih elementov GJI

V tem delu smo ocenili dostopnost in verodostojnost semantičnih elementov za izbrane sloje zbirke nepremičninskih podatkov GURS, ki z vidika dostopnosti in semantične celovitosti in skladnosti, vplivajo na semantično povezljivost zbirk. Ocenili jih smo z izbranimi merami, kot so tehnična (sintaktična) dostopnost, organizacijska dostopnost, stopnja uradnosti in semantična bogatost vira. V sklopu posamezne zbirke smo primarno upoštevali vse vire, ki so bili z visoko stopnjo odprtosti pridobljeni na javnih spletnih straneh upravljavca evidence; drugotno smo pozvali skrbnike evidence, da nam priskrbijo interno dokumentacijo (Preglednica 51).

Preglednica 51 - Analiza dostopnosti gradiva v evidenci in šifrantih GJI

GJI	Evidenca	Točke	Šifranti	Točke
Tehnična dostopnost vira pomena (Sintaktična dostopnost)	Spletno dostopni standardizirani servisi	6	Pisni viri v digitalni obliki	3
Organizacijska dostopnost vira	Splošno mednarodno dostopna in sprejeta vsebina	5	Opisi so del dokumentacije evidence v vsebinski domeni (strokovna javnost)	3
Stopnja uradnosti	Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti..., Pravna praksa (precedenčne sodbe))	6	Opisi so del strokovnih specifikacij in posebnih vokabularjev	4
Semantična bogatost vira	odstavek	3	podroben opis s primeri, skicami ali slikovnim gradivom	5
Starost vira	Trenutno veljavno	5	Trenutno veljavno	5
TOČK		25		20
DELEŽ (%)		86		69

Na ravni zbirke nepremičninskih podatkov imamo strojno razumljive vire, ki so objavljeni tako na slovenskem INSPIRE portalu, kot tudi na mednarodnih portalih. Omogočen je strojni dostop do opisov RDF, kar je že strojno zelo razumljiva raven. Zato je ocena teh virov dosegla 86 točk od 100, kar že predstavlja visoko osnovo za semantično povezljivost

Na ravni posamičnih slojev oziroma kategorij in šifrantov, so podatki večinoma odprto dostopni preko spleta, vendar v ne odprtokodnih in nestrukturiranih oblikah. Določen del dokumentacije, je na voljo v odprtokodnem standardiziranem formatu, z vključenimi unikatnimi identifikatorji, kar povečuje možnost strojne povezljivosti. Večina gradiva temelji na strokovnih specifikacijah in je tudi del pravnih podlag, kar kaže na visoko stopnjo družbene sprejetosti. Semantično podatki niso na napredni ravni, dostopna dokumentacija sicer omogoča nadaljnji razvoj in nadgradnjo v avtomatizirano semantično povezljivost.

V naslednjem poglavju smo obravnavane podatke umestili v razviti semantični referenčni okvir, kar nam bo v nadaljevanju pomagalo pri oceni možnosti povezovanja tematsko podobnih evidenc glede na njihovo tematsko prekrivanje.

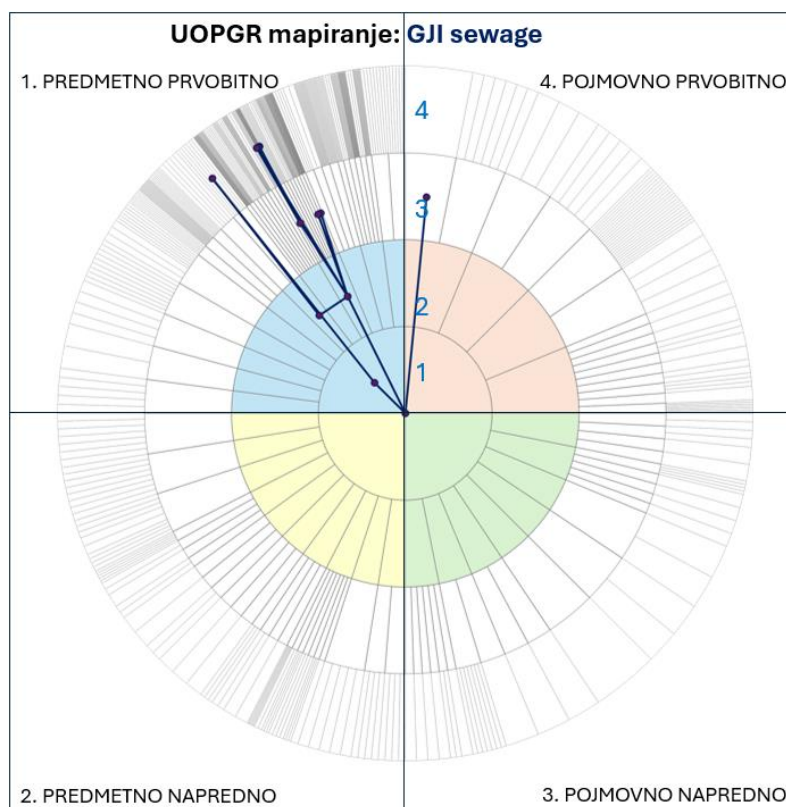
3.4.4 Umestitev šifrantov in njihovih vrednosti v taksonomijo UOPGR

V tem delu smo izbrane podatke in vrednosti v njihovih šifrantih umestili v razviti semantični referenčni sistem UOPGR, glede na njihovo tematiko in semantično globino.

3.4.4.1 Umestitev šifrantov GJI Kanalizacija (Linije in Točke)

Preglednica 52 - Umeščenosti vrednosti šifrantov sloja GJI Sewage v UOPGR s semantično globino

UOPGR	UOPGR OPIS KATEGORIJE	Sem. Globina	RABA_ID	RABA_OPIS
10402	transportne signalne in nadzorne naprave	3	3207	Oprema
1040903	kanalizacija	4	3201	Kanalizacijski vodi
1040903	kanalizacija	4	3299	Drugi objekti kanalizacijske infrastrukture
1040906	crpalisce	4	3202	Črpališče
1040907	izliv	4	3205	Izpust iz kanalizacijskega sistema
1040912	Jasek	4	3206	Jašek
1040912	Jasek	4	3203	Razbremenilnik
1040912	Jasek	4	3209	Zadrževalnik
1040912	Jasek	4	3210	Ponikovalnica padavinske vode
1060104	cistilna	4	3204	Čistilna naprava za odpadno vodo
40402	Sheme aktivnosti GOSPODARJENJA Z GRAJENIMI OBJEKTI	3	3208	Območje objekta kanalizacijskega sistema



Slika 21 - Graf umeščenosti vrednosti šifrantov sloja GJI Sewage v UOPGR

Opomba: Trenutni model UOPGR se deli na 4 kategorije, ki se skupno delijo na 34 razredov, ki se delijo skupno na 105 podrazredov in obsegajo 446 entitet (objektov, shem ali pojavov) v četrti globini. Skupno 590 konceptov.

V podatkovnem sloju GJI Sewage, smo od skupno 11 opisov identificirali 9 opisov iz šifrantov, ki ustrezajo entitetam na 4. semantični globini UOPGR. Zaradi presplošnega opisa smo 2 opisa uvrstili v šifrantih v podrazreda na 3. ravni (glej Slika 21). Umeščanje smo izvedli glede na trenutno raven razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov, bodo te povezave lahko bolj številčne in bolj bogate in s tem doprinesle k več možnostim povezovanja.

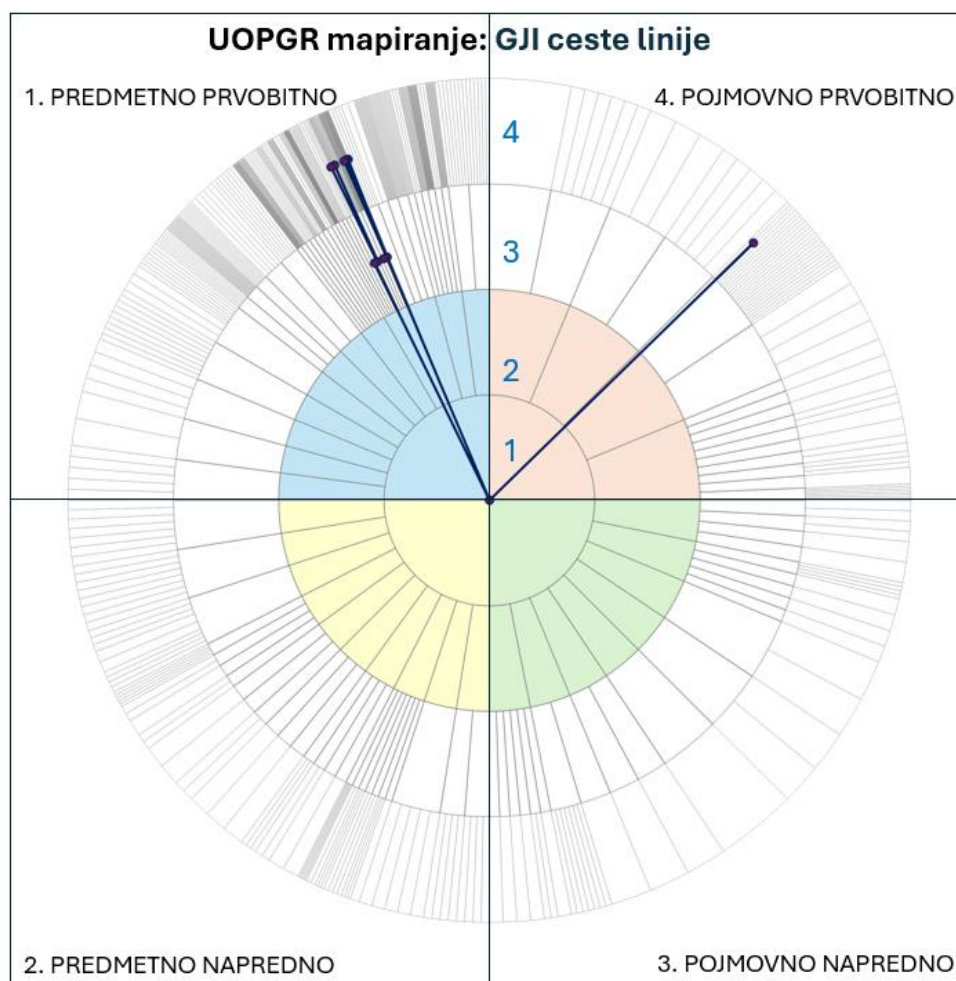
Na grafu je razvidno, da se opisi iz šifrantov večinoma umeščajo med Predmetne objekte Prvobitnih človeških aktivnosti (1. Kategorija). En atribut pa opisuje Pojemne sheme Prvobitnih človeških aktivnosti (4. Kategorija). V 2. in 3. kategoriji ni nobenega opisa. Iz tega sklepamo, da je podatek dobro umeščen v referenčni model, hkrati je GJI zelo specializirana evidenca.

3.4.4.2 GJI Ceste

Preglednica 53 - Umeščenosti vrednosti šifrantov sloja GJI Ceste v UOPGR s semantično globino

UOPGR	UOPGR OPIS KATEGORIJE	Sem. Globina	RABA_ID	RABA_OPIS
10401	Objekti aktivnosti PREMOSKANJA	3	1102	Objekt cestne infrastrukture
1040101	most	4	1	Most
1040102	predor	4	6	Predor
1040107	nadvoz	4	2	Nadvoz
1040108	podvoz	4	3	Podvoz
1040110	viadukt	4	5	Viadukt
1040111	galerija	4	6	galerija
10402	transportne signalne in nadzorne naprave	3	1199	Drugi objekti cestne infrastrukture
1040401	cesta	4	1101	Cesta ali pot (os ceste ali poti)
1040401	cesta	4	1	Avtocesta
1040401	cesta	4	2	Hitra cesta
1040401	cesta	4	3	Glavna cesta I. reda
1040401	cesta	4	4	Glavna cesta II. reda
1040401	cesta	4	5	Regionalna cesta I. reda
1040401	cesta	4	6	Regionalna cesta II. reda
1040401	cesta	4	7	Regionalna cesta III. reda
1040401	cesta	4	8	Turistična cesta
1040401	cesta	4	9	Lokalna cesta
1040401	cesta	4	11	Glavna mestna cesta
1040401	cesta	4	12	Zbirna mestna ali krajevna cesta
1040401	cesta	4	13	Mestna ali krajevna cesta
1040401	cesta	4	17	Gozdna cesta
1040401	cesta	4	18	Nekategorizirana cesta
1040407	pot	4	10	Javna pot
1040407	pot	4	14	Daljinska kolesarska pot

1040407	pot	4	15	Glavna kolesarska pot
1040407	pot	4	16	Javna pot za kolesarje
10405	tirni transport	3	4	Železnica, če je na nadvozu
4020221	stacionaža	4	1	Stacionaža upravnega tipa
4020221	stacionaža	4	2	Dodana stacionaža
4020221	stacionaža	4	3	Navidezna stacionaža



V podatkovnem sloju GJI Sewage, smo od skupno 31 opisov identificirali 28 opisov iz šifrantov, ki ustrezajo entitetam na 4. semantični globini UOPGR. Zaradi presplošnega opisa smo uvrstili 3 opise v šifrantih v podrazreda na 3. globini (glej Slika 22). Umeščanje smo izvedli glede na trenutno raven razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov, bodo te povezave lahko bolj številčne in bolj bogate in s tem doprinesle k več možnostim povezovanja.

Na grafu je razvidno, da se večina opisi iz šifrantov umeščajo med Predmetne objekte Prvobitnih človeških aktivnosti (1. Kategorija). En atribut pa opisuje Pojemne sheme Prvobitnih človeških aktivnosti (4. Kategorija). V 2 in 3 kategoriji ni nobenega opisa. Iz tega sklepamo, da je podatek dobro umeščen v referenčni model, hkrati je to zelo specializirana evidenca.

3.5 Rezultati naloga 3.4: Testna analiza možnosti semantične povezljivosti kmetijskih, gozdarskih zbirk podatkov

3.5.1 Opredelitev podatkov za analizo kmetijskih, gozdarskih zbirk podatkov

V tej skupini kmetijskih in gozdarskih podatkov smo obravnavali naslednje evidence:

Preglednica 54 - Opredelitev izbranih slojev/kategorij kmetijskih in gozdarskih podatkov za analizo semantične povezljivosti

Ime sloja	Opis	Vir
RABA_2024_09_30.shp	Grafični podatki evidence dejanske rabe kmetijskih zemljišč	https://rkg.gov.si/vstop/
odseki gozdni (wfs)	Grafični podatki gozdnih odsekov	https://prostor.zgs.gov.si/pregledovalnik/
Sestoji gozdni (wfs)	Grafični podatki sestojev	https://prostor.zgs.gov.si/pregledovalnik/
Gozdne ceste (wfs)	Grafični podatki gozdnih cest	https://prostor.zgs.gov.si/pregledovalnik/

3.5.1.1 Evidenca dejanske rabe kmetijskih zemljišč

Evidenca dejanske rabe kmetijskih in gozdnih zemljišč (v nadaljevanju: Evidenca dejanske rabe) je enotna državna evidenca o dejanski rabi zemljišč. Primarno je namenjena ugotavljanju dejanskega stanja rabe zemljišč kot pogoja za izvajanje ukrepov skupne kmetijske politike Evropske unije. Posredno in neposredno se podatki Evidence dejanske rabe uporabljajo tudi pri določanju dejanske rabe na parcelo v zemljiškem katastru, višine katastrskega dohodka, odškodnine za spremembo namembnosti kmetijskih zemljišč, spremembi bonitete zemljišč, ugotavljanju upravičenosti do socialnih podpor, za statistične analize, raziskave, ipd. Podatki o gozdu v Evidenci dejanske rabe se usklajujejo s podatki iz evidenc Zavoda za gozdove Slovenije (ZGS).

Evidenco in njeno dokumentacijo smo v shp obliki pridobil na spletni strani <https://rkg.gov.si/vstop/>. Hkrati je za pregled dostopna tudi na njihovem spletnem pregledovalniku

(<https://rkg.gov.si/GERK/WebViewer/>).

3.5.1.2 Podatki iz pregledovalnika zavoda za gozdove

Iz velikega obsega podatkov o gozdovih in evidencah za njihovo upravljanje, pregledovalnika Zavoda za gozdove, smo izbrali tri sloje. Odseki gozdni in sestoji, z namenom pregleda možnosti povezovanja s podatki o pokrovnosti tal in v vrste rabe. Dodatno smo izbrali še sloj gozdne ceste, za možnost primerjave z ostalimi sloji, ki se nanašajo na ceste.

3.5.2 Ocena pogojev za semantično povezovanje

Da bi podatke med seboj lahko semantično povezovali, je treba poznati tudi ostale tehnične in vsebinske omejitve, ki se ne nanašajo samo na razumevanje konteksta. Zato smo pred samo analizo povezljivosti semantičnih elementov, opravili pregled tehničnih oziroma sintaktičnih lastnosti, ki omogočajo

povezovanje, obstoj informacij o izbranih elementih kakovosti, ter pregled obstoja semantičnih elementov.

3.5.2.1 Sintaktična povezljivost kmetijskih, gozdarskih zbirk podatkov

V tem delu smo opredelili sintaktične elemente za izbrane sloje kmetijske in gozdarskih zbirk podatkov, ki posledično vplivajo na svojo semantično povezljivost. Določili smo format zapisa geometrije in pripadajoče dokumentacije, ocenili razčlenjenost zbirke na eno ali več tabel, ter ugotovili, ali za posamezni zapis obstaja enolični identifikator in na kakšnem ravni (lokalno, državno, globalno). S tem smo opredelili, kako dobro so objavljeni podatki pripravljeni za strojno razumevanje (Preglednica 55).

Preglednica 55 - Sintaktična povezljivost virov evidence Raba tal kmetijskih zemljišč

Sintaktična povezljivost: Kmetijska raba			
KRITERIJ	Opis Evidence	Podatkovni sloj/kategorija	Šifranti, Atributi
Format zapisa geometričnega sloja in atributov	<i>Ne nanaša se na to kategorijo</i>	- ESRI Shp datoteka (zip) - Spletna aplikacija (ni WFS)	- Dbf, wfs
Dodatna dokumentacija	- Pdf dokumenti - Html strani	Pdf dokument	- Pdf tabele
Enostavnost/ Kompleksnost	<i>Ne nanaša se na to kategorijo</i>	Enostavno (navezava na šfrante)	- Pdf dokumenti - Html strani - Ni za vse attribute
Obstoj enoličnega identifikatorja:	/	ID na ravni evidence; raven organizacije	DA – Id vrste rabe
Strojna berljivost (5 Star)	1- PDF, Html	1- ESRI shp	- DBF, PDF

Glede na opisano, lahko zaključimo, da je podatkovni vir Evidence dejanske rabe sintaktično relativno enostaven, vendar je zapisan v formatih, ki imajo najnižjo stopnjo strojne razumljivosti (1 zvezdica), kar pomeni, da bo za semantično razumevanje potrebna interpretacija vsebine v pdf dokumentih, preden bo mogoče nadaljnje povezovanje. Obstoj enoličnega identifikatorja objekta predstavlja dobro izhodišče za nadaljnje povezovanje vsaj na državni ravni.

Preglednica 56 - Sintaktična povezljivost gozdarskih podatkov Pregledovalnika ZGS

Sintaktična povezljivost: ZGS Portal			
KRITERIJ	Opis Evidence	Podatkovni sloj/kategorija	Šifranti, Atributi
Format zapisa geometričnega sloja in atributov	<i>Ne nanaša se na to kategorijo</i>	WFS servis	Pdf
Dodatna dokumentacija	Pdf, html	Pdf	Pdf
Enostavnost/ Kompleksnost	<i>Ne nanaša se na to kategorijo</i>	Enostavno (ni zunanjih šifrantov)	<i>Ne nanaša se na to kategorijo</i>
Obstoj enoličnega identifikatorja:	/	Na ravni objekta,	<i>Ne nanaša se na to kategorijo</i>
Strojna berljivost (5 Star)	1 – Html spletna stran	3 – WFS	1- pdf

Podatki iz portala ZGS so dostopni preko WFS Servisa, vsa dokumentacija o slojih in njihovih šifrantih, je navolga v pdf dokumentu, ki smo ga prenesli s spletne dokumentacije. Podatki so organizirani

preprosto, brez zunanjih šifrantov. Vsi objekti imajo na ravni evidence ZGS svoje enolične identifikatorje. S tehničnega vidika, je evidenca enostavna za dostop preko strukturiranega, odprtokodnega, objavljenega formata (WFS), njena dokumentacija pa le v neodprtokodnem in nestrukturiranem formatu.

V naslednjem razdelku smo pregledali, kako so za podatkovni vir opisani elementi kakovosti, kar nam bo pri semantičnem povezovanju dalo informacijo o uporabnosti in zanesljivosti semantičnega povezovanja.

3.5.2.2 Opis elementov kakovosti

V tem delu smo opredelili opis izbranih elementov kakovosti podatkovnega vira za Evidenco dejanske rabe, ki posledično vplivajo na njeno uporabnost pri semantičnem povezovanju predvsem z vidika opredelitve konteksta. Ugotovili smo obstoj informacij o izvoru podatkovnega vira (angl. lineage) na ravni geometrije podatkov o lastnostih objekta (atributov), o zgodovini transformacij podatkov in metodologijah zajema, kako je opisana časovna umeščenost veljavnosti informacij o geometriji in lastnostih objekta, ter o konsistentnosti objave podatkov znotraj sloja in evidence.

3.5.2.2.1 Elementi kakovosti evidence dejanske rabe kmetijskih zemljišč

Preglednica 57 - Analiza obstoja elementov kakovosti za podatke evidence dejanske rabe kmetijskih zemljišč

Elementi kakovosti: Kmetijska raba		
OPIS KRITERIJA	Elementi kakovosti	Pomanjkljivosti
Izvor, metodologija	Izvor je podrobno opisan v: RABA_Beri_Me.pdf RABA_IntKljuc_20131009.pdf Vrste_dejanske_rabe_zemljišč.pdf	Nismo zasledili opisa metodologije zajema geometrije – nedoločena natančnost podatka
Natančnost geometrije	Zajem v merilu 1:5000	
Časovna umeščenost	Datum zajema v atributu [D_OD] Dostopnost arhivskih podatkov Opis spremembe metodologije	Časovni zamik med dejanskim stanjem in evidenco znaša do 3 leta
Konsistentnost podatkov	Ključni atributi so izpolnjeni 100%	Atributi o preverjanju kakovosti ponekod niso izpolnjeni v atributih USER_ID_ZA; MANAGER; STANJE in VERIFIED ; Evidenca prikazuje različen podatkovni model v shp datoteki za prenos glede na objavljen sloj v aplikaciji

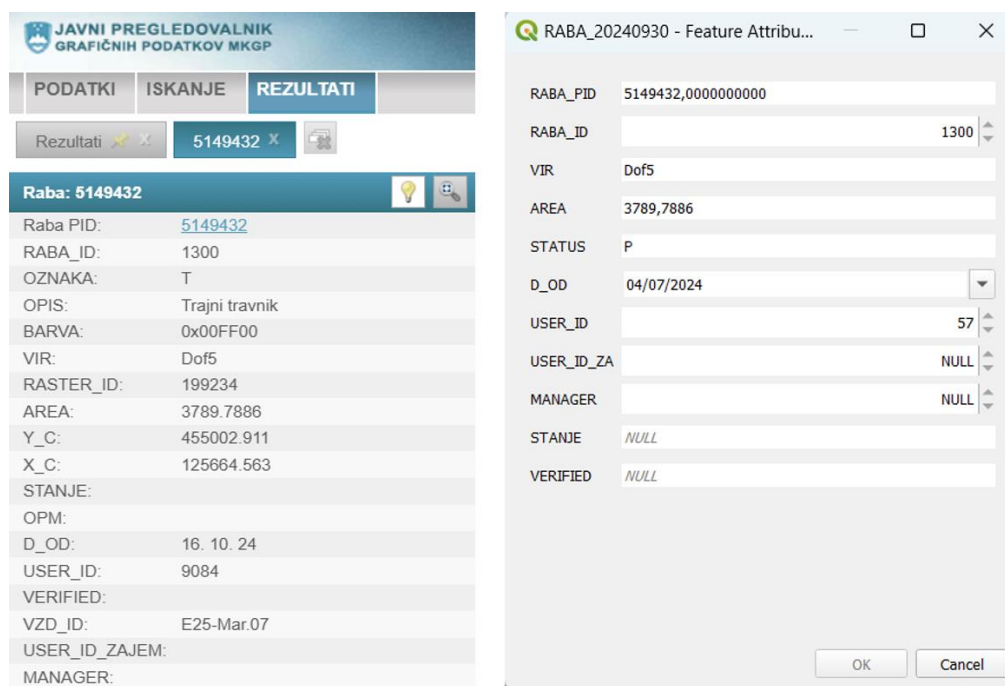
- **Izvor Evidence dejanske rabe kmetijskih zemljišč**, je na splošno opisan v dokumentu RABA_Beri_Me.pdf, ker je okvirno opisana metodologija zajema in viri. Primarni vir je računalniško podprta fotointerpretacija digitalnih ortofoto posnetkov, dodatni viri pa so še zajemi iz satelitskih posnetkov, terenske meritve (ARSKTRP, MKGP, IRSKGLR), vloge strank ter podatki Zavoda za gozdove Slovenije (za gozdne površine). Iz dokumentacije pa ni razvidna podrobna metodologija zajema geometrije. Metodologija zajema ključne lastnosti posameznega objekta (Snemalna Pravila in kriteriji za izbora), so izčrpno opisane v ločenem dokumentu Interpretacijski ključ (RABA_IntKljuc_20131009.pdf) in delno tudi v dokumentu Vrste_dejanske_rabe_zemljišč.pdf, ki je na voljo na isti spletni strani. Na voljo so tudi povezave do

arhivskih podatkov, kot tudi povezave na dokumentacijo o spremembi metodologije zajema skozi čas.

- **Časovna umeščenost** je prav tako opredeljena v atributu [D_OD], ki vsebuje datum zadnje spremembe ali pregleda poligona evidence dejanske rabe. Poleg tega, je iz dokumentacije razvidno, da je časovni zamik med dejanskim stanjem in evidenco lahko znaša do 3 leta, zaradi možne razlike v starosti virov za zajem.
- **Konsistentnost, evidence in podatkov** je pomembna za semantično povezljivost z vidika zagotavljanja enovitosti povezovanja na isti objekt na enak način. Ta parameter kakovosti smo preverjali na ravni enovitosti podatkov evidence, objavljenih na različnih virih.

Ugotovljene neskladnosti podatkovnega modela iz različnih virov

Na spletni strani MKGP je možno evidenco RABA pregledovati v Javnem pregledovalniku grafičnih podatkov MKGP-RKG ali pa prenesti v obliki shape datoteke. Ob pregledu obeh virov, smo ugotovili, da podatkovni model prenesene datoteke ni identičen podatkovnemu modelu objavljenega sloja na javnem pregledovalniku, kot je razvidno iz (Slika 23).



Slika 23 - Na levi primer podatkovnega modela iz MKGP pregledovalnika, na desni struktura atributov s prenesenega sloja

Iz slike je razvidno, da podatkovni model na javnem pregledovalniku vsebuje več elementov (npr. OZNAKA, OPIS, BARVA, RASTERID, X_C, Y_C, OPM ter VZD_ID), ki niso opisani v dokumentaciji podatkovnega modela.

Iz opisanega je razvidno, da so izbrani elementi kakovosti za evidenco na voljo in omogočajo opredelitev kakovosti opisanega objekta predvsem s pogleda časovne umeščenosti kot tudi z vidika zagotavljanja informacij o virih podatkov. Z vidika časovne umeščenosti, je možno iz podatkov in dokumentacije izbrati časovni okvir, znotraj katerega je bil podatek veljaven. V osnovnih podatkih ni informacije o časovni nehomogenosti virov za določena področja, so pa ti podatki na voljo v arhivskih podatkih.

V dostopni dokumentaciji ni podrobnih informacij o natančnem postopku izvajanja metod zajema s primarne metode ali ostalih postopkov, predvsem z vidika natančnosti določitve geometrije. Metoda klasifikacije območij je izčrpno opisana v dokumentu »Interpretacijski ključ«, kar jasno opredeli mejne pogoje za umestitev v klasifikacijo objektov v podatkovnem sloju. Dodatno je iz podatkovne strukture razvidno tudi dejstvo, ali je bila izvedena dodatna kontrola, kar za tiste objekte poda dodatno informacijo o stopnji zaupanja.

Z vidika konsistentnosti distribuiranih različic, smo ugotovili neskladnosti, ki se kažejo predvsem v razlikah med objavljenimi podatkovnimi modeli, ter njihovimi opisi. Evidenca RABA, ki jo prenesemo v obliki shp datoteke ima podatkovni model relativno dobro dokumentiran, medtem, ko ima, na javnem portalu objavljeni sloj RABA, razširjen podatkovni model, za katerega pa nismo našli dodatnih opisov.

V naslednjem razdelku smo pregledali, kateri semantični artefakti so na voljo v javno dostopni dokumentaciji, ki nam bo predstavljala osnovo za končno analizo semantične povezljivosti.

3.5.2.2.2 Elementi kakovosti slojev iz zbirke Zavoda za Gozdove RS

Preglednica 58 – Dokumentacija elementov kakovosti gozdarskih podatkov Pregledovalnika ZGS

Elementi kakovosti: ZGS		
KRITERIJA	Elementi kakovosti	Pomanjkljivosti
Izvor, metodologija	/	Ni opisa v dostopni dokumentaciji
Natančnost geometrije	/	V dokumentaciji, tega podatka ni. Upravljaavec zbirke nam tega podatka ni posredoval
Časovna umeščenost	Časovna umeščenost podatka je vidna samo v aplikaciji z izboromo prikaza metapodatkov	Tega podatka ni v dokumentaciji niti v podatkovnem modelu
Konsistentnost podatkov	- Standardizirani formati - enotni CRS - Unikatni ID	

Za podatke zbirke ZGS na javno dostopnih spletnih straneh nismo našli podrobnejše dokumentacije o večini elementov kakovosti (Izvor, metodologija zajema, Natančnost geometrije). Časovna umeščenost podatkov je dostopna samo skozi funkcionalnost aplikacije, kar ni praktično za obdelavo večjih količin podatkov. Podatki so celoviti, imajo unikatni ID za objekte in enoten koordinatni sistem.

Z vidika kakovosti, za dobro semantično povezovanje manjka precej podatkov. Glede na ostalo strokovno literaturo, so ti podatki nekje zajeti, a bi jih bilo treba obdelati in objaviti.

3.5.2.3 Obstoje semantičnih elementov

V tem delu smo opredelili obstoj semantičnih elementov za izbrane sloje kmetijske in gozdarskih zbirk podatkov, ki omogočajo razumevanje konteksta in so ključni za semantično povezljivost podatkov. Preverili smo:

- obstoj metapodatkov, ključnih besed, opisov in dokumentacije na ravni evidence, posameznih slojev, ter na ravni opisov posameznih objektov v šifrantih,
- povezanost v širše sprejete klasifikacije (ontologije, taksonomije, šifranti, vokabularji na nacionalnem oziroma mednarodni ravni),
- dokumentiranost konteksta uporabe, kot so uporabniški primeri, okoliščine nastanka vrste deležnikov, ter pomenov besed, ki opisujejo posamezne entitete,

- obstoj naprednih semantičnih elementov, kot so ontologije, taksonomije, grafi znanja na ravni organizacije, evidence, sloja ali tipa objektov.

S tem smo opredelili, kako dobro so objavljeni podatki pripravljeni za strojno razumevanje.

3.5.2.3.1 Semantični elementi evidence rabe kmetijskih zemljišč

Preglednica 59 - Obstoj semantičnih elementov: Evidenca rabe kmetijskih zemljišč

Analiza obstoja semantičnih elementov evidence rabe kmetijskih zemljišč			
OPIS KRITERIJA	Evidenca	Podatkovni sloj	Šifranti
Obstoj metapodatkov	<i>Inspire portal (sloj ni enak tistemu na straneh MK)</i>	<i>- Inspire portal (sloj ni enak tistemu na straneh MK)</i>	Podatkovni katalog je podrobno opisan za del atributov [RABA]
Opisi , ključne besede, dokumentacija	INSPIRE metapodatki; Zakonodaja	- ESRI Shp datoteka (zip) Spletna aplikacija (ni WFS)	Podroben opis tipov rabe z besedilom in slikami
Opredelitev entitet	/	/	Vrste entitet (tip rabe tal) so opredeljene s podrobnimi opisi in slikami
Povezanost v širše klasifikacije	Tematski opis skozi INSPIRE	Tematski opis skozi INSPIRE	/
Kontekst	Zakonodaja Pravilnik	/	Podroben opis tipov rabe z besedilom in slikami
Napredni sem. Elementi	NE	NE	NE

S pregledom obstoja semantičnih elementov, smo ugotovili, da je na ravni evidence obstajajo strojno berljivi metapodatki v INSPIRE metapodatkovnem portalu, kar omogoča lažje odkrivanje vira podatkov, a hkrati ta vir ni enak viru, ki smo ga prenesli s spletne strani upravljavca podatkov, kar kaže na nekonsistentnost. Na ravni šifrantov, je ta podatkovni vir eden izmed najbolj semantično opisanih, saj za posamezen tip vrste rabe obstaja podroben opis pomena in načina zajema podatka, skupaj s slikovnim primerom.

Hkrati so vsi opisi na nižji stopnji strojne razumljivosti, zapisani v pdf formatu so sicer digitalno berljivi. Semantična povezljivost bi bila boljša, če bi obstajali napredni semantični elementi, kot so ontologije, ali če bi se šifranti jasno navezovali na mednarodno standardizirane klasifikacije.

3.5.2.4 Semantični elementi podatkov iz zbirke Zavoda za Gozdove RS

Preglednica 60 -Obstoj semantičnih elementov: podatki s portala ZGS

GJI Sloji			
OPIS KRITERIJA	Evidenca	Podatkovni sloj/kategorija	Šifranti
Obstoj metapodatkov	NE	Na ravni aplikacije	Opis podatkov in strukture, ter pomen šifrantov
Opisi , ključne besede, dokumentacija	Spletna stran ZGS	Opis v odstavkih v pridobljeni dokumentaciji	Opis podatkov in strukture, ter pomen šifrantov
Opredelitev entitet	<i>Ne nanaša se na to kategorijo</i>	Opis v odstavkih v pridobljeni dokumentaciji	Opis podatkov in strukture, ter pomen šifrantov
Povezanost v širše klasifikacije	NE	NE	NE
Kontekst	NE	NE	NE
Napredni sem. Elementi	NE	NE	NE

Semantičnih elementov za podatke s portala ZGS je zelo malo. Metapodatkov v strojno berljivi, standardizirani obliki ni, Opisi posameznih evidenc (slojev) in njihovih atributov je dostopen v dokumentaciji v obliki pdf, a je podrobno opisana le na ravni sloja/kategorije. Na ravni šifranta je samo enostavna poved, ki pogosto ne obrazloži pomena laiku. Prav tako nismo zasledili umeščanja v širše klasifikacije, razen 2 atributov, ki se navezujeta na GU.

Semantičnih elementov, razen podrobnih opisov evidenc in »samopisujočih« se vrednosti v šifrantih, ter unikatnega ID-ja nismo našli. Na poziv upravljavcu zbirke za dodatno dokumentacijo ni bilo pozitivnega odziva.

V naslednjem odstavku smo, glede na razvite mere opredelili dostopnost do semantičnih elementov, s čimer smo dobili informacijo, kako dobre pogoje imajo podatki za semantično povezovanje z vidika skladnosti in dostopnosti.

3.5.3 Mere dostopnost semantičnih elementov

V tem delu smo ocenili dostopnost in verodostojnost semantičnih elementov za izbrane sloje kmetijske in gozdarskih zbirk podatkov, ki z vidika dostopnosti in semantične celovitosti in skladnosti, vplivajo na semantično povezljivost zbirk. Ocenili jih smo z izbranimi merami, kot so tehnična (sintaktična) dostopnost, organizacijska dostopnost, stopnja uradnosti in semantična bogatost vira. V sklopu posamezne zbirke smo primarno upoštevali vse vire, ki so bili z visoko stopnjo odprtosti pridobljeni na javnih spletnih straneh upravljavca evidence; drugotno smo pozvali skrbnike evidence, da nam priskrbijo interno dokumentacijo.

3.5.3.1 Evidenca dejanske rabe kmetijskih zemljišč

Preglednica 61 - Analiza dostopnosti semantičnih elementov: Evidenca rabe tal kmetijskih zemljišč

Kmetijska raba	Evidenca	Točke	Šifranti	Točke
Tehnična dostopnost vira pomena (Sintaktična dostopnost)	Lokalni Standardizirani strukturirani viri v digitalni obliki	5	Pisni viri v digitalni obliki	3
Organizacijska dostopnost vira	Opisi so del dokumentacije evidence v vsebinski domeni (strokovna javnost)	3	Opisi so del dokumentacije evidence Organizacije	2
Stopnja uradnosti	Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti..., Pravna praksa (precedenčne sodbe))	6	opisi so del besedil znanstvenih in strokovnih knjig, učbenikov, priročnikov ter tehničnih načrtov, risb in skic	3
Semantična bogatost vira	odstavek	3	podroben opis s primeri, skicami ali slikovnim gradivom	4
Starost vira	2019 (trenutno veljavno)	5	2013 (trenutno veljavno)	5
TOČK		22		17
DELEŽ(%)		76		59

Na ravni evidence rabe kmetijskih zemljišč, imamo dokumentacijo, ki nam omogoča semantično razumevanje evidence. Ti podatki so večinoma dostopni le na ravni organizacije, delno tudi skozi portal INSPIRE, vendar zaradi prej opisane neskladnosti med objavljenimi podatki tega nismo upoštevali. Časovna natančnost virov je mlajša od 5 let, vendar je trenutno še vedno veljavna, kar jih umešča med najvišjo časovno natančnost. Zato je ocena teh virov 76 točk od 100, kar samo po sebi daje evidenci dobro oceno za semantično povezovanje.

Na ravni šifrantov rabe kmetijskih zemljišč, imamo zelo izčrpne opise o vrednostih v šifrantih, ki so opremljene s slikami, kar nam omogoča dobro semantično osnovo za razumevanje podatkov. Strojnim aplikacijam je vsebina dostopna pretežno v pdf formatu, kar ni najbolj dostopno, hkrati pa so opisi le organizacijske narave, čeprav javno dostopni. Relativna starost opisov vrste rabe (več kot 10 let), bi lahko bila z vidika časovne natančnosti zahtevala premislek, vendar če se metodologija opredelitve vrste

rabe ni spreminjala in je trenutno veljavna jo to še vedno uvršča med časovno najbolj natančne vire. Zato je ocena teh virov 59 točk od 100, kar samo po sebi daje evidenci dobro oceno za semantično povezovanje, vendar z zmanjšano strojno razumljivostjo.

3.5.3.2 Gozdarske zbirke podatkov Zavoda za gozdove

Preglednica 62 - Analiza dostopnosti semantičnih elementov: gozdarski podatki zbirke GZS

ZGS	Evidenca	Točke	Šifranti	Točke
Tehnična dostopnost vira pomena (Sintaktična dostopnost)	Lokalni Standardizirani strukturirani viri v digitalni obliki	5	Pisni viri v digitalni obliki	3
Organizacijska dostopnost vira	Opisi so del dokumentacije evidence v vsebinski domeni (strokovna javnost)	3	Opisi so del dokumentacije evidence Organizacije	2
Stopnja uradnosti	Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti..., Pravna praksa (precedenčne sodbe))	6	opisi so del besedil znanstvenih in strokovnih knjig, učbenikov, priročnikov ter tehničnih načrtov, risb in skic	3
Semantična bogatost vira	odstavek	3	podroben opis s primeri, skicami ali slikovnim gradivom	4
Starost vira	2019	5	2013	1
TOČK		22		13
DELEŽ(%)		76		45

Dostopnost do majhnega nabora semantičnih elementov, ki smo jih identificirali v prejšnjem poglavju, smo pridobile dokumentacije s spletne strani, iskali še dodatne informacije v pravnih podlagah in drugih virih, kot je prosto dostopna strokovna literatura. Odkrili smo, da je kar nekaj pravne podlage, kar bi lahko izkoristili kot semantični vir. Na ravni evidence, je zadnja dokumentacija starejša od 3 let in mlajša kot 5.

V naslednjem poglavju smo obravnavane podatke umestili v razviti semantični referenčni okvir, kar nam bo v nadaljevanju pomagalo pri oceni možnosti povezovanja tematsko podobnih evidenc glede na njihovo tematsko prekrivanje.

3.5.4 Umestitev šifrantov in njihovih vrednosti v UOPGR taksonomijo

V tem delu smo izbrane podatke in vrednosti v njihovih šifrantih umestili v razviti semantični referenčni sistem UOPGR, glede na njihovo tematiko in semantično globino.

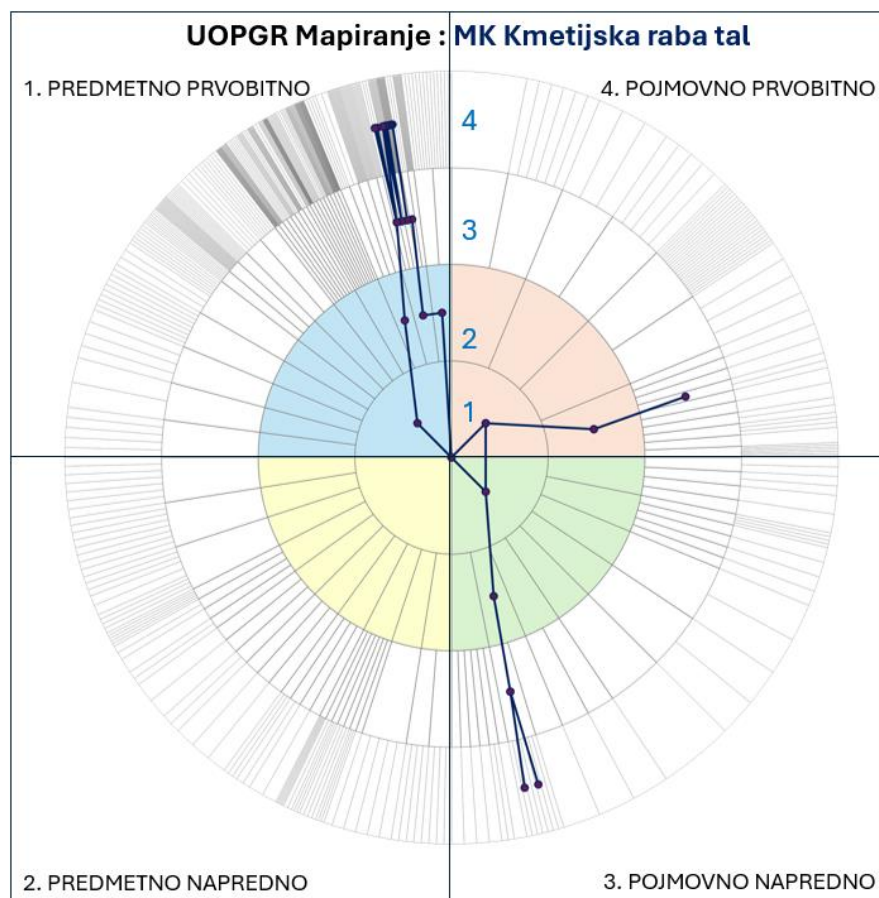
Trenutni semantični referenčni sistem UOPGR se deli na 4 kategorije, ki se skupno delijo na 34 razredov, ki se delijo skupno na 105 podrazredov in obsegajo 446 entitet (objektov, shem ali pojavov) v četrti globini. Skupno 590 konceptov.

3.5.4.1 Evidenca kmetijske rabe tal

Preglednica 63 - Umeščenosti vrednosti šifrantov Kmetijske rabe tal v UOPGR s semantično globino

UOPGR	UOPGR OPIS KATEGORIJE	Sem. Globina	RabaID	RABA_OPIS
101	Objekti aktivnosti bivanja	2	3000	Pozidano in sorodno zemljišče (25 m2)
102	Objekti aktivnosti PRIDOBIVANJA surovin in energije	2	3000	Pozidano in sorodno zemljišče (25 m2)
103	Objekti aktivnosti PROIZVODNJE izdelkov (industrijsko/obratno)	2	3000	Pozidano in sorodno zemljišče (25 m2)
10202	Pridobivanje vode	3	7000	Voda
10203	Zemljišča/objekti za pridobivanje rastlinskih surovin	3	1240	Trajni nasadi
10203	Zemljišče/objekt za pridobivanje zivalskih surovin	3	1600	Neobdelano kmetijsko zemljišče (1000 m2)
10205	Zemljišče/objekt za pridobivanje lesa	3	1800	Kmetijsko zemljišče, poraslo z gozdnim drevjem (1000 m2)
40107	Opazovanje naravnih okolij - habitatov	3	4100	Barje (5000 m2)
40107	Opazovanje naravnih okolij - habitatov	3	4210	Trstičje (5000 m2)
40107	Opazovanje naravnih okolij - habitatov	3	4220	Ostalo zamočvirjeno zemljišče (5000 m2)
1020301	njiva	4	1100	Njiva (1000 m2)
1020303	hmeljišče	4	1160	Hmeljišče (500 m2)
1020304	oljčnjak	4	1230	Oljčnik (500 m2)
1020305	sadovnjak	4	1221	Intenzivni sadovnjak (1000 m2)
1020305	sadovnjak	4	1222	Ekstenzivni oziroma travniški sadovnjak (1000 m ²)
1020306	vinograd	4	1211	Vinograd (500 m ²)

1020306	vinograd	4	1212	Matičnjak (500 m ²)
1020310	greda topla	4	1190	Rastlinjak (25 m ²)
1020311	drevesnica	4	1180	Trajne rastline na njivskih površinah (1000 m ²)
1020312	vrtnarija	4	1180	Trajne rastline na njivskih površinah (1000 m ²)
1020401	pašnik	4	1300	Trajni travnik (1000 m ²)
1020404	travnik	4	1300	Trajni travnik (1000 m ²)
1020404	travnik	4	1321	Barjanski travnik (1000 m ²)
1020501	gozd	4	1500	Drevesa in grmičevje (1000 m ²)
1020501	gozd	4	2000	Gozd (2500m ²)
1020502	plantaža	4	1420	Plantaža gozdnega drevja (1000 m ²)
3020102	enota varovanja vodnih virov	4	4210	Trstičje (5000 m ²)
3020105	biotop	4	4220	Ostalo zamočvirjeno zemljišče (5000 m ²)
3020102	enota varovanja vodnih virov	4	7000	Voda (25 m ²)
3020105	biotop	4	4210	Trstičje (5000 m ²)



Slika 24 - Graf umeščenosti vrednosti šifrantov Kmetijske rabe tal v UOPGR

V evidenci rabe kmetijskih zemljišč smo identificirali 20 objektov, ki ustrezajo entitetam na 4. semantični globini UOPGR, določene tipe, pa smo zaradi presplošnega opisa uvrstili samo v 7 podrazredov na 3. ravni (glej Slika 24)

Umeščanje smo izvedli glede na trenutno raven razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov bodo te povezave lahko bolj številčne in bolj bogate in s tem doprinesle k več možnostim povezovanja.

Na grafu je razvidno, da se večina tipov rabe v osnovi umešča med Predmetne objekte Prvobitnih človeških aktivnosti (1. Kategorija). Določene tipe pa lahko obravnavamo tudi kot bolj kompleksne objekte na področjih Pojemovnih shem prvobitnih in naprednih aktivnostih (4 in 3 kategorija).

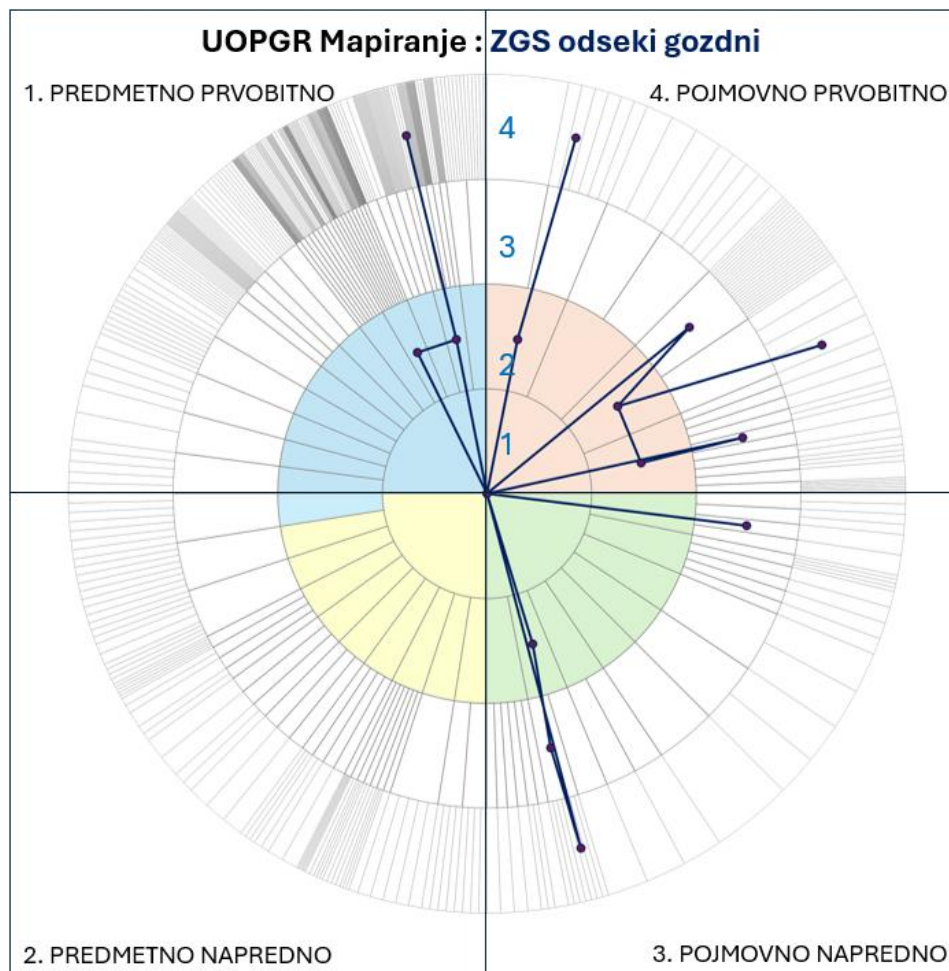
3.5.4.2 Umestitev Gozdarske zbirke podatkov Zavoda za gozdove v UOPGR

3.5.4.2.1 Gozdni Odseki

Preglednica 64 - Umeščenosti vrednosti šifrantov sloja odseki gozdni ZGS v UOPGR s semantično globino

UOPGR	UOPGR OPIS KATEGORIJE	Sem. Globina	Feature	Description
104	Objekti aktivnosti PRENOSA (TRANSPORTA)	2	spravilo	šifra vrste spravila
104	Objekti aktivnosti PRENOSA (TRANSPORTA)	2	spravilo_naziv	vrsta spravila
104	Objekti aktivnosti PRENOSA (TRANSPORTA)	2	razdalja	razdalja spravila [m]
30201	Varovanje NARAVNIH virov	3	carb_tot_c	skupni vezani ogljik [t/ha]
30201	Varovanje NARAVNIH virov	3	ponor_c	letni ponor ogljika [t/ha/leto]
30201	Varovanje NARAVNIH virov	3	odprtost	delež odprte površine odseka [%]
30802	Sheme aktivnosti določanja OBMOČIJ POMEMBNIH ZA POŽARNO VARNOST	3	pozar	šifra požarne ogroženosti (povzeta po območnem načrtu, na podlagi katerega je izdelan GGN GGE)
30802	Sheme aktivnosti določanja OBMOČIJ POMEMBNIH ZA POŽARNO VARNOST	3	pozar_naziv	naziv požarne ogroženosti (povzeta po območnem načrtu, na podlagi katerega je izdelan GGN GGE)
40106	Opazovanje mineralov	3	kamnina	šifra vrste kamnine
40106	Opazovanje mineralov	3	kamnina_naziv	vrsta kamnine
40106	Opazovanje mineralov	3	kamnit	delež kamnitosti [%]
40106	Opazovanje mineralov	3	skalnat	delež skalovitosti [%]
40202	Vzpostavljanje geometričnih shem v prostoru geografskih razseznosti	3	polpokr	šifra položaja v pokrajini
40202	Vzpostavljanje geometričnih shem v prostoru geografskih razseznosti	3	polpokr_naziv	položaj v pokrajini
40202	Vzpostavljanje geometričnih shem v prostoru geografskih razseznosti	3	relief	šifra reliefa
40202	Vzpostavljanje geometričnih shem v prostoru geografskih razseznosti	3	relief_naziv	relief
40202	Vzpostavljanje geometričnih shem v prostoru geografskih razseznosti	3	nvod	spodnja nadmorska višina gozda [m]
40202	Vzpostavljanje geometričnih shem v prostoru geografskih razseznosti	3	nvdo	zgornja nadmorska višina gozda [m]
1020501	gozd	4	katgozd	šifra kategorije gozda
3020105	BIOTOP	4	katgozd_naziv	kategorija gozda

3020105	BIOTOP	4	ohranjen	šifra ohranjenosti gozda
3020105	BIOTOP	4	ohranjen_naziv	ohranjenost gozda
3020105	BIOTOP	4	grt1	šifra prevladujočega gozdnega rastiščnega tipa
3020105	BIOTOP	4	grt1_naziv	prevladujoč gozdni rastiščni tip
4020101	toponim	4	krajime	krajevno ime
4020219	naklon	4	lega	šifra lege
4020219	naklon	4	lega_naziv	Lega
4020219	naklon	4	nagib	nagib terena [°]
4040105	Gospodarjenje z LESOM	4	tarifa_sm	tarifa za skupino drevesnih vrst smreka [%]
4040105	Gospodarjenje z LESOM	4	tarifa_je	tarifa za skupino drevesnih vrst jelka [%]
4040105	Gospodarjenje z LESOM	4	tarifa_oi	tarifa za skupino drevesnih vrst ostali iglavci [%]
4040105	Gospodarjenje z LESOM	4	tarifa_bu	tarifa za skupino drevesnih vrst bukev [%]
4040105	Gospodarjenje z LESOM	4	tarifa_hr	tarifa za skupino drevesnih vrst hrasti [%]
4040105	Gospodarjenje z LESOM	4	tarifa_pl	tarifa za skupino drevesnih vrst plemeniti listavci [%]
4040105	Gospodarjenje z LESOM	4	tarifa_tl	tarifa za skupino drevesnih vrst trdi listavci [%]
4040105	Gospodarjenje z LESOM	4	tarifa_ml	tarifa za skupino drevesnih vrst mehki listavci [%]
4040105	Gospodarjenje z LESOM	4	odprt_gurs	odprtost gozdov modificirana za potrebe GURS
4040105	Gospodarjenje z LESOM	4	intgosp	šifra intenzivnosti gospodarjenja
4040105	Gospodarjenje z LESOM	4	intgosp_naziv	kategorija intenzivnosti gospodarjenja
4040105	Gospodarjenje z LESOM	4	rk	rastiščni koeficient prevladujočega GRT*
4040105	Gospodarjenje z LESOM	4	rk_gurs	rastiščni koeficient GRT z najnižjim rastiščnim koeficientom
4040105	Gospodarjenje z LESOM	4	odprtost	delež odprte površine odseka [%]



Slika 25 - Graf umeščenosti vrednosti šifrantov sloja odseki gozdni ZGS v UOPGR

V podatkovnem sloju odseki gozdni, smo od 42 opisov identificirali 24 opisov iz šifrantov, ki ustrezajo entitetam na 4. semantični globini UOPGR. Zaradi presplošnega opisa smo uvrstili 15 opisov v šifrantih v podrazrede na 3. globini in 3 opise v razred na 2. globini. (glej Slika 25).

Umeščanje smo izvedli glede na trenutno raven razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov, bodo te povezave lahko bolj številčne in bolj bogate in s tem doprinesle k več možnostim povezovanja.

Na grafu je razvidno, da se opisov iz šifrantov umešča med 1. 3. in 4. kategorijo.

Veliko število konceptov s semantično globino 2. ali 3. nakazuje, da je na teh področjih prostor za nadgradnjo referenčnega modela.

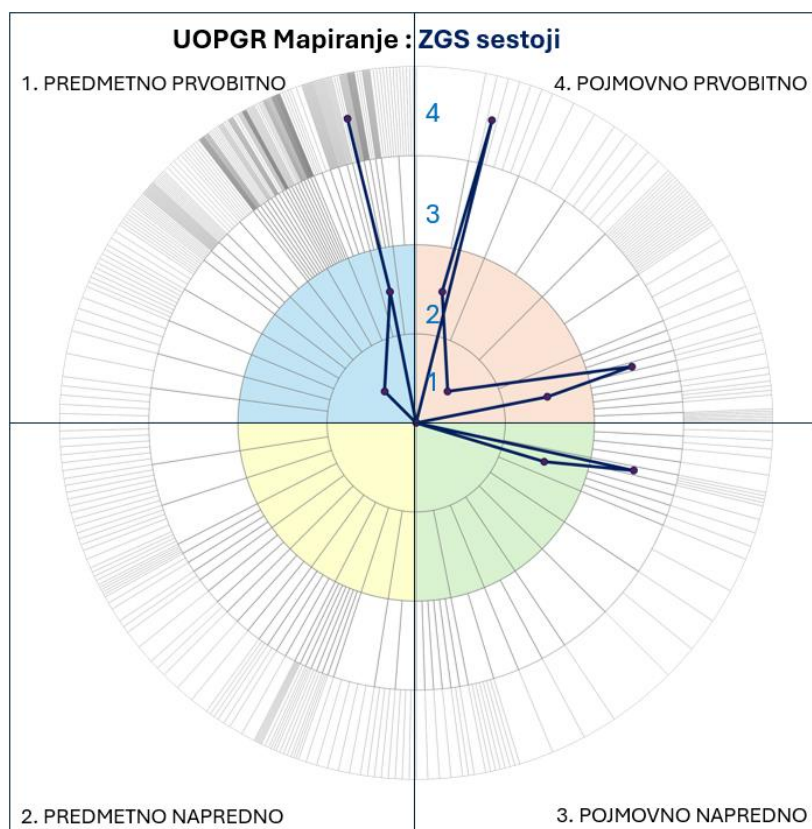
3.5.4.2.2 Gozdni Sestoji

Preglednica 65 - Umeščenosti vrednosti šifrantov sloja sestoji GZS v UOPGR s semantično globino

UOPGR	UOPGR OPIS KATEGORIJE	Sem. Globina	Sestoji Feature	Sestoji Opis
30705	Sheme aktivnosti DOLOČANJA MEJA OBMOČIJ UPRAVLJANJA VIROV	3	ggo	šifra gozdnogospodarskega območja
30705	Sheme aktivnosti DOLOČANJA MEJA OBMOČIJ UPRAVLJANJA VIROV	3	ggo_naziv	naziv gozdnogospodarskega območja
30705	Sheme aktivnosti DOLOČANJA MEJA OBMOČIJ UPRAVLJANJA VIROV	3	odsek	šifra odseka
30705	Sheme aktivnosti DOLOČANJA MEJA OBMOČIJ UPRAVLJANJA VIROV	3	sestoj	šifra sestoja
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv11	delež skupine drevesnih vrst smreka [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv11_m	delež skupine drevesnih vrst smreka v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv21	delež skupine drevesnih vrst jelka [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv21_m	delež skupine drevesnih vrst jelka v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv30	delež skupine drevesnih vrst bor [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv30_m	delež skupine drevesnih vrst bor v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv34	delež skupine drevesnih vrst macesen [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv34_m	delež skupine drevesnih vrst macesen v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv39	delež skupine drevesnih vrst ostali iglavci [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv39_m	delež skupine drevesnih vrst ostali iglavci v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv41	delež skupine drevesnih vrst bukev [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv41_m	delež skupine drevesnih vrst bukev v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv50	delež skupine drevesnih vrst hrasti [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv50_m	delež skupine drevesnih vrst hrasti v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv60	delež skupine drevesnih vrst plemeniti listavci [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv60_m	delež skupine drevesnih vrst plemeniti listavci v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv70	delež skupine drevesnih vrst trdi listavci [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv70_m	delež skupine drevesnih vrst trdi listavci v mladju [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv80	delež skupine drevesnih vrst mehki listavci [%]
40107	Opazovanje naravnih okolij - habitatov	3	lzkdv80_m	delež skupine drevesnih vrst mehki listavci v mladju [%]

4040105	Gospodarjenje z LESOM	4	rfaza	šifra razvojne faze
4040105	Gospodarjenje z LESOM	4	rfaza_naziv	razvojna faza
4040105	Gospodarjenje z LESOM	4	sksmerni	šifra gozdnogojitvene smernice
4040105	Gospodarjenje z LESOM	4	sksmerni_naziv	gozdnogojitvena smernica
4040105	Gospodarjenje z LESOM	4	zasnova	šifra zasnove sestoja
4040105	Gospodarjenje z LESOM	4	zasnova_naziv	zasnova sestoja
4040105	Gospodarjenje z LESOM	4	sklep	šifra sestojnega sklepa
4040105	Gospodarjenje z LESOM	4	sklep_naziv	sestojni sklep
4040105	Gospodarjenje z LESOM	4	negovan	šifra negovanosti sestoja
4040105	Gospodarjenje z LESOM	4	negovanost_naziv	negovanost sestoja
4040105	Gospodarjenje z LESOM	4	pompov	površina pomladka v sestoju [ha]
4040105	Gospodarjenje z LESOM	4	pomzas	zasnova pomladka v sestoju
1020501	gozd	4	lzigl	lesna zaloga iglavcev v sestoju [m3]
1020501	gozd	4	lzlst	lesna zaloga listavcev v sestoju [m3]
1020501	gozd	4	lzsku	skupna lesna zaloga v sestoju [m3]
1020501	gozd	4	prigl	prirastek iglavcev v sestoju [m3]
1020501	gozd	4	prlst	prirastek listavcev v sestoju [m3]
1020501	gozd	4	prsku	skupni prirastek v sestoju [m3]
1020501	gozd	4	etigl	možni posek iglavcev v sestoju [m3]
1020501	gozd	4	etlst	možni posek listavcev v sestoju [m3]
1020501	gozd	4	etsku	skupni možni posek v sestoju [m3]
1020501	gozd	4	lzskdv11	delež skupine drevesnih vrst smreka [%]
1020501	gozd	4	lzskdv11_m	delež skupine drevesnih vrst smreka v mladju [%]
1020501	gozd	4	lzskdv21	delež skupine drevesnih vrst jelka [%]
1020501	gozd	4	lzskdv21_m	delež skupine drevesnih vrst jelka v mladju [%]
1020501	gozd	4	lzskdv30	delež skupine drevesnih vrst bor [%]
1020501	gozd	4	lzskdv30_m	delež skupine drevesnih vrst bor v mladju [%]
1020501	gozd	4	lzskdv34	delež skupine drevesnih vrst macesen [%]
1020501	gozd	4	lzskdv34_m	delež skupine drevesnih vrst macesen v mladju [%]
1020501	gozd	4	lzskdv39	delež skupine drevesnih vrst ostali iglavci [%]
1020501	gozd	4	lzskdv39_m	delež skupine drevesnih vrst ostali iglavci v mladju [%]
1020501	gozd	4	lzskdv41	delež skupine drevesnih vrst bukev [%]
1020501	gozd	4	lzskdv41_m	delež skupine drevesnih vrst bukev v mladju [%]

1020501	gozd	4	lzskdv50	delež skupine drevesnih vrst hrasti [%]
1020501	gozd	4	lzskdv50_m	delež skupine drevesnih vrst hrasti v mladju [%]
1020501	gozd	4	lzskdv60	delež skupine drevesnih vrst plemeniti listavci [%]
1020501	gozd	4	lzskdv60_m	delež skupine drevesnih vrst plemeniti listavci v mladju [%]
1020501	gozd	4	lzskdv70	delež skupine drevesnih vrst trdi listavci [%]
1020501	gozd	4	lzskdv70_m	delež skupine drevesnih vrst trdi listavci v mladju [%]
1020501	gozd	4	lzskdv80	delež skupine drevesnih vrst mehki listavci [%]
1020501	gozd	4	lzskdv80_m	delež skupine drevesnih vrst mehki listavci v mladju [%]



Slika 26 - Graf umeščenosti vrednosti šifrantov sloja sestoji ZGS v UOPGR

V podatkovnem sloju odseki gozdni, smo od 55 opisov identificirali 31 opisov iz šifrantov, ki ustrezajo entitetam na 4. semantični globini UOPGR. Zaradi presplošnega opisa smo uvrstili 24 opisov v šifrantih v podrazrede na 3. globini. (glej Slika 26).

Umeščanje smo izvedli glede na trenutno raven razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov, bodo te povezave lahko bolj številčne in bolj bogate in s tem doprinesle k več možnostim povezovanja.

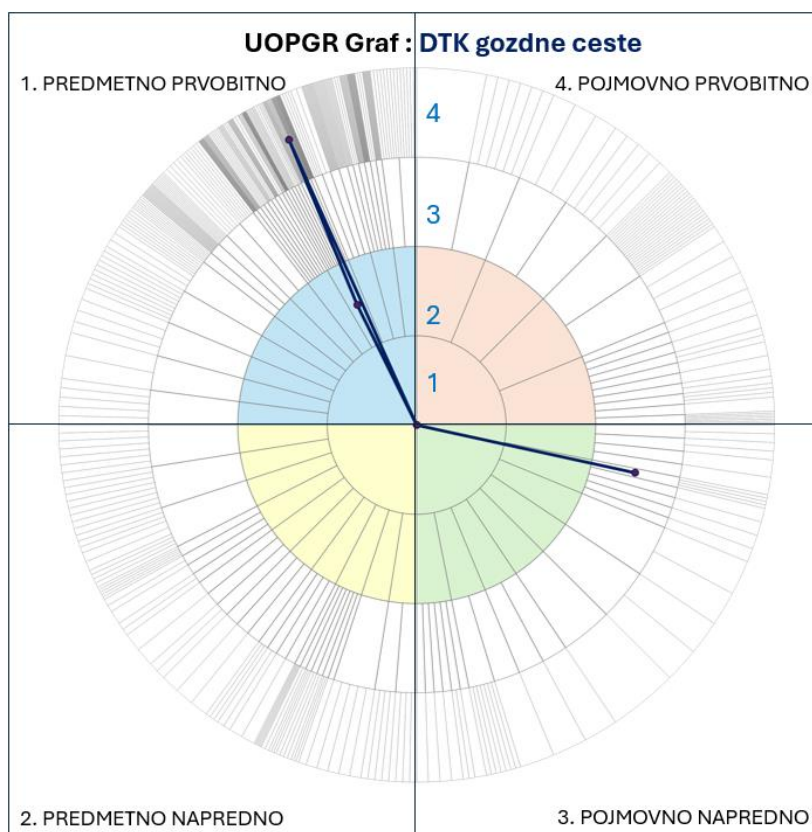
Na grafu je razvidno, da se opisov iz šifrantov umešča med 1. 3. in 4. kategorijo.

Veliko število konceptov s semantično globino 2 ali 3 nakazuje, da je na teh področjih prostor za nadgradnjo referenčnega modela.

3.5.4.2.3 Gozdne Ceste

Preglednica 66 - Umeščenosti vrednosti šifrantov sloja gozdne Ceste ZGS v UOPGR s semantično globino

UOPGR	UOPGR OPIS KATEGORIJE	Sem. Globina	Sestoji Feature	Sestoji Opis
30705	Sheme aktivnosti DOLOČANJA MEJA OBMOČIJ UPRAVLJANJA VIROV	3	ggo_naziv	naziv gozdnogospodarskega območja
30705	Sheme aktivnosti DOLOČANJA MEJA OBMOČIJ UPRAVLJANJA VIROV	3	ggo_naziv	naziv gozdnogospodarskega območja
1040301	steza	4	cesta_sifra	šifra gozdne ceste
1040301	steza	4	cesta_naziv	naziv gozdne ceste



Slika 27 - Graf umeščenosti vrednosti šifrantov sloja gozdnih Cest ZGS v UOPGR

V podatkovnem sloju odseki gozdni, smo od 4 opisov identificirali 2 opisa iz šifrantov, ki ustrezata entitetam na 4. semantični globini UOPGR. Zaradi presplošnega opisa smo 2 opisa uvrstili v podrazred na 3. globini. (glej Slika 27).

Umeščanje smo izvedli glede na trenutni ravni razvoja modela in glede na svojo interpretacijo povezljivosti z ostalimi področji. S povezovanjem različnih domenskih ekspertov bodo te povezave lahko bolj številčne in bolj bogate in s tem prispevale k več možnostim povezovanja.

Na grafu je razvidno, da se opisov iz šifrantov umešča med 1. 3. kategorijo ter, da ima zelo skromno semantično širino.

V naslednjem poglavju smo izvedli skupni povezovalni test, kjer smo uporabili novo razvite mere dostopnosti do semantičnih elementov, ter evidence primerjali v referenčnem sistemu.

3.6 Skupni povezovalni test

V skupnem povezovalnem povezovalnem testu smo uporabili novo razvite mere dostopnosti do semantičnih elementov, ter obravnavane evidence primerjali v referenčnem sistemu UOPGR.

3.6.1 Opredelitev slojev za primerjavo (DTK Pokrovnost tal, MK RABA, GZS Sestoji)

V tem testu smo izbrali podatkovne sloje iz različnih evidenc, ki prikazujejo objekte uporabljene za določanje pokrovnosti tal, oziroma njihove dejanske rabe. Izbrali smo sloje, ki smo jih že analizirali v prejšnjem poglavju.

Preglednica 67 - Seznam slojev za skupni povezovalni test

Ime sloja	Opis	Vir
RABA_2024_09_30.shp	Grafični podatki evidence dejanske rabe kmetijskih zemljišč	https://rkg.gov.si/vstop/
odseki gozdni (wfs)	Grafični podatki gozdnih odsekov	https://prostor.zgs.gov.si/pregledovalnik/
Sestoji (wfs)	Grafični podatki sestojev	https://prostor.zgs.gov.si/pregledovalnik/
LU_PokritostTal.shp	Podatkovni sloj Pokritost tal (Linija), objektnega področja POKRITOST TAL	https://ipi.eprstor.gov.si/jgp/data

3.6.2 Primerjava mer dostopnosti semantičnih elementov

V tem delu analize smo primerjali dostopnost semantičnih elementov glede na razvite mere dostopnosti za Kmetijsko rabo tal, DTK Pokritost tal in GZS – Odseki gozd.

Preglednica 68 - Primerjava mer dostopnosti semantičnih elementov

	Kmetijska raba tal			DTK Pokritost tal			GZS - Odseki Gozd		
	Evidenca		Šifra	Evidenca		Šifra	Evidenca		Šifra
Tehnična dostopnost vira pomena (Sintaktična dostopnost)	Lokalni Standardizirani strukturirani viri v digitalni obliki	5	Pisni viri v digitalni obliki	3	Spletno dostopni standardizirani servisi	6	Pisni viri v digitalni obliki	3	Pisni viri v digitalni obliki
Organizacijska dostopnost vira	Opisi so del dokumentacije evidence v vsebinski domeni (strokovna javnost)	3	Opisi so del dokumentacije evidence Organizacije	2	Splošno mednarodno dostopna in sprejeta vsebina	5	Opisi so del dokumentacije evidence Organizacije	2	Opisi so del dokumentacije evidence Organizacije
Stopnja uradnosti	Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti..., Pravna praksa (precedenčne sodbe))	6	opisi so del besedil znanstvenih in strokovnih knjig, učbenikov, priročnikov ter tehničnih načrtov, risb in skic	3	Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti..., Pravna praksa (precedenčne sodbe))	6	Opisi so del pravnih podlag (Zakon, pravilnik, uredba, utemeljitev pravnih podlag, akti..., Pravna praksa (precedenčne sodbe))	4	opisi so del besedil znanstvenih in strokovnih knjig, učbenikov, priročnikov ter tehničnih načrtov, risb in skic
Semantična bogatost vira	odstavke	3	podroben opis s primeri, skicami ali slikovnim gradivom	5	odstavke	3	podroben opis s primeri, skicami s kotiranjem ali slikovnim gradivom	6	odstavke
Starost vira	2019 (trenutno veljavno)	5	2013 (trenutno veljavno)	5	Trenutno veljavno	5	do 3 leta	4	2019
TOČK		22		18		25		19	
DELEŽ(%)		76		62		86		66	

Iz vrednosti v tabeli smo ocenili, da so viri za Kmetijsko rabo in za Odseke Gozd (evidenca ZGS) v večini kriterijev manj dostopni, kot podatki iz topografske zbirke. Čeprav so vsebinsko bolj podrobni.

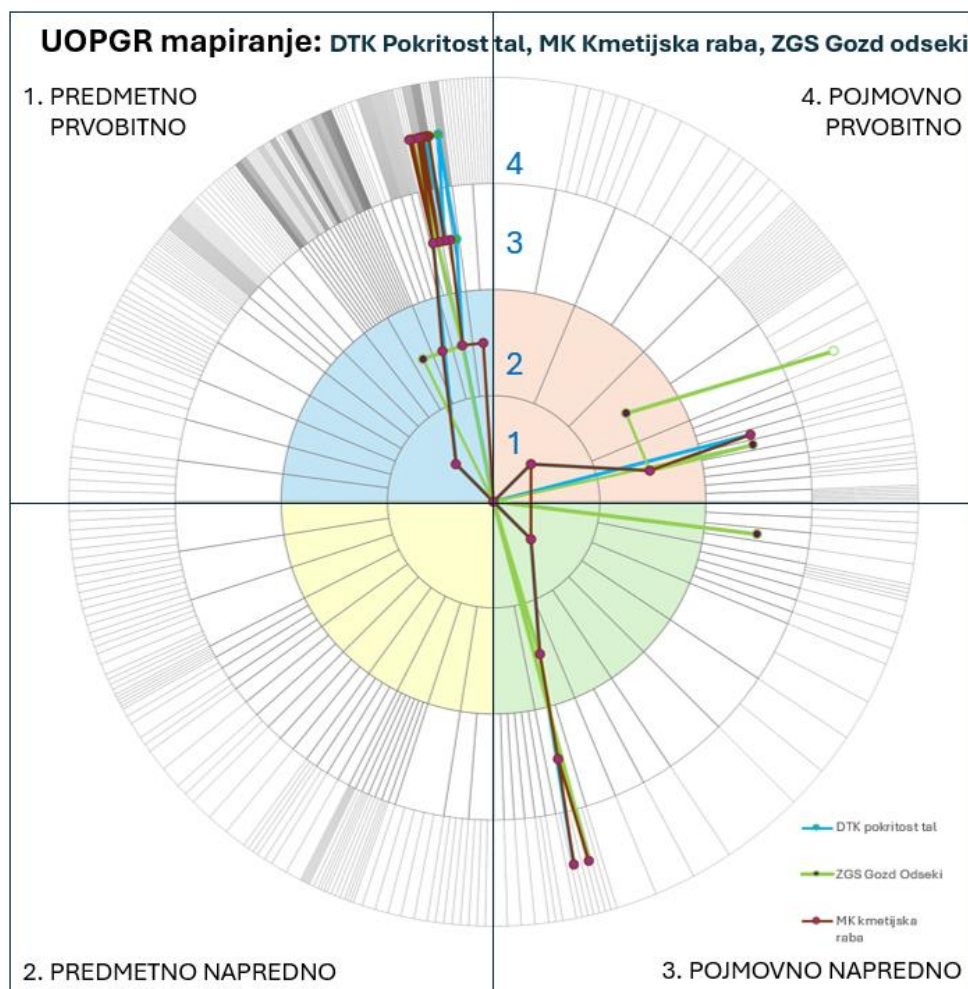
3.6.3 Primerjava skozi referenčni okvir UOPGR

Dodatno smo podatke primerjali še preko novo razvitega orodja referenčnega semantičnega okvirja.

3.6.3.1 Evidentiranje skupnih konceptov

Preglednica 69 - Evidentiranje skupnih konceptov po UOPGR za sloje tematike pokrovnosti oziroma rabe tal

UOPGR	UOPGR OPIS KATEGORIJE	DTK pokritost tal	MK kmetijska raba	ZGS Gozd Odseki
101	Objekti aktivnosti BIVANJA		X	
102	Objekti aktivnosti PRIDOBIVANJA surovin in energije		X	
103	Objekti aktivnosti PROIZVODNJE izdelkov (industrijsko/obrtno)		X	
10202	Pridobivanje vode		X	
10203	Zemljišca/objekti za pridobivanje rastlinskih surovin		X	
10205	Zemljišce/objekt za pridobivanje lesa		X	
40107	Opazovanje naravnih okolij - habitatov	X	X	
1020301	njiva		X	
1020303	hmeljišce	X	X	
1020304	olcnjak	X	X	
1020305	sadovnjak	X	X	
1020306	vinograd	X	X	
1020310	greda		X	
1020311	drevesnica		X	
1020312	vrtnarija		X	
1020401	pasnik		X	
1020404	travnik		X	
1020501	gozd	X	X	X
1020502	plantaza		X	
3020102	VODA	X	X	
3020105	BIOTOP		X	X
104	Objekti aktivnosti PRENOSA (TRANSPORTA)			X
30201	Varovanje NARAVNIH virov			X
30802	Sheme aktivnosti določanja OBMOČIJ, POMEMBNIH ZA POŽARNO VARNOST			X
40106	Opazovanje mineralov			X
40202	Vzpostavljanje geometričnih shem v prostoru geografskih razsežnosti			X
4020101	toponim			X
4020219	naklon			X
4040105	Gospodarjenje z LESOM			X
30705	Sheme aktivnosti DOLOČANJA MEJA OBMOČIJ UPRAVLJANJA VIROV			



Slika 28 - Graf umeščenosti vrednosti šifrantov slojev tematike rabe tal v UOPGR

Glede na gostoto porazdelitev skupnih konceptov, smo iz zgornje tabele in grafa (glej Slika 28) ugotovili, da je evidenca Kmetijska raba MK, glede na število konceptov UOPGR semantično bolj bogata, kot DTK Pokrovnost tal ali Odseki Gozd. Kmetijska raba pokriva vse koncepte, ki jih pokriva sloj DTK in še več, čeprav ima glede na prejšnjo analizo mer dostopnosti nižjo raven dostopnosti evidence kot šifrantov.

Iz tega lahko sklepamo, da bi bilo smiselno tehnološko in z dodatnimi opisi povečati dostopnost do semantičnih elementov Kmetijske rabe, saj ima visoko semantično vrednost.

3.7 Zaključek

V poglavju 3 smo opravili analizo semantične povezljivosti štirih skupin podatkovnih zbirk. Analizirali smo podatke v topografski bazi, nepremičninske zbirke IS Kataster, zbirnega katastra gospodarske javne infrastrukture ter podatkov v kmetijski in gozdarski domeni. Analiza temelji na več kriterijih, kot so sintaktična povezljivost, dokumentacija kakovosti podatkov, prisotnost semantičnih elementov in glede na različne dostopnosti teh elementov. Poleg tega dokument umešča podatkovne zbirke v razviti skupni semantični okvir UOPGR, s čimer ocenjuje njihovo skladnost ter povezljivost na semantični ravni.

3.7.1 Ključne ugotovitve o semantični povezljivosti analiziranih zbirk

Na ravni zbirke Topografskih podatkov, smo za analizo izbrali sloje za stavbe, ceste in pokrovnost tal. Le ti so na voljo v strojno berljivih formatih, vendar pogosto na osnovni ravni. Medtem ko so metapodatki na voljo v formatu RDF, in šifranti dostopni v formatih JSON ali CSV, je podrobna dokumentacija pogosto omejena in dostopna le na zahtevo. Podatki so opremljeni z edinstvenimi identifikatorji, kar omogoča povezovanje s podatki drugih nacionalnih evidenc. Tehnični potencial za semantično povezovanje je prisoten, vendar zahteva predhodno pripravo podatkov.

Nepremičninski podatki so visoko standardizirani na ravni evidenc in so skladni z direktivami INSPIRE. Na voljo so v strojno berljivih formatih (npr. XML, JSON), vendar posamezni sloji in pripadajoči šifranti pogosto zanemarijajo napredne semantične koncepte. Identificiranih je bilo več kot 60 šifrantov, med katerimi so nekateri ustrezno semantično opisani in semantično bogati, določenim pa bi predlagali izboljšave. Podatki vključujejo edinstvene nacionalne identifikatorje, kar omogoča povezovanje med podatki. Vsebinsko je semantična povezljivost na visokem ravni, vendar pogosto ni javno dostopna prek spleta. Tehnični potencial za semantično povezovanje je prisoten na visoki ravni, vendar zahteva predhodno pripravo podatkov.

GJI podatki so tehnološko dostopni prek standardov INSPIRE in so na voljo v naprednih strojno berljivih formatih, kot so GeoPackage, WMS, WFS in Atom. Metapodatki so na voljo v RDF obliki, vendar je ostala dokumentacija, vključno z informacijami o kakovosti podatkov, večinoma v osnovnih formatih, kot so Word in PDF. Povezovanje z zunanjimi šifranti je kompleksno in zahteva pregled dokumentacije, ki je v osnovni nestrukturirani in neodprtokodni obliki. Podatki vključujejo edinstvene identifikatorje in so dobro dokumentirani, vendar predlagamo bolj sistematičen pristop k semantičnim povezavam.

Podatki o kmetijskih zemljiščih so sintaktično preprosti, vendar predstavljeni v formatih z nizko stopnjo strojne berljivosti. Semantično razumevanje teh podatkov pogosto zahteva analizo in interpretacijo dokumentacije v PDF formatu. Opisi posameznih kategorij rabe zemljišč so ustrezno podani skupaj s primeri, vendar je njihova uporabnost omejena, zaradi formata in neskladij med podatkovnimi modeli prikaza teh podatkov v spletni aplikaciji naproti podatkom za prenos.

Gozdarski podatki so dostopni prek storitev WFS, vendar je dokumentacija omejena na osnovne opise v PDF formatu. Organizacija podatkov je razmeroma preprosta, brez integracije zunanjih šifrantov ali napredne semantične strukture. Edinstveni identifikatorji omogočajo osnovno povezovanje, vendar mehanizmi za integracijo z drugimi viri podatkov niso podani. Dokumentacija o kakovosti podatkov je pomanjkljiva, kar omejuje njihovo uporabnost za napredne analize.

3.7.2 Splošna dostopnost semantičnih elementov:

Na ravni evidenc so podatki Geodetske uprave na splošno dobro podprti z osnovnimi semantičnimi elementi, skladnimi z INSPIRE smernicami. To vključuje metapodatke, ki so pogosto na voljo v strojno berljivem formatu RDF. Žal pri ostalih zbirkah to ni tako. Metapodatkov v strojno berljivi obliki večinoma ni, je pa prosto dostopne precej dokumentacije, na podlagi katere bi z obdelavo te informacije lahko pridobili.

Na ravni posameznih podatkovnih struktur in šifrantov *primanjkuje natančnih in podrobnih opisov* v strojno berljivih formatih, kar otežuje njihovo samodejno interpretacijo. Te informacije pogosto niso javno dostopne, če pa so, so le ti v strojno manj razumljivi obliki (nestrukturirani, neodprtokodni). Podrobnejšo dokumentacijo je pogosto treba zahtevati neposredno pri upravljavcih podatkov.

3.7.3 Predlogi za izboljšanje semantične povezljivosti

- **Izboljšanje strojne berljivosti:** Dokumentacijo, predvsem opise šifrantov in atributov, je treba pretvoriti iz statičnih formatov, kot so PDF in DOCX, v strukturirane in strojno berljive formate, kot so JSON, CSV ali RDF.
- **Standardizacija podatkovnih modelov:** Poenotenje podatkovnih modelov med različnimi dostopnimi točkami, da se zmanjšajo neskladja, ki zavirajo interoperabilnost.
- **Opisovanje šifrantov z izboljšano semantiko:** Šifranti in njihove opisne vrednosti bi morali vsebovati *podrobnejše informacije ter pozitivne in negativne primere uporabe*, da se omogoči lažjo strojno obdelavo.
- **Razvoj ontologij in taksonomij:** Razvoj ali prilagoditev standardnih ontologij in taksonomij za vsestransko semantično strukturiranje podatkovnih zbirk. Razviti referenčni semantični okvir, kot je UOPGR, bi lahko služil kot temelj za te pobude.
- **Izboljšanje dostopnosti dokumentacije:** Ključna dokumentacija, kot so podatki o kakovosti in metodologiji, mora biti na voljo vsem zainteresiranim stranem prek javnih platform.
- **Implementacija principov povezanih podatkov:** *Izkoriščanje unikatnih identifikatorjev za implementacijo principov povezanih podatkov*, kar bi omogočilo boljšo integracijo med podatkovnimi zbirkami.
- **Vključitev semantičnih anotacij:** Semantične anotacije naj se dodajo neposredno k podatkom, kar omogoča samodejno razumevanje njihovega pomena s strani strojnih sistemov. S sistemom za javno komentiranje teh podatkov se lahko bogastvo semantičnega opisa razširi in komentarje uporablja za izboljšave.
- **Uvedba mehanizmov za povezovanje šifrantov:** Jasen opis postopkov za semantično povezovanje geometrijskih slojev z zunanjimi šifranti, zlasti v okviru kompleksnih podatkov, kot so GJI.

Z implementacijo teh ukrepov bi se semantična skladnost in povezljivost prostorskih podatkov v Sloveniji znatno izboljšana, kar bi omogočilo učinkovitejšo integracijo, analizo in uporabo podatkov na vseh ravneh – od tehničnih uporabnikov do končnih analiz. Takšne izboljšave bi prav tako okrepile skladnost z evropskimi in mednarodnimi standardi ter omogočile naprednejše aplikacije, kot so umetna inteligenca in strojno učenje v prostorskih podatkovnih okoljih.

4 Delovni sveženj (DS4): Izdelava usmeritev, vodil za upravljanje procesa semantičnega povezovanja za GURS

Cilj doseganja ponovne uporabe že obstoječih odprtih programskih rešitev iz prve faze ter analize ovir ter predlog rešitev druge faze bosta predstavljala osnova za doseganje cilja enotnih usmeritev za interno in zunanje semantično povezovanje s povečano stopnjo medopravilnosti. Ključni koncepti za doseganje povečanje stopnje medopravilnosti so podatkovne platforme, *podatkovni prostori* in podatkovne tržnice, ki *temeljijo na pomensko – semantično razumljivih podatkih in storitvah*. Za potrebe implementacije umetne inteligence (umetna inteligenca namreč zahteva ogromno množico podatkov za učenje) je treba zagotoviti večjo razpoložljivost in usklajenost obsežnih in raznovrstnih podatkovnih ekosistemov. Podatkovne pobude EU, kot so Direktiva o odprtih podatkih, Nabori podatkov visoke vrednosti (HVD), Globalni statistični geoprostorski okvir Evrope (GSGF-E), Evropski okvir interoperabilnosti (EIF) in Infrastruktura geoprostorskega znanja (GKI) obravnavata izzive v zvezi z integracijo geoprostorskih podatkov.

4.1 Semantična kakovost poslovnih pravil pogojuje semantično kakovost podatkov

Shranjeni podatki so zgolj sistemski ali podatkovni *ostanek dogodkov, ki so se že zgodili*, spomin na pretekle dogodke. Za resnično odpravo težav s "kakovostjo podatkov" je potrebna obravnava vidika poslovanja, poslovnih procesov, premik žarišča od načrtovanja podatkov ali čiščenja podatkov, na *dejavnost in znanje o organizaciji sami* (Ross, 2024).

V obeh primerih običajno obstaja *skupen kontekst*, znotraj katerega se lahko *razlaga, interpretira pomen sporočil*, pa tudi bolj ali manj *izmenjava pojasnil* v realnem času.

Podatki niso samo podatki; to je sredstvo za komunikacijo. *Dejanje ustvarjanja podatkov* je dejanje *ustvarjanja sporočila ljudem v prihodnosti*; skoraj zagotovo ne v komunikaciji s prejemniki sporočila neposredno, »iz oči v oči« ali v povezavi v živo prek interaktivnega omrežja. V tem smislu je *ustvarjanja sporočila ljudem v prihodnosti, torej ustvarjanje podatkov, slepa komunikacija*.

Posledično morajo podatki, ki jih nekdo ustvari, dobesedno govoriti sami zase, sami pojasnjevati svoj pomen.

Tipični ukrepi zagotavljanja kakovosti podatkov se osredotočajo le na kakovost vsebine shrambe podatkov, namesto da bi se osredotočili predvsem na *semantično kakovost namenjenih poslovnih komunikacij*. Običajni parametri ocenjevanja kakovosti podatkov so na primer popolnost, edinstvenost, pravočasnost in podobno. Po svoji naravi so takšni parametri retroaktivni (zagledani v preteklost), namesto proaktivni, ter obravnavajo podatke z IT systemskega vidika ne pa semantično komunikacijskega, poslovnega vidika.

Poudarek pri ustvarjanju kakovostnih podatkov mora biti na *učinkovitosti komuniciranja* – torej na *pomenski, semantični kakovosti*. Pred izvedbo projektov semantičnega povezovanja podatkov je treba ugotoviti, kateri načini ustvarjanja podatkov lahko zagotavljajo *dobro poslovno komunikacijo*.

Ocenjevanje semantične kakovosti podatkov se osredotoča na to, *ali je pomen sporočila, podatka jasn*. Če pomen podatka, semantika ni jasna, poslovne komunikacije ni mogoče pravilno razumeti. Z drugimi besedami, *Za sporočane miselne zasnove, koncepte* je treba zagotoviti *jasnost pomena uporabljenih besed, simbolov* – ne samo kakovost podatkov. Torej za zagotavljanje semantične kakovosti podatkov ni dovolj le pravilno oblikovanje podatkov (sintaksa) in uporaba podatkovnih omejitev (kontrol).

Kakovost podatkov v podatkovni shrambi nikoli ne more biti boljša od *kakovosti poslovnega komuniciranja, ki proizvaja podatke*. Dogaja se, da je kakovost podatkov visoko ocenjena, čeprav je

bilo poslovno komuniciranje, ki je proizvedlo podatke, slabo: zmedeno, protislovno, nerazumljivo ali drugače neučinkovito. Visoka ocena kakovosti podatkov ob slabi komunikaciji je torej nesmiselna.

Če želimo zagotoviti visokokakovostne podatke je treba zagotoviti proaktivne ukrepe za oblikovanje visokokakovostnih sporočil, namesto, da bi se osredotočali na že ustvarjene in oblikovane podatke – ne glede na to, ali gre za strukturirane podatke (v bazah podatkov) ali nestrukturirane podatke iz pisne poslovne komunikacije.

Tako imenovana »slepa komunikacija«, s katero *se ustvarjajo sporočila, podatki ljudem v prihodnosti*, mora zagotavljati naslednje lastnosti:

- Zelo kakovostni *dokazi* (evidence) o tem, *kaj vsebina pomeni*.
- *Vsebina* dokazov *predstavlja točno resničnost*.
- *Nepotrebnost dodatnih predpostavk* za dopolnjevanje dokazov.

Ključne oblike dokazov za prejemnike sporočil oziroma podatkov, so naslednje;

- *besede* (termini, izrazi),
- *opredelitve* terminov (definicije, opisi),
- poslovno *besedišče* (vokabular),
- *pravila poslovanja* (poslovna, snemalna pravila).

Iz ključnih oblik dokazov (sporočil, evidenc) izhajajo ključni parametri semantične, pomenske kakovosti (Slika 29).

Parametri semantične kakovosti dokazov	Opis: »dokaz...«	Pojasnila	Negativni vidiki
BERLJIVOST (radibility)	ni kodirano ali kriptirano	IZRAŽEN V NARAVNEM JEZIKU Pomen ni prikrit z označevalci (terminov/kod/besed). točno pravo ime.	kodiranje , kratice, kriptiranje
RAZUMLJIVOST (understandability)	zlahka razumljivo	Uporabljeni so le termini s TRDNO poslovno OPREDELITVIJO	opredelitve pojmov manjkajo, so nejasne, nenatančne, nepopolne in/ali neposlovne
NATANČNOST (precision)	pomensko stabilno	Uporabljeni so le termini iz poslovnega BESEDNJAKA , osnovani na konceptualnem modelu (poslovni ONTOLOGIJI). Izrazi (vključno s sinonimi) se morajo vedno nanašati samo na en koncept v danem kontekstu.	Izbira besede za neki koncept v sporočilu je enostavno napačna ; Homonimi – včasih beseda pomeni eno, včasih pa drugo; Beseda mogoče obsega širok siv pas pomena
ZANESLJIVOST (reliability)	brez napak , ki bi lahko zmanjšale uporabnost; pravilni podatki nastanejo, ker se poslovna dejavnost izvaja pravilno .	Spoštovana so vsa POSLOVNA PRAVILA: Poslovna pravila se nanašajo na pravilno vodenje ustanove. Če je ustanova pravilno vodena, mora biti njeno poslovno komuniciranje pravilno oblikovano. Če je njeno poslovno komuniciranje pravilno oblikovano , mora biti pravilna tudi vsebina njenih podatkovnih shramb .	Podatkovne/sistemske arhitekture imajo pogosto tehnični pogled na poslovna pravila ; Poslovna pravila se ne nanašajo na podatke , nanašajo na poslovno znanje in poslovno dejavnost ; Pri kakovosti podatkov pravzaprav ne gre za kakovost vaših podatkov , temveč za kakovost vaših poslovnih pravil .

Slika 29 Parametri semantične kakovosti dokazov (evidenc), podatkov, sporočil (povzeto po Ross, 2024)

Usmeritve za uporabo parametrov semantične kakovosti podatkov so navedene v nadaljevanju.

Upoštevajte, da so podatki sporočilo, *komunikacija z drugimi ljudmi v nekem trenutku v prihodnosti*.

Poimenovanje stvari mora vedno temeljiti na načinih komuniciranja o stvareh z besedami v *naravnem jeziku*.

Skrbno se je treba izogibati neustreznim ali zavajajočim imenom ali tistim besedam, ki bi jih zlahka napačno razumeli – izbira besed lahko močno vpliva na končno učinkovitost komunikacije.

Razumljivo sporočilo, za podatke uporablja le izraze, termine, besede s trdnimi poslovnimi opredelitvami (definicijami).

Če opredelitve besed (terminov, izrazov) ni, ali pa so nejasne, nenatančne, nepopolne in/ali neposlovne, zlahka pride do napačne komunikacije. Točno opredeljevanje besed je osrednja veščina za strokovnjake vseh vrst.

Natančno sporočilo je tisto, ki pravilno uporablja opredeljene izraze iz poslovnega besednjaka.

Včasih je izbira besede za določen koncept v sporočilu preprosto napačna. Takšna uporaba je lahko zelo zavajajoča. Še hujša napaka je nedoslednost v rabi (homonimi) – na primer, včasih ista beseda pomeni eno, včasih pa drugo.

Izrazi, besede (vključno s sinonimi) se morajo vedno nanašati samo na en koncept v danem kontekstu. Za to potrebujete čvrst (dobro opredeljen) poslovni besednjak, ki pa zahteva robusten konceptni model (poslovna ontologija).

Zanesljivo sporočilo je tisto, ki je v skladu z vsemi ustreznimi poslovnimi pravili.

Obstaja bistvena razlika med *komuniciranjem v poslovnem smislu* in *komuniciranjem v obliki podatkov*. Poslovna pravila se nanašajo na poslovno *znanje* in poslovno *dejavnost* (aktivnosti), ne na podatke – vsaj ne neposredno.

Strokovnjaki, ki delajo s podatkovnimi/sistemskimi arhitekturami, imajo pogosto tehnični (IT) pogled nanje, kar ni dovolj ustrezno. *Poslovna pravila niso* podatkovna ali sistemska pravila. *Pravo poslovno pravilo* je merilo za vodenje organizacije in njenih procesov.

Poslovna pravila se nanašajo na pravilno vodenje organizacije (procesov). Če je organizacija pravilno vodena, mora biti njeno *poslovno komuniciranje pravilno oblikovano*. Če je njeno poslovno komuniciranje pravilno oblikovano, je pravilna tudi vsebina njegovih podatkovnih zbirk.

Kakovostni podatki so rezultat *konsistentnih poslovnih pravil*. Pomembneje, pravilni podatki nastanejo, ker se poslovna dejavnost sploh izvaja pravilno. Pri kakovosti podatkov pravzaprav *ne gre za kakovost vaših podatkov, temveč bolj za kakovost vaših poslovnih pravil*.

Očitno je, da potrebujete tudi sintaktične kontrole in omejitve, vendar se te le bežno dotikajo *resničnih poslovnih pravil*. O sintaktičnih omejitvah je preprosto govoriti, ker vključujejo vrednosti samo enega podatkovnega polja. Sintaktične omejitve dejansko ne obravnavajo kakovosti podatkov.

Temeljni vzroki za težave s *neustrezno kakovostjo podatkov* ter za napačno komunikacijo v pisni ali drugi poslovni komunikaciji so dejansko semantične narave, v *nekonsistentnosti poslovnih pravil*.

Usmeritev za GURS: Svetujemo, da se izvede preskuse skladnost poslovnih pravil, ki vodijo poslovne procese za posamezne zbirke podatkov, v skladu z Načeli neodvisnosti poslovnih pravil, dosegljivih v Manifestu poslovnih pravil, ki ga je objavila nekomercialna skupina imenovana Business Rules Group leta 2003.

Nabor smernic (RuleSpeak®) za izražanje poslovnih pravil na jedrnat, poslovno prijazen način je namenjen *doseganju večje jasnosti in doslednosti pri sporočanju pravil*, bodisi med poslovneži/eksperti, bodisi kot zahteve za IT sistema. Prav tako je nabor smernic namenjen učinkovitemu zajemanju, izražanju in ohranjanju meril odločanja in poslovnega znanja. RuleSpeak je bil eden od treh referenčnih zapisov, uporabljenih pri ustvarjanju standardne semantike za poslovni besednjak in poslovna pravila (Semantics for Business Vocabulary and Business Rules, SBVR), ki ga je objavila skupina za upravljanje objektov (Object Management Group, OMG) leta 2007. Nabor smernic je popolnoma skladen s to specifikacijo (zadnja različica 1.3, 2015).

Smernice RuleSpeak® so oblikovane v obliki hierarhiije desetih členov in pod členov kot sledi v nadaljevanju (Copyright, 2006–2025. Business Rules Group).

1. člen. **Primarne zahteve, ne sekundarne**

- 1.1. *Poslovna Pravila so prvorazredni elementi postavljanja zahtev.*
- 1.2. *Pravila so bistvenega pomena za poslovne modele in tehnološke modele ter od njih ločen del.*

2. člen. **Pravila naj bodo ločena od procesov, ne vsebovana v njih**

- 2.1. *Pravila so eksplicitne omejitve vedenja in/ali nudijo podporo vedenju.*
- 2.2. *Pravila niso proces in ne postopek. Pravila ne bi smela biti vsebovana v nobenem od teh (procesov ali postopkov).*
- 2.3. *Pravila veljajo za vse procese in postopke. Obstajati mora en usklajen, koheziven sklop pravil, ki se dosledno izvajajo na vseh ustreznih področjih poslovne dejavnosti.*

3. člen. **Namerno znanje, ne stranski proizvod**

- 3.1. *Pravila temeljijo na dejstvih, dejstva temeljijo na konceptih, izraženih z izrazi (besedami).*
- 3.2. *Izrazi (besede) izražajo poslovne koncepte; dejstva podajajo trditve o teh konceptih; pravila omejujejo in podpirajo ta dejstva.*
- 3.3. *Pravila morajo biti izrecna, eksplicitna. O nobenem konceptu ali dejstvu se nikoli ne domneva nobeno pravilo.*
- 3.4. *Pravila so osnova tistega, kar organizacija ve o sebi – to je osnovno poslovno znanje.*
- 3.5. *Pravila je treba negovati, varovati in upravljati.*

4. člen. **Deklarativno, ne postopkovno**

- 4.1. Pravila naj bodo *deklarativno izražena v stavkih v naravnem jeziku* za poslovno občinstvo.
- 4.2. Če nečesa ni mogoče izraziti, potem to ni pravilo.
- 4.3. *Niz stavkov je deklarativen* le, če *niz stavkov nima implicitnega zaporedja*.
- 4.4. Vse *izjave o pravilih*, ki zahtevajo *konstrukcije*, ki niso izrazi in dejstva, *pomenijo predpostavke o implementaciji sistema*.
- 4.5. Pravilo se razlikuje od katerega koli *uveljavljanja*, ki je zanj opredeljeno. *Pravilo in njegovo uveljavljanje sta ločeni zadevi*.
- 4.6. *Pravila je treba opredeliti neodvisno od odgovornosti* za to, kdo, kje, kdaj ali kako jih uveljavlja.
- 4.7. *Izjeme od pravil* so izražene z drugimi pravili.

5. člen. **Dobro oblikovan izraz, ne na hitro izmišljen (ad hoc)**

- 5.1. Poslovna pravila morajo biti *izražena tako*, da jih je *mogoče preveriti glede pravilnosti*.
- 5.2. Poslovna pravila morajo biti izražena tako, da jih je *mogoče med seboj preveriti glede skladnosti*.
- 5.3. Formalne logike, kot je *predikatna logika*, so temeljne *za dobro oblikovano izražanje poslovnih pravil*, pa tudi za tehnologije, ki izvajajo poslovna pravila.

6. člen. **Na pravilih temelječa arhitektura, ne posredna implementacija**

- 6.1. Aplikacija poslovnih pravil je namenoma zgrajena tako, da se prilagaja stalnim spremembam poslovnih pravil. Platforma, na kateri teče aplikacija, bi morala podpirati takšne stalne spremembe.
- 6.2. Neposredno izvajanje pravil – na primer v *mehanizmu pravil* – je boljša implementacijska strategija kot prepis pravil v neko proceduralno obliko.
- 6.3. Sistem poslovnih pravil mora biti *vedno sposoben razložiti sklepanje*, na podlagi katerega pride do sklepov ali ukrepa.
- 6.4. Pravila temeljijo na *resničnih vrednostih (v skladu stvarnimi dejstvi)*. Uporabnikom je skrito, *kako se določi ali vzdržuje resnično-vrednostna pravila*.
- 6.5. *Razmerje med dogodki in pravili* je na splošno *mного proti mnogo*.

7. člen. **Postopki, vodeni s pravili, ne programiranje na podlagi izjem (»razen če«)**

- 7.1. *Pravila določajo mejo* med sprejemljivo in nesprejemljivo poslovno dejavnostjo.

7.2. Pravila pogosto zahtevajo posebno ali selektivno obravnavo zaznanih kršitev. Takšna *dejavnost kršitve pravil* je dejavnost kot vsaka druga dejavnost.

7.3. Da bi zagotovili največjo doslednost in ponovno uporabnost, bi morale biti *obravnavanje nesprejemljive poslovne dejavnosti ločeno od obravnavanja sprejemljive poslovne dejavnosti*.

8. člen. **Zaradi posla, ne zaradi tehnologije**

8.1. Pravila se nanašajo na *poslovno prakso in smernice*; zato so *pravila motivirana s poslovnimi cilji* in so oblikovana z različnimi vplivi.

8.2. Pravila organizacijo vedno nekaj stanejo.

8.3. *Stroške* uveljavljanja pravil je treba *uravnotežiti s poslovnimi tveganji* in poslovnimi priložnostmi, ki bi jih sicer lahko izgubili.

8.4. „Več pravil“ ni boljše. Običajno je *manj, vendar 'dobrih pravil' boljše*.

8.5. Učinkovit sistem lahko temelji na majhnem številu pravil. Dodatna, *bolj diskriminatorna* pravila se lahko *dodajo naknadno* tako, da *sčasoma postane sistem pametnejši*.

9. člen. **Od, s strani in za poslovneže, ne za IT strokovnjake**

9.1. Pravila naj *izhajajo iz dobro obveščenih poslovnježev (ekspertov)*.

9.2. Poslovneži bi morali imeti na voljo *orodja*, ki bi jim pomagala pri *oblikovanju, potrjevanju in upravljanju pravil*.

9.3. Poslovneži bi morali imeti na voljo *orodja*, ki bi jim pomagala *preveriti skladnost poslovnih pravil med seboj*.

10. člen. **Upravljanje poslovne logike, ne platform strojne/programske opreme**

10.1. *Poslovna pravila so ključno poslovno sredstvo*.

10.2. Dolgoročno so *pravila za organizacijo pomembnejša od strojne/programske platforme*.

10.3. Poslovna pravila morajo biti *organizirana in shranjena tako, da jih je mogoče brez težav prenesti na nove platforme strojne/programske opreme*.

10.4. *Pravila in sposobnost njihovega učinkovitega spreminjanja*, so temeljnega pomena za izboljšanje poslovne prilagodljivosti.

Poslovna pravila je mogoče izraziti z uporabo pristopov modeliranja, kot so Unified Modeling Language (UML), z notacijo Business Process Execution Language (BPEL) ali notacijo Business Process

Modeling Notation (BPMN), Odločitveni model in notacija (DMN) ali semantika poslovnega besednjaka in Poslovna pravila (SBVR).

S tem naborom smernic za doseganje večje jasnosti in doslednosti pri sporočanju pravil z namenom zagotavljanja semantične kakovosti podatkov, ki je pogojena s to kakovostjo opredelitve poslovnih pravil, zaključujemo predlog usmeritev v zvezi s pomembnostjo kakovosti poslovnih pravil. Za uspešno upravljanje procesov semantičnega povezovanja je treba upoštevati tudi sektorske, poslovne podatkovne prostore, kar je opisano v nadaljevanju.

4.2 Sektorski podatkovni prostori in podatkovna ekonomija

Podatkovni prostor je mogoče definirati kot *podatkovni ekosistem*, zgrajen *okoli skupno dogovorjenih gradnikov*, ki omogočajo učinkovito in zaupanja vredno izmenjavo podatkov med udeleženci za *ustvarjanje vrednosti* (Data Space Business Alliance). To je *sodelovalni trg za podatke*! Trenutno (2024) je opredeljenih štirinajst skupnih evropskih podatkovnih prostorov.

Cilj EU je *odpraviti obstoječe pravne in tehnične ovire* z ustvarjanjem *vseevropskih skupnih in medopravilnih podatkovnih prostorov* ter *doseči sinergijsko vrednost med podatkovnimi prostori z njihovo povečano povezljivostjo*. Evropska komisija opredeljuje podatkovni prostor kot *ogrodje*, ki *podpira skupno rabo podatkov znotraj podatkovnega ekosistema*. Udeležencem zagotavlja jasno strukturo za skupno rabo, trgovanje in *sodelovanje na podatkovnih sredstvih* na način, ki je v skladu z *ustreznimi zakoni in predpisi* ter zagotavlja pošteno obravnavo za vse vpletene. Cilj EU je tudi zagotoviti *enotni digitalni trg podatkov* in razvoj *podatkovnega gospodarstva*, ki stremi k povečanemu razvoju izdelkov na podlagi podatkov in k povpraševanju po njih.

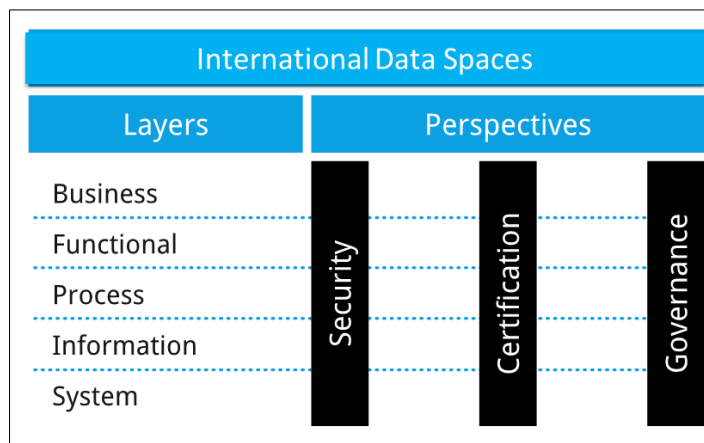
Podatkovni prostori so podatkovne *infrastrukture in okviri upravljanja*, ki omogočajo *razširjeno združevanje in skupno rabo podatkov*. Zagotavljajo *orodja in storitve* za obdelavo in skupno rabo podatkov. Običajno je zbirka podatkov (Data Base DB) je *centralizirana zbirka semantično homogenih podatkov* v togem sistemu za upravljanje podatkov, v nasprotju s *podatkovnim prostorom*, ki je zbirka *decentraliziranih heterogenih podatkov*.

Vzpostavljajo *semantično pregledne strukture upravljanja podatkov* v skladu z zakonodajo EU za izboljšanje *razpoložljivosti, kakovosti in interoperabilnosti podatkov v različnih sektorjih*. Podatkovni prostori delujejo kot tržnice, kar omogoča lastnikom podatkov in ponudnikom storitev, da *prejmejo finančno nadomestilo za svoje prispevke*.

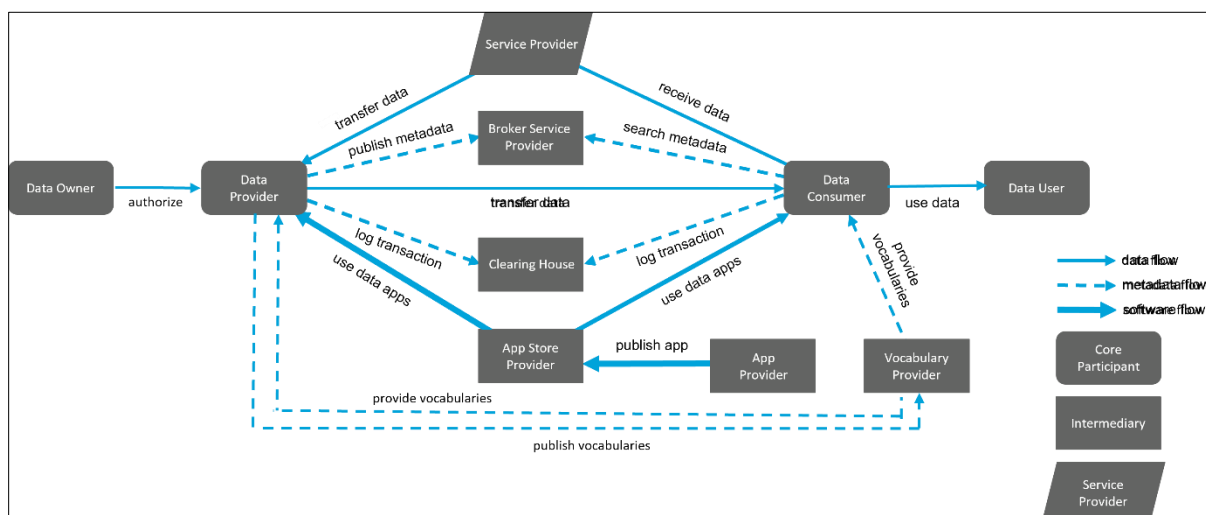
4.3 Mednarodni podatkovni prostori (The International Data Spaces - IDS)

Mednarodni podatkovni prostori (The International Data Spaces - IDS) so *virtualni podatkovni prostori*, ki izkoriščajo obstoječe standarde in tehnologije ter modele upravljanja, ki so dobro sprejeti v podatkovnem gospodarstvu, da olajšajo varno in standardizirano *izmenjavo podatkov in povezovanje podatkov v zaupanja vrednem poslovnem ekosistemu*. Pobuda Mednarodni podatkovni prostori predlaga model referenčne arhitekture (Reference Architecture Model - RAM). Referenčna arhitektura IDS je na *višji ravni abstrakcije* kot običajni arhitekturni modeli konkretnih programskih rešitev.

Neprofitna koalicija podatkovnih prostorov (International Data Space Association - IDSA) je, prisotna v 28 državah. Vzpostavila je arhitekturo, vmesnike in vzorčno kodo za odprt, varen podatkovni ekosistem zaupanja vrednih partnerjev in je namenjena promociji IDS-RAM za *vzpostavitev mednarodnega standarda*. Referenčni arhitekturni model IDS - RAM 4 (IDS Reference Architecture Model 4) vsebuje pet slojev ter tri vidike mednarodnih podatkovnih prostorov (Slika 30).



Slika 30 Pet plasti referenčnega arhitekturnega modela in trije vidiki (vir: International Data Space Association - IDSA)



Slika 31 Osnovne interakcije za izmenjavo in deljenje podatkov v mednarodnih podatkovnih prostorih (vir: International Data Space Association - IDSA)

Objekti IDS, ki jih morajo upravljati udeleženci IDS, so:

- **Povezovalnik (Connector):** tehnična osnovna komponenta, ki je potrebna, da se udeleženec pridruži mednarodnim podatkovnim prostorom.
- **Podatki (Data):** tukaj sinonim za podatkovno sredstvo, tj. vsebino, ki jo ponudnik podatkov izpostavi izmenjavi.
- **Besednjak (Vocabulary):** ontologije, referenčni podatkovni modeli ali elementi metapodatkov, ki jih je mogoče uporabiti za označevanje in opis naborov podatkov, pravilnikov uporabe, aplikacij, virov podatkov o storitvah in podobno.
- **Identiteta (Identity):** podatki o udeležencih in za udeležence IDS.
- **Aplikacija (App):** aplikacije, ki jih je mogoče namestiti znotraj povezovalnika. Aplikacije olajšajo delovne tokove obdelave podatkov. Certificira jih lahko certifikacijski organ po certifikacijskih postopkih, opredeljenih v perspektivi certificiranja.
- **Transakcija (Transaction):** zajema vse dejavnosti, ki se izvajajo med izmenjavo podatkov.
- **Storitev (Service):** programska oprema, ki se izvaja v konektorju in je na voljo kot storitev (algoritem in računalniški čas).

Konektor mednarodnih podatkovnih prostorov, ki je *osrednja komponenta arhitekture*, je implementiran v različnih različicah in ga je mogoče kupiti pri različnih prodajalcih. Kljub temu lahko vsak konektor komunicira s katerim koli drugim konektorjem (ali drugo tehnično komponento) v ekosistemu mednarodnega podatkovnega prostora, kar bo zagotovilo *standardizirano medopravilnost*.

Ekosistem podatkov: Arhitektura mednarodnih podatkovnih prostorov ne zahteva zmogljivosti osrednjega shranjevanja podatkov. Namesto tega zasleduje idejo decentralizacije shranjevanja podatkov, kar pomeni, da *podatki fizično ostanejo pri posameznem lastniku podatkov*, dokler niso preneseni na zaupanja vredno osebo. Ta pristop zahteva *celovit opis vsakega vira podatkov ter vrednost in uporabnost podatkov* za druga podjetja, skupaj z zmožnostjo *integracije besednjakov* podatkov, *specifičnih za domeno*. Poleg tega *metapodatkovni posredniki* v ekosistemu zagotavljajo storitve za *iskanje podatkov v realnem času*.

Vzpostaviti bo treba nove poslovne modele za te storitve z zagotavljanjem klirinških mehanizmov in *funkcij zaračunavanja* ter z ustvarjanjem *posredniških rešitev in tržnic za metapodatke*, specifičnih za domeno.

Neprofitna koalicija podatkovnih prostorov (International Data Space Association - IDSA) s sedežem v Dortmundu v Nemčiji, ima trenutno (november 2024) 183 institucij članic.

Usmeritev za GURS: Predlagamo, da se kot neprofitna organizacija včlani v Neprofitno koalicijo podatkovnih prostorov in spremlja ter vpliva na razvoj mednarodnih podatkovnih prostorov (<https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4>).

Strategija Digitalna Slovenija 2030

Tudi krovna strategija Digitalna Slovenija 2030 opredeljuje podatke kot strateško surovino in gonilo pametne družbe 5.0 in poudarja izjemni pomen zagotavljanja posodobljenih, pravočasnih in celovitih, *medopravilnih* podatkov. V Sloveniji smo v začetku leta 2024 naredili prvi korak do vzpostavitve podatkovnih prostorov na posvetu v organizaciji Ministrstva za digitalno preobrazbo. Po vzoru EU bo treba pripraviti podatkovne prostore (za najmanj deset sektorjev), trenutno ima EU že 14 sektorskih podatkovnih prostorov, in sicer *zdravje, mobilnost, finance, proizvodnja, okolje, energija, kmetijstvo, javna uprava, spretnosti, znanost in inovacije (Open Science Cloud), kulturna dediščina, pametna mesta in skupnosti (zeleni dogovor), jezik, mediji in turizem*. Izzivi za vsakega od načrtovanih podatkovnih prostorov bodo odvisni od posebnosti vsakega sektorja s ponavljajočimi se elementi, kot so interoperabilnost in prenosljivost podatkov ter mehanizmi za dostop do podatkov.

4.4 Usmeritve za interno (nacionalno) semantično povezovanje GURS

V *analizi problemov* smo si zastavili več raziskovalnih vprašanj (glej 1.1.5) in na osnovi njih prišli do pomembnih ugotovitev.

Prehod na politiko odprtih podatkov predstavlja izziv za poslovni model nacionalnih kartografskih in katastrskih agencij po vsem svetu. Odprti, povezani podatki (OLD) bi morali biti del poslovnih razprav, vključno z izjavami o strateški viziji in ambicijah.

Najpogostejše omenjena težava je heterogenost podatkov, to je pomenska, strukturna in sintaktična nedoslednost. Zaradi semantične heterogenosti (Bucher et al., 2020) pravnih definicij in terminologije med neodvisnimi vladnimi agencijami za upravljanje zemljišč je pristop od zgoraj navzdol oviran (ali celo nemogoč). Heterogenost razpoložljivih prostorskih podatkov uporabnikom podatkov, spletnim aplikacijam in storitvam otežuje odkrivanje, interpretacijo in uporabo informacij v velikih in porazdeljenih spletnih sistemih (van den Brink et al., 2018). Tudi za katastrske proizvajalce in podatke so značilne težave glede heterogenosti. Prav tako velja, da tudi metapodatki, ki so standardizirani in urejeni z direktivami, kot je "Infrastruktura za prostorske informacije v Evropski skupnosti (INSPIRE)," 2007, niso pomensko definirani znotraj kontekstno dovolj bogatega referenčnega sistema. Naša

raziskava je razkrila tudi opredelitev termina »*okoliščine podatka*«, (angleško »data context«) kot enega od problemov pri uveljavljanju paradigme LD v praksi geodetskih uprav, saj so *metapodatki pogosto preslabo pripravljeni*. Ena od rešitev težave integracije, povezovanja podatkov, je *zagotoviti te podatkovne silose kot povezane, odprte podatke* in izvesti integracijo z uporabo povezovalnih semantičnih tehnologij.

Rešitve

Treba je oblikovati formalne besednjake za ustvarjanje in objavo LD zahteva razumevanje področja znanja in predvidene uporabe (Čaгдаš in Stubkjær, 2015b; Ronzhin et al., 2019). Vzpostaviti je treba standardni nabor dogovorjenih *ontologij, besednjakov in nizov ključnih besed geografskih metapodatkov za opis odnosov med prostorskimi značilnostmi* (Wiemann & Bernard, 2016).

Najproduktivnejše države glede objavljanja LD so Nizozemska, Švica in Finska. Najpogosteje objavljene v EU (odprtokodno in pripravljeno za povezovanje) so teme: upravne enote, sledijo imena krajev, identifikatorji zgradb in naslovi. Sklepamo, da dajejo Geodetske uprave prednost objave temam, ki omogočajo posredno sklicevanje (poizvedba, navigacija) za zunanje uporabnike.

Največji izzivi pri izvajanju objavljanja odprtih povezanih podatkov OLD so *odločanje o prioritetah razvoja*, sledenje tehnološkemu trendu, *kompetentnost in spretnosti osebja* ter razpoložljivost smernic najboljše prakse (kot jo predstavlja ta raziskava), *ustvarjanje ontologij* iz obstoječih modelov UML, zagotavljanje *ustrezne kakovosti metapodatkov* ter uporaba razširljivega pristopa za zmanjšanje časa in stroškov, povezanih s postopkom objave.

Za sodelovanje geodetske uprave z drugimi državnimi organi, kot so statistika, državne agencije in občine je potreben vzpostavljen Okvir za sodelovanje, kot platforma (na primer NEO), ki bo prispevala k zagotavljanju medopravilnosti in integraciji različnih vrst informacij javnega sektorja in drugih geoprostorskih informacij. Pri tem ostaja dejstvo, da gre za *zelo kompleksno dejavnosti modeliranja za objavljanje povezanih podatkov (OLD) z vidika obravnave celotne vladne infrastrukture (različna ministrstva)* tudi zaradi različnih pravnih in upravljaljskih zahtev, ki obkrožajo posamezen podatkovni model.

Za odprto objavo geoprostorskih podatkov je treba uporabiti načela celostnega pristopa, iz sheme petih oziroma sedmih »zvezdic« (Tim Berners-Lee, 2010):

1. Omogočite dostop do podatkov na spletu v kateri koli obliki.
2. Dajte podatke na voljo kot strukturirane podatke (npr. Excel).
3. Uporabljajte nelastniške formate (npr. CSV namesto formata Excel).
4. Uporabite URI-je za označevanje stvari, da lahko ljudje pokažejo na vaše podatke.
5. Povežite svoje podatke z drugimi podatki, da zagotovite kontekst.
6. Naboru podatkov zagotovite eksplicitno shemo metapodatkov, ki pojasnjuje nabor podatkov.
7. Preverite podatke glede na shemo za boljšo kakovost.

Ta celovit pristop omogoča ponovno uporabo podatkov za boljše razumevanje značilnosti nabora podatkov in *ovrednotenje kakovosti podatkov* za predvideni namen (Hyvönen et al., 2014). Z izvajanjem ustreznih smernic, orodij in strokovnega znanja lahko geodetske uprave zagotovijo, da so njihova prizadevanja za objavljanje LD uspešna in izpolnjujejo potrebe njihovih deležnikov.

- | |
|---|
| <ol style="list-style-type: none"> 1. Prepoznati je treba potrebo po <i>boljšem ujemanju med poslovnim in tehničnim (programerskim) vidikom povezanih podatkov</i>. 2. Vzpostavljena mora biti izrecna razlika med modelom (grafom) znanja (poslovnih pravil domene) in informacijskim modelom: <i>povezan podatkovni model (OLD) sposoben opisati pomen podatkov</i> |
|---|

zunanjemu svetu z uporabo modela znanja, hkrati pa v informacijskem modelu opisati vidike modela, specifične za organizacijo.

3. Prvi korak k grafu znanja narejen tako, da so različni ključni registri geodetske uprave na voljo kot povezani podatki (OLD). Grafa znanja je rezultat združevanja registrov ključev in drugih semantično transformiranih naborov podatkov. Graf znanja naj nastane z ustvarjanjem plasti na vrhu ključnih registrov, v njihovi povezani (LOD) podatkovni obliki.
4. Začetna različica grafa znanja Geodetske uprave naj uporablja nov, uporabniku prijazen model osredotočen na temo/koncept stavbe oziroma zgradbe, ker je na splošno najbližje temu, kako neizkušeni uporabniki dojemajo svoje okolje.
5. Model znanja za isto bazo podatkov naj opredeljuje vsako generično in izmenljivo znanje (poslovna pravila), ki ga je pomembno obdržati v organizaciji, vendar ga je treba deliti z drugimi.
6. Treba je razviti infrastrukturo, treba za objavo povezanih podatkov v organizaciji na osnovi obstoječih standardov geoprostorske skupnosti in drugih inovativnih tehnološke rešitev.
7. Treba je uporabiti uporabljati obstoječe knjižnice programskih rešitev in se tako zato izogibati potrebi po razvoju lastnih rešitev po meri. Za proces preslikave med modeloma je najbolje uporabiti besednjake in ontologije formatov RDF(S) [6], OWL [7] in/ali SKOS [8], ki služijo temu, da se model poveže z zunanjim svetom, v interesu ponovne uporabe in interoperabilnosti.
8. Uporabiti je treba predstavitevni jezik SHACL, ki je najbolj optimiziran za podatke v obliki grafov in povezane podatke.
9. Uporabljati je treba načelo konfiguracije nad kodo, ki zagotavlja, da se isti » informacijski cevovod« (pipeline) uporablja za vse povezane projekte objavljanja podatkov, pri čemer je treba le konfiguriranje komponent, kjer je to treba.
10. Treba je omogočiti, da se ta pristop lahko ponovi na več naborih podatkov, pri čemer izvajanje od vira do objave poteka v kratkem časovnem okviru.
11. Treba je uporabiti tako imenovan pretočni pristop za proces preslikave med modeloma, kar v praksi pomeni, da so vsi povezani podatkovni modeli čim bližje izvornemu modelu in da lahko izbrani viri podpirajo funkcijo pretakanja v interesu dostave podatkov v realnem času.
12. Zagotoviti je treba da upravitelj primarne baze ostane odgovoren za povezano (LD) podatkovno različico svojega silosa, kar omogoča, da graf znanja spada pod drug (poslovni) model upravljanja.
13. Integrirani podatkovni model potreben in bi ga morala razviti in vzdrževati Geodetska uprava sama
14. Proces pretvorbe/preslikave relacijskih podatkov za dano geoprostorsko bazo (zbirko podatkov) v povezane podatke se zaključi v več korakih, izvedenih med ekstrakcijo, pretvorbo in nalaganjem (ETL) procesa.

Ta pristop poudarja izboljšano stroškovno učinkovitost in učinkovitost virov ter krepi poslovni primer za povezane podatke v organizaciji, kot je Geodetska uprava.

Usmeritve za GURS: Zasnovali smo usmeritve za semantično povezovanje baz podatkov geodetske uprave z vidika sektorskih podatkovnih prostorov, opredeljenih s strani EU, v obliki preglednice (Slika 32). Baze podatkov geodetske uprave in storitve (stolpci v tabeli) imajo ključno vlogo v večini sektorskih podatkovnih prostorov (vrstice v tabeli) preko uporab podatkov, aplikacij v posameznih sektorjih (celice na presečiščih vrstic in stolpcev v tabeli).

Applikacije	Podatki in storitve GURS								
	katastri								
Sektorski podatkovni prostori EU	Topografski podatki - zemljevidi, karte	GNSS omrežje SIGNAL, pretvorbe in transformacije	Sistemski načrti in kataster zemljišč	Sistemski načrt in kataster stavb	Fizikalne lastnosti nepremičnega premoženja	Kataster grajene (komunalne) infrastrukture	Množično vrednotenje nepremičnin	Register prostorskih enot in EHIŠ	Register zemljepisnih imen (REZI)
1 Zdravje	zdravstvene storitve, nujne službe, upravljanje zdravstvene krize in pandemije	nujne službe, upravljanje zdravstvene krize in pandemije			nujne službe, upravljanje zdravstvene krize in pandemije	upravljanje prekrške s pimo, meteorološko in fekalno vodo		nujne službe, upravljanje zdravstvene krize in pandemije, geoprostorska statistika	podlage v GIS
2 Mobilnost	podlage v GIS	navigacija v realnem času, zmanjševanje zastojev v prometu, zmanjševanje emisij, prehod na zelenjši transport	množične katastrske preureditve za transportna omrežja	navigacija v stavbi	navigacija v stavbi	množične katastrske preureditve za transportna omrežja	vrednoti mobilnostnih omrežij	geoprostorska statistika	podlage v GIS
3 Finance			upravljanje tveganj, prostorske razsežnosti pravic na zemljiščih in stavbah, preprečevanje zlorab	upravljanje tveganj, prostorske razsežnosti pravic na zemljiščih in stavbah, preprečevanje zlorab	lastnosti nepremičnega premoženja za izračun vrednosti in hipoteke	ocen vrednosti grajene infrastrukture	zpremljanje trga nepremičnin, vrednosti, razvoj in upravljanje nepremičnin	ocena kapitala v nepremičnem premoženju, geoprostorska statistika	
4 Proizvodnja	transportni sistemi, podlage v GIS	geodetska izmera in optimizacija preskrbnih verig	prostorske razsežnosti pravic na zemljiščih in stavbah	prostorske razsežnosti pravic na zemljiščih in stavbah	geodetska izmera, kartiranje in popis	geodetska izmera in kartiranje	razvoj in upravljanje nepremičnin	logistika, obvladovanje nesreč	podlage v GIS
5 Okolje	zpremljanje okolja in vplivi, podlage v GIS		upravljanje tveganj, prostorske razsežnosti pravic na zemljiščih in stavbah, preprečevanje zlorab	upravljanje tveganj, prostorske razsežnosti pravic na zemljiščih in stavbah, preprečevanje zlorab	zpremljanje okolja in vplivi nepremičnega premoženja na okolje	zpremljanje vplivov komunalne infrastrukture na okolje	prostorski razvoj in upravljanje nepremičnin	nujne službe, obvladovanje nesreč, geoprostorska statistika	podlage v GIS
6 Energija	raba energije, potencial proizvodnje zelene energije, podlage v GIS	optimizacija preskrbnih verig	upravljanje tveganj, prostorske razsežnosti pravic na zemljiščih in stavbah, preprečevanje zlorab	upravljanje tveganj, prostorske razsežnosti pravic na zemljiščih in stavbah, preprečevanje zlorab	zpremljanje prekrške, načrtovanje prekrške,	zpremljanje prekrške, načrtovanje prekrške, zpremljanje vplivov na okolje	prostorski razvoj in upravljanje nepremičnin	obračun, nujne službe, odpravljanje napak, obvladovanje nesreč, geoprostorska statistika	podlage v GIS
7 Kmetijstvo	reliefni elementi proizvodnega potenciala zemljišč in odprtosti gozda	precizno kmetijstvo za pridelavo hrane, optimizacija preskrbnih verig	prostorske razsežnosti pravic na zemljiščih in stavbah, množične katastrske preureditve	prostorske razsežnosti pravic na kmetijskih, gozdanih zemljiščih in stavbah	lastnosti kmetijskih stavb in zgradb, vrednotenje	namakalni sistemi, osuševalni sistemi, protipoplavni sistemi	razvoj kmetijstva in upravljanje nepremičnin	geoprostorska statistika	podlage v GIS
8 Javna uprava	podlage v GIS	podpora mobilni inventarizaciji prostora in nepremičnin	prostorske razsežnosti pravic na zemljiščih in stavbah	prostorske razsežnosti pravic na zemljiščih in stavbah	lastnosti nepremičnega premoženja za izračun vrednosti za davke in socialne transfere	lastnosti nepremičnega premoženja za izračun komunalnega prispevka	razvoj in upravljanje nepremičnin	geoprostorska statistika	podlage v GIS
9 Pametna mesta in skupnosti	podlage v GIS	podpora mobilni inventarizaciji prostora mesta in nepremičnin	prostorske razsežnosti pravic na zemljiščih in stavbah	prostorske razsežnosti pravic na zemljiščih in stavbah	informatično načrtovanje, podpora inventarizaciji prostora mesta in vrednotenje premoženja	opremljenost, zpremljanje vplivov na okolje, ocena kapitala v infrastrukturi	razvoj in upravljanje nepremičnin, ocena kapitala v nepremičnem premoženju	geoprostorska statistika	podlage v GIS
10 Kulturna dediščina	podlage v GIS	podpora mobilni inventarizaciji kulturne dediščine	prostorske razsežnosti pravic na zemljiščih in stavbah	prostorske razsežnosti pravic na zemljiščih in stavbah	prostorske razsežnosti pravic na stavbah in varovanje dediščine		ocena kapitala v nepremičnin kulturni dediščini	geoprostorska statistika	podlage v GIS
11 Turizem	podlage v GIS								podlage v GIS
12 Spretnosti									
13 Znanost in inovacije	podlage v GIS	da	da	da	da	da	da	da	podlage v GIS
14 Jezik									da
15 Mediji	podlage v GIS							da	podlage v GIS

Slika 32 Pregled sektorskih podatkovnih prostorov, osnovnih zbirk podatkov GURS ter njihovih potencialnih sektorskih aplikacij (lasten prikaz)

Opomba: Za bolj čitljiv pregled glej prilogo podatkovni prostori EU in baze GIS.

4.4.1 Slovenski nacionalni interoperabilnostni okvir (NIO)

Potrebe po medopravilnosti se nenehno spreminjajo zaradi novih politik, pravnih aktov, pobud in novih nastajajočih tehnologij. V okviru projekta smo organizirali sestanek s predstavniki Ministrstva za digitalno preobrazbo, Direktorata za razvoj digitalnih rešitev in podatkovno ekonomijo, podrobneje Sektorja za upravljanje s podatki.

Predstavili so nam Nacionalni interoperabilnostni okvir (NIO), preko katerega zagotavljajo smernice javnim upravam, kako izboljšati upravljanje interoperabilnosti, vzpostaviti medorganizacijske odnose in kako zagotoviti racionalizacijo postopkov, ki podpirajo digitalne storitve. NIO je enotno spletno mesto za objavo nacionalnih politik in smernic za digitalno preobrazbo ter interoperabilnost virov in digitalnih storitev javnega sektorja, kjer se lahko izvajajo naslednje ključne vsebine:

- Objavljanje izdelkov interoperabilnosti.
- Objavljanje informacij o interoperabilnosti.
- Sistematično iskanje, prikazovanje in predstavljanje izdelkov interoperabilnosti.

V okviru svojih dejavnosti organizirajo tudi kolokvije za uslužbence v javni upravi (na primer za tiste, ki delajo z *dokumentarnim in arhivskim gradivom* – 16. kolokvij). Na kolokvijih *predstavljajo uredniško politiko NIO*, s katero določajo postopek od vnosa produkta do glasovanja za medopravilni izdelek.

Postopek bogatitve in nadgradnje metapodatkov za opis semantičnih interoperabilnostnih izdelkov naj se izvaja v skladu s standardom Asset description metadata schema (ADMS), kar pomeni generiranje datoteke RDF v skladu z ADMS, pri čemer navajajo uspešna primera zajema metapodatkov interoperabilnostnih izdelkov nad OpenScience.si in GURS (na primer Register zemljepisnih imen – REZI).

Zajem potrebnih metapodatkov za opis naborov odprtih podatkov v skladu s standardom »DCAT Application Profile for data portals in Europe«.

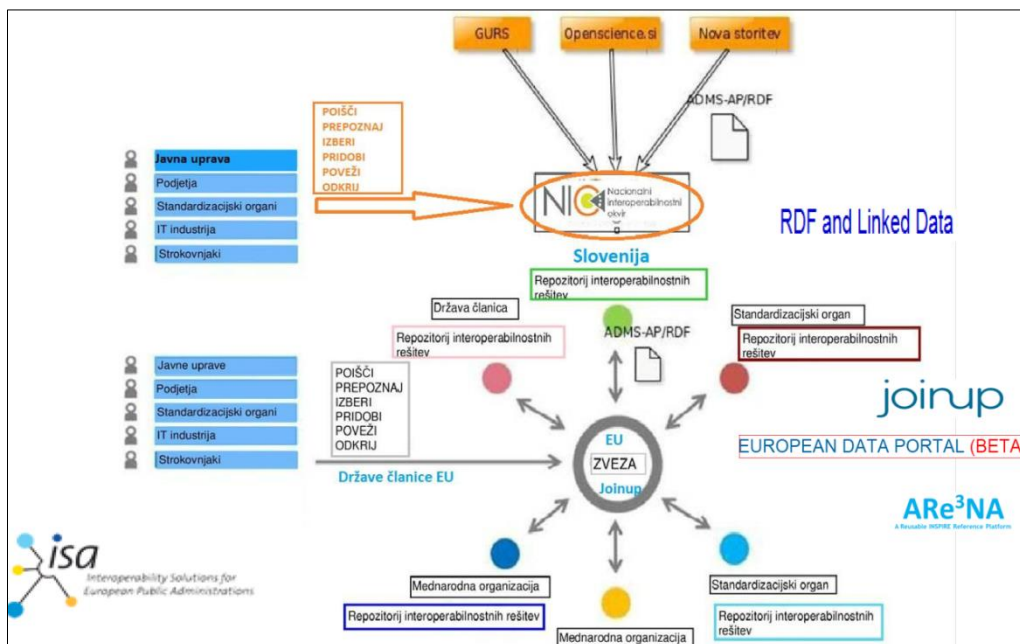
Cilj uredniške politke je tudi

- *označevanje semantičnih izdelkov* za objavo na drugih portalih,
- *označevanje naborov podatkov* za objavo na drugih portalih,
- podpora NIO za *objavljanje zunanjih podatkovnih zbirk*.

Za GURS navajajo, da zajem podatkov s portala GURS v DCAT poteka preko aplikacijskega vmesnika GURSa, bolj specifično URLjev, ki omogočajo izvoz vsebine GURS v XML, pri čemer vsi metapodatki, ki jih ima GURS na portalu, ustrezajo formatu DCAT.

Preslikavanje (Slika 33) poteka na portalu NIO preko uporabniškega vmesnika za podporo prenosa velikih podatkov. Primer izvedbe: ko se podatki z GURS uspešno preberejo in preslikajo v notranje podatkovne strukture NIO, se omogoči izvoz v poljubnem podprtem formatu (XML, RDF, text).

Usmeritve za GURS: Za zbirke podatkov, navedene v Slika 32 Pregled sektorskih podatkovnih prostorov, osnovnih zbirk podatkov GURS ter njihovih potencialnih sektorskih aplikacij preslikati v format DCAT.



Slika 33 Predstavitev strukture RDF in povezanih podatkov preko portala NIO (Šeponja, MJU, 2015)

4.4.2 Platforma za semantično interoperabilnost (PZSI)

Tehnična in semantična prenova registrov in evidenc, V okviru EU operacije »Povezljivost, odprtost, kakovost (POK)« je imela za posledico vzpostavitev Platforme za semantično interoperabilnost, katere uporabniki so razvijalci aplikativnih rešitev v javni upravi, ter pripravljavci zakonodaje in uradnih besedil. Namen platforme PZSI je na enem mestu zbirati in distribuirati:

- enotne *podatkovne modele* temeljnih registrov in ključnih informacijskih rešitev javne uprave,
- enotne *šifrante*, ki se uporabljajo v okviru javne uprave,
- enolično in jasno definirano *terminologijo*, ki se uporablja v okviru javne uprave.

4.4.3 NIO Centralni besednjak javne uprave

Osrednji besednjak zagotavlja konsistenten jezik in predstavlja referenco za snovalce informacijskih rešitev, aplikacij, registrov, evidenc ter oblikovalce zakonodajnih in drugih besedil. Osrednji besednjak enolično in jasno definira ključno terminologijo, ki se uporablja v okviru javne uprave. Vsi pojmi v centralnem besednjaku imajo jasno, nedvoumno in ne redundantno opredelitev. V osrednjem besednjaku so termini organizirani v hierarhično strukturo. Vsak pojem je lahko v enem ali več odnosih nadrejenosti ali podrejenosti do drugih pojmov (Tričkovič, 2019).

Centralni besednjak je objavljen s pomočjo spletne aplikacije Web Protege, ki je večjezično orodje za upravljanje ontologij in drugih besednjakov. Podpira standarde RDF, SKOS, OWL in druge sodobne standarde s področja semantičnega spleta. Besednjak je bil prvič objavljen v letu 2020, zadnja sprememba besednjaka pa je bila opravljena v juliju 2024. Centralni besednjak v obliki spletne aplikacije, v oktobru 2024 ne omogoča povezave zunanjim uporabnikom (<https://cnb.sigov.si/webprotege/#login>) vendar je dostop do Centralnega besednjaka zaenkrat mogoč le znotraj omrežja državnih organov HKOM, zato ga v nalogi ne moremo analizirati in komentirati.

Ministrstvo za digitalno preobrazbo, Direktorat za razvoj digitalnih rešitev in podatkovno ekonomijo je pripravil več usmeritev (zadnjič posodobljenih v letu 2020), in sicer Usmeritve za uporabo centralnega besednjaka, Seznam ontologij ter Navodila za uporabo orodja WebProtege.

Usmeritve za GURS: Pripraviti pregled in analizo pomenskih relacij terminov s področij delovanja (katastrska poslovna pravila, topografska poslovna/snemalna pravila, GJI poslovna pravila in podobno) GURS in sorodnih, sektorskih področij:

- (Slika 32 Pregled sektorskih podatkovnih prostorov, osnovnih zbirk podatkov GURS ter njihovih potencialnih sektorskih aplikacij (lasten prikaz),
- 4.4.3 NIO Centralni besednjak javne uprave in
- besednjakov, ki smo jih navedli v razdelku 4.4.6 Seznam referenčnih terminoloških virov NIO za javno upravo RS.

4.4.4 Metodološka navodila za pripravo izdelka medopravilnosti in vnos v Portal NIO

V skladu s pravili Nacionalnega interoperabilnostnega okvira (NIO) in v skladu s skrbništvom izdelkov interoperabilnosti, glede načina izbiranja primernih *izdelkov interoperabilnosti* se uporablja Vprašalnik CAMSS, ki zajema 80 vprašanj. V skladu z usmeritvami vsebin je priporočeno, da se v repozitorij portala NIO vključijo tisti izdelki, ki dosegajo vsaj 60 % pozitivnih odgovorov na *vprašanja o skladnosti*.

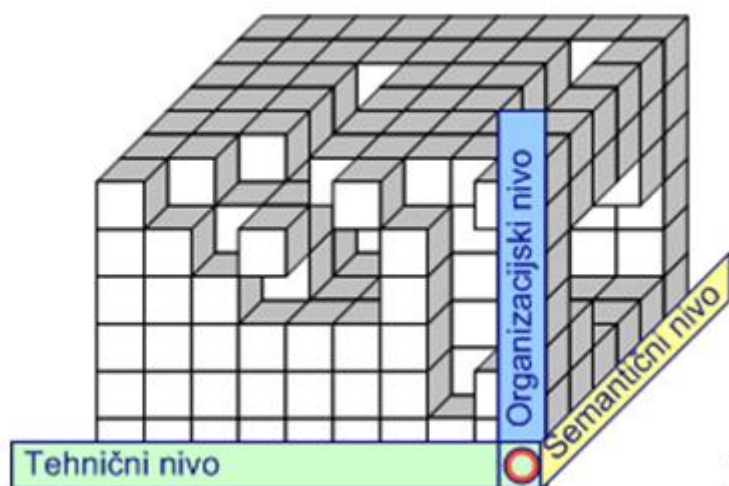
Pri pripravi podatkov za *postopek potrjevanja* izdelka interoperabilnosti je treba pripraviti *dva sklopa metapodatkov*, in sicer splošne in posebne metapodatke. V portal NIO vnašajo izdelki različnih ravni in pod ravni interoperabilnosti.

Strategija elektronskega poslovanja (SREP), kot *politični okvir interoperabilnosti*, predvideva elektronsko poslovanje javne uprave s pomočjo *horizontalne povezanosti* sistemov informacijsko komunikacijske tehnologije in s *prenovljenimi poslovnimi procesi*. Vsaka nacionalna javna uprava prispeva k zagotavljanju evropskih javnih storitev znotraj svojega nacionalnega pravnega okvira z usklajeno zakonodajo (memorandumi, odločbe, pogodbe, pravilniki, sklepi, soglasja, uredbe in zakoni) med sodelujočimi javnimi upravami držav članic EU, ker predstavlja *pravno interoperabilnost*.

Sodelujoči se morajo sporazumeti glede skupnega *standarda za opis poslovnih procesov*, kar vključuje tudi *prenovo poslovnih procesov*, ki se začne z njihovo analizo, s ciljem, da se poenoti razumevanje *taksonomije procesov*, ker lahko zagotovi *organizacijsko interoperabilnost*. Glede povezovanja informacijskih sistemov, lastništva podatkov, omejitev in pridobivanja podatkov iz podatkovnih zbirk se vzpostavijo *medorganizacijski dogovori*.

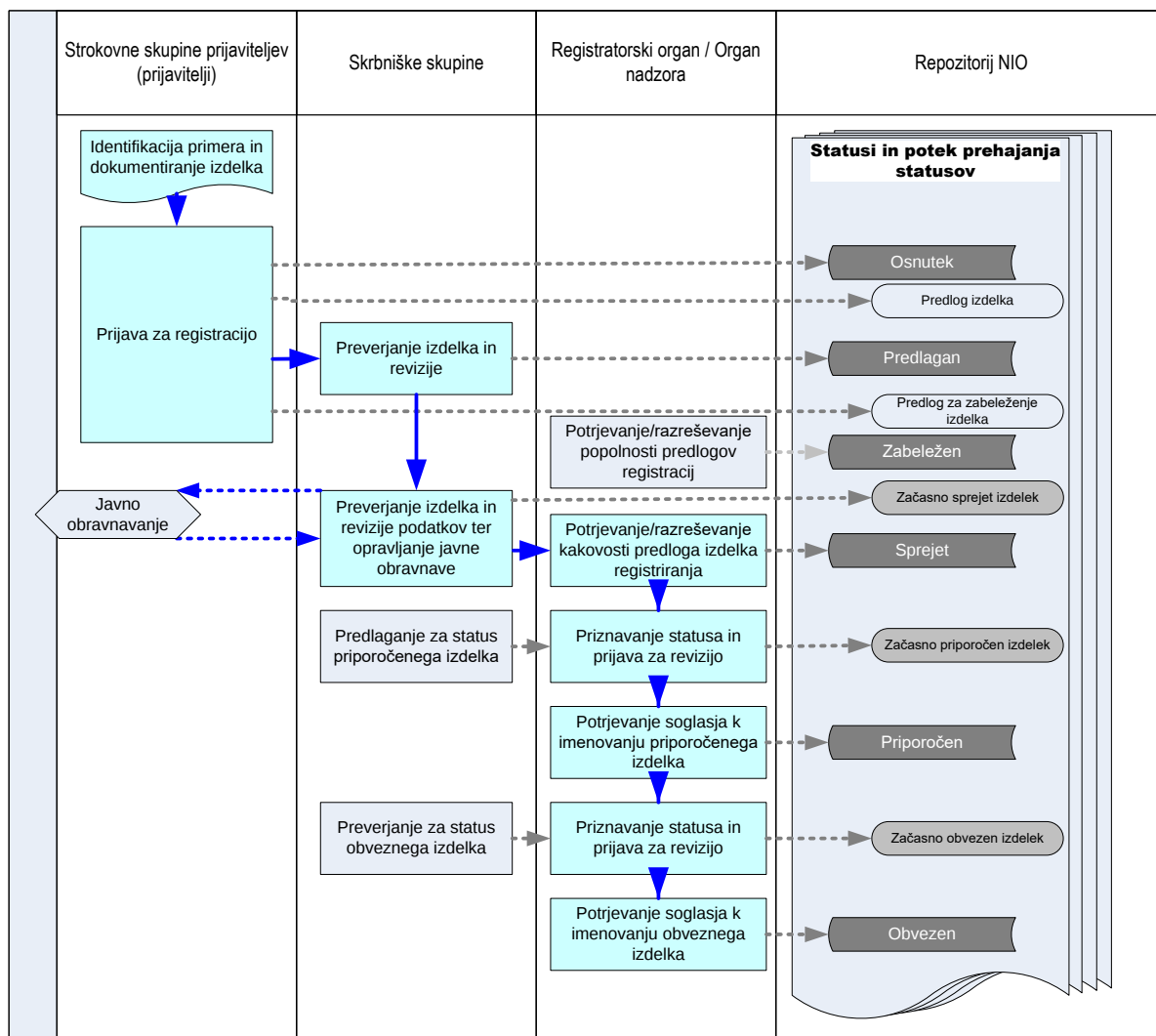
Da se lahko *elementi informacijskih sistemov* (entitete, podatkovni vmesniki, semantična orodja, tehnični semantični meta modeli – XML sheme, zbirke podatkov) pri nadaljnjem razvoju ustrezno *uskladijo v pomenskem smislu* je treba opredeliti in *določiti vsebine in opise*, ki bodo nedvoumno pokazali *pomen posameznega elementa*.

Na podlagi določil NIO se uskladijo in se dogovorijo standardizirani elementi za *tehnične interoperabilnosti*, kot so arhitekture porazdeljenih rešitev, dostopnost, EDI in XML standardi, standardi integracije, mrežnih storitev, naborih znakov, posredovanje in varnost sporočil, predstavitev in izmenjava podatkov, tipi datotek, formati dokumentov, avtorstva rešitev, domenskih imen, storitve javnih imenikov, storitve medsebojnih povezav, večkanalni dostopi ter zgoščevanje datotek.



Slika 34 Glavne ravni interoperabilnosti (vir: NIO portal)

Uporabniki iz institucij državnih organov lahko dobijo dodatne vloge v okviru Portala NIO, glede na pravice za urejanje vsebin, ki jim jih dodeli glavni urednik v skladu s politiko NIO. Nadzorni uredniki odločajo glede primernosti posameznega izdelka in njegovi vključitvi med potrjene in sprejete izdelke interoperabilnosti, ki jih morajo ostali ustvarjalci, vključeni v izvajanje interoperabilnosti, upoštevati in delovati v skladu z njimi.



Slika 35 Krovni proces prijave izdelkov v portalu NIO (vir: Metodološka navodila NIO)

Prijavitelji izdelkov interoperabilnosti spremljajo komentarje *nadzornih urednikov* in strokovne javnosti v času pred in ob javni obravnavi, in na osnovi pripomb uskladi izdelek (Slika 35).

Skrbnik izdelka, urednik repozitorija (v odgovorni organizaciji, ki ima sklenjen partnerski NIO sporazum o sodelovanju) spremlja svoj izdelek tudi potem, ko je le-ta že v enem izmed končnih statusov (sprejet, priporočen ali obvezen) in je v uporabi. Pri tem redno *spremlja področje*, ki ga izdelek pokriva; v primeru sprememb ali dopolnitev *sproži postopke* za dopolnitev obstoječega izdelka, ponovno obravnavo; skrbi tudi za *obveščanje* uporabnikov o spremembah.

Usmeritev za GURS: za ključne zbirke podatkov (Slika 32 Pregled sektorskih podatkovnih prostorov, osnovnih zbirk podatkov GURS ter njihovih potencialnih sektorskih aplikacij) naj GURS imenuje skrbnike izdelkov, ki naj bodo obenem tudi uredniki repozitorija z zgoraj naštetimi funkcijami.

4.4.5 NIO Terminološke smernice za javno upravo in terminološki dogovori (2023)

V praksi pogosto ni lahko priti do *terminološkega dogovora med strokovnjaki*, saj je treba najti kompromis med različnimi mnenji in stališči o posameznih pojmi, hkrati pa mora biti izbrani termin ustrezen tudi z jezikovnega vidika. V javni upravi je nedosledna in *neustrezna raba terminologije* dokaj pogosta. S ciljem zagotavljanja uporabe enotne terminologije v javni upravi je bil oblikovan urejeni

seznam terminoloških virov, ki naj bi se primarno uporabljali. Ti viri dosegajo ustrezno raven kakovosti in medsebojne usklajenosti.

Če iskanega pojma ni v nobenem referenčnem ali drugem kakovostnem terminološkem viru, je treba izpeljati ustrezni *postopek definiranja novega pojma oziroma termina* zanj. Pri tem se lahko uporabijo postopki, ki so opisani v smernicah. Referenčni terminološki vir je treba izbrati *glede na področje*, na katerem termin uporabljamo. Enak termin ima namreč lahko na različnih področjih različen pomen. Če bi potrebovali isti termin v drugem, strokovnem kontekstu, bi uporabili vir, ki pojem bolj določno in podrobno definira v tem kontekstu.

Ena glavnih nalog v terminologiji je definiranje pojma po terminoloških načelih. Treba je ločiti med *opredelitvijo* (definicijo) in *razlago termina*, čeprav se izraza ne uporabljata dosledno razločevalno (Gajšek, 2022). *Terminološka načela* so orodje za doseganje terminološkega dogovora, katerega namen je kompromis med različnimi mnenji različnih strokovnjakov, in za izbiro najustrežnejšega termina izmed predlaganih. Glavna terminološka načela so (Gajšek, 2022):

1. načelo *ustaljenosti*,
2. načelo *gospodarnosti*,
3. načelo *jezikovno sistemske ustreznosti*,
4. načelo *jezikovnokulturne ustreznosti*.

Opredelitve terminov naj čim bolj razumljivo in natančno pripravijo *strokovnjaki za konkretno področje*. Predlog termina naj nato skupaj z definicijo posredujejo na naslov elektronske pošte terminologija.ju@gov.si. V primeru potrebe po definiranju novega termina s področja informatike, je treba ravnati v skladu z navodili za urednike na spletni strani www.islovar.org/islovar/navodila-urednikom. V postopku oblikovanja novega termina je treba zagotoviti tudi *sodelovanje pristojnega vsebinskega skrbnika*. Če organ, pri katerem je nastala potreba po definiranju novega termina, ugotovi, da vsebinsko zase ni pristojen, naj naveže *stik s pristojnim področnim organom* (na primer z ministrstvom za okolje in prostor za terminologijo s področja prostorskih podatkov). Za splošno terminologijo je treba ravnati v skladu z navodili na spletni strani Terminološka svetovalnica | ZRC SAZU (zrc-sazu.si).

Nezaželenost večpomenskosti (sinonimov) za isto miselno zasnovo (koncept, pojem)

Po terminološki teoriji naj bi vsakemu pojmu (konceptu) posamezne stroke pripadalo le eno poimenovanje/razumevanje/razlaga/opredelitev. V terminologiji sinonimi niso zaželeni, ker zmanjšujejo možnost enopomenskega oziroma enournega sporazumevanja, z vidika učinkovitega sporazumevanja pa so nefunkcionalni. Zato slovarji usmerjajo bralca na *prednostni izraz*. Splošni izrazi imajo več možnih pomenov, za termine pa se pričakuje, da natančno in nedvoumno izražajo določeno vsebino, kar se izraža z enopomenskostjo in izogibanjem večpomenskosti.

Jezikovno-semantična načela poudarjajo prizadevanje za jasno *izražanje v domačem jeziku*. Pri uvajanju izrazov za nova dognanja je treba izraze iz tujih jezikov sprejemati kritično in s premislekom. Namesto nekritičnega prevzemanja tujk je treba z *domačimi besedotvornimi postopki* ustvariti nove izraze ali že *obstoječim pripisati nov pomen*. Pri trku dveh nasprotujočih načel se je vedno treba odločati na podlagi

konkretnega primera, kar pomeni, da terminološka načela niso samodejno prednostna, ampak je treba raziskati konkretni termin in na podlagi tega izbrati tista načela, ki jih je utemeljeno upoštevati.

Vsak, ki se priključi posamezni stroki, se mora priključiti tudi *terminološkemu dogovoru*. Pri tem ne gre za apriorno slovenjenje mednarodnih strokovnih izrazov, temveč multidisciplinarno sodelovanje posamezne stroke, ki ureja izrazje, in jezikoslovja.

4.4.6 Seznam referenčnih terminoloških virov NIO za javno upravo RS

Terminološki viri, ki so priporočeni za uporabo v javni upravi, morajo dosegati ustrezno raven kakovosti, da bi lahko bili uvrščeni na seznam referenčnih terminoloških virov.

Usmeritev za GURS: S področij, ki smo jih semantično analizirali v tej nalogi, v seznamu najdemo naslednje vire, pomembne za semantično analizo baz podatkov v skrbništvu GURS:

- Gozdarski slovar (gozd)
- Pravni terminološki slovar (tudi nepremičnine, ZRC SAZU)
- Urbanistični terminološki slovar (Urbanistični inštitut Republike Slovenije, Založba ZRC, 2015)
- Slovarček - omrežje SIGNAL (Geodetski inštitut Slovenije)
- Geografski terminološki slovar (ZRC SAZU)
- Slovar toponimske terminologije (Geodetska uprava Republike Slovenije, 1995)
- Seznam tujih zemljepisnih imen v slovenskem jeziku (Geodetska uprava Republike Slovenije, 2001)

Drugi viri, ki niso vključeni v seznam, vendar predstavljajo pomembno strokovno terminološko osnovo za raziskovanje semantične medopravilnosti v geoprostorskih domeni, so še naslednji terminološki viri:

- *Večjezični slovar za področje geoinformatike – v nastajanju (projekt CRP V2-2133: GURS, ZRC SAZU, UL FGg, 2023, še ni v seznamu)*
- Kmetijski tehniški slovar (ZRC SAZU, 1997, ni na seznamu)
- Topografski ključ za izdelavo in prikaz vsebine geodetskih načrtov (GURS, 2006, ni v seznamu)
- Nepremičninski slovar (GZS, Zbornica za poslovanje z nepremičninami, ni v seznamu)

Obstajajo tudi drugi viri, ki vsebujejo pomenske opredelitve terminov, predvsem zakoni in podzakonski predpisi in priročniki. V delovnem sklopu 3 smo navedli, kateri so bili vključeni v semantično analizo zbirk podatkov.

4.4.7 OPSI nacionalni portal in stičišče odprtih podatkov Slovenije

Portal OPSI zagotavlja pravico do brezplačne in preproste ponovne uporabe podatkov, ki so objavljeni na način "odprtih podatkov", in predstavlja *osrednji katalog zbirk podatkov v državi* in je osrednje spletno mesto za *objavo samih podatkov v strojno berljivih formatih*.

Na OPSI portalu je trenutno (leto 2024) za področje Okolje in prostor objavljenih 130 *odprtih občinskih zbirk podatkov* s strani 122 občin (od 212), in sicer v formatih CSV (129 zbirk), XLS (26), HTML (3), KML (2), PDF (1), programski vmesnik OPSI (7 zbirk), s skupno oceno odprtosti 3/5.

V katalogu Informacij javnega značaja (IJZ) na portalu OPSI, za področje Okolje in prostor je objavljenih 203 zbirk podatkov od tega podatki objavljeni kot odprti (157), podatki, ki niso javno dostopni (27 zbirk), na voljo je spletni vpogled v podatke (16), podatki so javno dostopni in na voljo za dostop in ponovno uporabo (3).

Stičišče odprtih podatkov Slovenije – OPSI HUB je spodbujevalni mehanizem za razvoj odpiranja podatkov javnega sektorja in ponovne uporabe odprtih podatkov v Sloveniji.

V letu 2024 je Slovenija v indeksu odprtih podatkov OURData ostala med prvimi desetimi državami OECD, ki vztrajno in sistematično skrbijo za ustrezno zagotavljanje dostopa do odprtih podatkov in spodbujajo uporabo verodostojnih podatkov v korist celotne družbe in gospodarstva. Slovenija je od objave prejšnjega indeksa leta 2019 napredovala iz 10. na 7. mesto na lestvici«.

4.5 Usmeritve za zunanje (mednarodno) semantično povezovanje GURS

Za področje prostorskih podatkov se upoštevajo standardi za prostorske podatke, priporočila OGC in določila direktive INSPIRE. Za zbirke prostorskih podatkov, ki so opredeljene v Zakonu o infrastrukturi za prostorske informacije, pa je obvezno treba upoštevati izvedbena pravila, kot jih določa direktiva INSPIRE.

Slovenski geoportal INSPIRE (<http://www.geoportal.gov.si/>) je namenjen vsem slovenskim institucijam, ki so dolžne zagotavljati z INSPIRE skladne metapodatke, podatke in storitve, institucijam EU ter drugim uporabnikom, ki iščejo informacije o prostorskih podatkih in storitvah nad prostorskimi podatki. Portal zagotavlja informacije o prostorskih podatkih in storitvah, omogoča upravljanje metapodatkov o podatkih in storitvah, avtomatično zbiranje metapodatkov iz drugih skladnih metapodatkovnih sistemov in avtomatičen prenos metapodatkov v druge skladne metapodatkovne sisteme, kot so evropski geoportal INSPIRE (<https://inspire-geoportal.ec.europa.eu>) in evropski podatkovni portal (<https://www.europeandataportal.eu>). Predstavlja centralno točko za informacije o prostorskih podatkih, z INSPIRE skladnih prostorskih podatkih, njihovih metapodatkih in storitvah.

Podatkovne specifikacije, ki so osnova za harmonizacijo podatkov in izvedbo storitev, so pripravljene v UML modelirnem jeziku, kot standardni izmenjevalni format je uporabljena XML/GML notacija. Ključne so predvsem storitve iskanja podatkov oziroma metapodatki (OGC CSW standard), storitve vpogleda (OGC WMS standard) in storitve prenosa (OGC FWS standard).

Povzetek drugih priporočil

- Za opisovanje prostorskih zbirk podatkov, serij zbirk podatkov in storitev svetujemo uporabo formata GeoDCAT.
- Format zagotavlja sintakso RDF za povezavo zveze metapodatkovnih elementov, določenih v osnovnem profilu standarda ISO 19115:2014 in v okviru direktive INSPIRE. Omogoča iskanje metapodatkov na splošnih podatkovnih portalih.
- Predlagamo uporabo standardov za prostorske podatke SIST/TC GIG, CEN/TC 287 GI, ISO/TC 211; OGC-priporočila za prostorske podatke in storitve ter <https://www.ogc.org/> proučitev dobrih praks prostorskih podatkov na spletu: <https://www.w3.org/TR/sdw-bp/>

Za institucije v RS, ki so na podlagi Zakona o infrastrukturi za prostorske informacije dolžne zagotavljati z INSPIRE skladne metapodatke in storitve ter svojo objavo, za EU-institucije, ki želijo pridobivati informacije o slovenskih prostorskih podatkih, za uporabnike, ki potrebujejo informacije o INSPIRE, prostorskih podatkih in storitvah.

4.5.1 Evropska strategija za podatke (Evropska komisija, 2020) in podatkovni prostori

Cilj strategije je, da EU postane vodilna družba, ki bo temeljila na podatkih. Ustvarjanje enotnega trga za podatke bo omogočilo prost pretok podatkov znotraj EU, kjer bo še poseben poudarek pretoku in uporabi podatkov med sektorji v korist podjetij, prebivalcev, raziskovalcev in javnih uprav. Implementacija strategije skozi zakonodajne in tehnične ukrepe bo opolnomočila ljudi, podjetja in organizacije za sprejemanje boljših odločitev na podlagi vpogledov iz neosebnih podatkov, ki bi morali biti na voljo vsem pod določenimi pogoji. Ključni zakonodajni ukrepi s tega področja so Akt o upravljanju podatkov, Akt o podatkih in Akt o umetni inteligenci (gaia-x Hub Slovenia).

Na osnovi načel, določenih v strategiji, se oblikuje Evropski podatkovni ekosistem. Krovni elementi ekosistema so *podatkovne platforme* (Data Platforms), *podatkovni prostori* (Federated Data Spaces) in *podatkovne tržnice* (Data Marketplaces), okolje za semantično razumevanje podatkov.

Evropska komisija opredeljuje, kot smo že navedli devet oziroma deset *podatkovnih prostorov*, ki so ključni za razvoj družbe in gospodarstva: zdravstvo, industrijska proizvodnja, kmetijstvo, finance, mobilnosti, okolje-zeleni dogovor, energija, javna uprava in kompetence (usposobljenosti, veščine). Podatke je treba povezati oziroma zagotoviti interoperabilnost in dostopnost.

Poslovno združenje podatkovnih prostorov (Data Spaces Business Alliance - DSBA) združuje štiri najbolj pomembna *evropska združenja* na področju podatkov: BDVA (Big Data Value Association), FIWARE, Gaia-X in IDSA (International Data Spaces Association). BDVA se osredotoča na vrednost podatkov in podatkovno ekonomijo, FIWARE z implementacijo tehničnih rešitev in standardov, Gaia-X z upravljanjem podatkov in protokoli za posredovanje med ponudnikom in uporabnikom podatkov, IDSA pa postavlja referenčne arhitekture in pravila za združevanje podatkov.

Gaia-X Hub Slovenia (GXH-SI) je federativno infrastrukturno omrežje za varne podatke namenjena uveljavljanju uporabniških ekosistemov na nacionalni ravni, s cilji razviti ekosisteme, povezati nacionalne pobude in zagotoviti osrednjo kontaktno točko zainteresiranim stranem v Sloveniji.

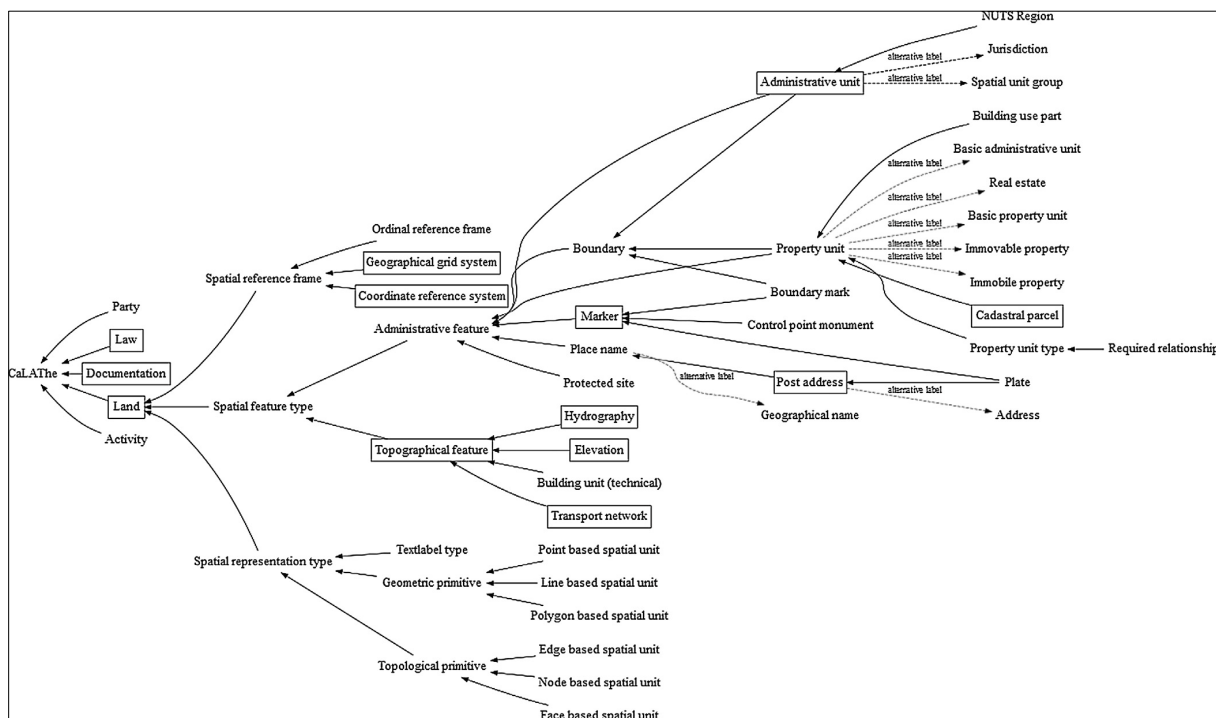
4.5.2 Povezovanje podatkov za zemljiško administracijo

Okrepitev mednarodne komunikacije zahteva večjezične terminološke baze podatkov in druga jezikovna orodja. Konzorcij svetovnega spleta je oblikoval nadaljnjo specifikacijo sistema preproste organizacije znanja (Simple Knowledge Organization System - SKOS), ki je formalni jezik, zasnovan za standardizirano predstavitev strukturiranih besedišč, kot tudi metodologijo povezanih podatkov za objavljanje strukturiranih podatkov v strojno berljivem in med seboj povezani način, da postanejo bolj uporabni. Z uporabo pristopa povezanih podatkov je Evropska komisija nedavno vzpostavila temeljne besednjake e-uprave, tudi *tezavra za področje katastra in zemljiške administracije*.

Glavni namen je prispevati k razvoju „povezane zemljiške uprave“, ki sprejema tehnologije povezanih podatkov za semantično upravljanje naborov podatkov, shranjenih v javnih registrih, ter znanstvenih in zakonodajnih virov, shranjenih v knjižnicah.

Kontrolirani besednjak »Tezaver kataster in zemljiška administracija« (Cadastre and Land Administration Thesaurus - CaLaThe) v formatu SKOS se lahko uporablja za specifikacijo metapodatkovnih zapisov znanstvenih in zakonodajnih virov ter obogatitev teh virov s pomenskimi opombami. Zagotavlja tudi osnovo za nadaljnje pobude za razvoj ontologije. Tezaver v glavnem izhaja iz izrazov standarda ISO 19152:2012 Model domene zemljiške administracije (Land Administration Domain Model - LADM), ki predstavlja *statični vidik domene*. Poročila projekta Oblikovanje lastnine v nordijskih državah (Property Formation in the Nordic countries, Kort og Matrikelstyrelsen, 2006) zagotavljajo osnovo za *časovni ali dejavnostni vidik domene*, medtem ko ANSI/NISO, 2005 Smernice za konstrukcijo, obliko in upravljanje enojezičnih kontroliranih besedišč (Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies) nudi metodološke nasvete. Tezaver je sestavljen iz 143 pojmov, katerih razmerja so zapisana po omenjenem standardu SKOS. Približno ena četrtna izrazov je prevzeta iz obstoječih tezavrov, vključno z večjezičnim

kmetijskim besednjakom AGROVOC, tezavrom GEMET s temami prostorskih podatkov INSPIRE in tezavrom STW za ekonomijo.



Slika 36 Vizualna predstavitev koncepta »zemljišče« (Land), (izrazi v okvirčkih sozeti iz uveljavljenih besednjakov, Cagdas, Stubkear, 2015)

Primer: pojem »prostorske enote« (LADM) je bil označen s tipom prostorske značilnosti, z ožjimi topografskimi značilnostmi in administrativnimi značilnostmi, kot je prikazano na Slika 36. Paket prostorskih enot LADM vključuje pomembno razlikovanje med prostorskimi enotami in njihovo prostorsko predstavitevjo.

4.5.3 INSPIRE skupni slovar konceptov

INSPIRE Feature Concept Dictionary (IFCD) deluje kot skupni slovar konceptov funkcij za vse specifikacije podatkov INSPIRE. Slovar konceptov skupnih značilnosti vsebuje izraze in definicije, ki so potrebni za določanje tematskih vrst prostorskih objektov, njegova glavna vloga pa je zlasti podpiranje prizadevanj za usklajevanje in prepoznavanje nasprotij med specifikacijami vrst prostorskih objektov v različnih temah.

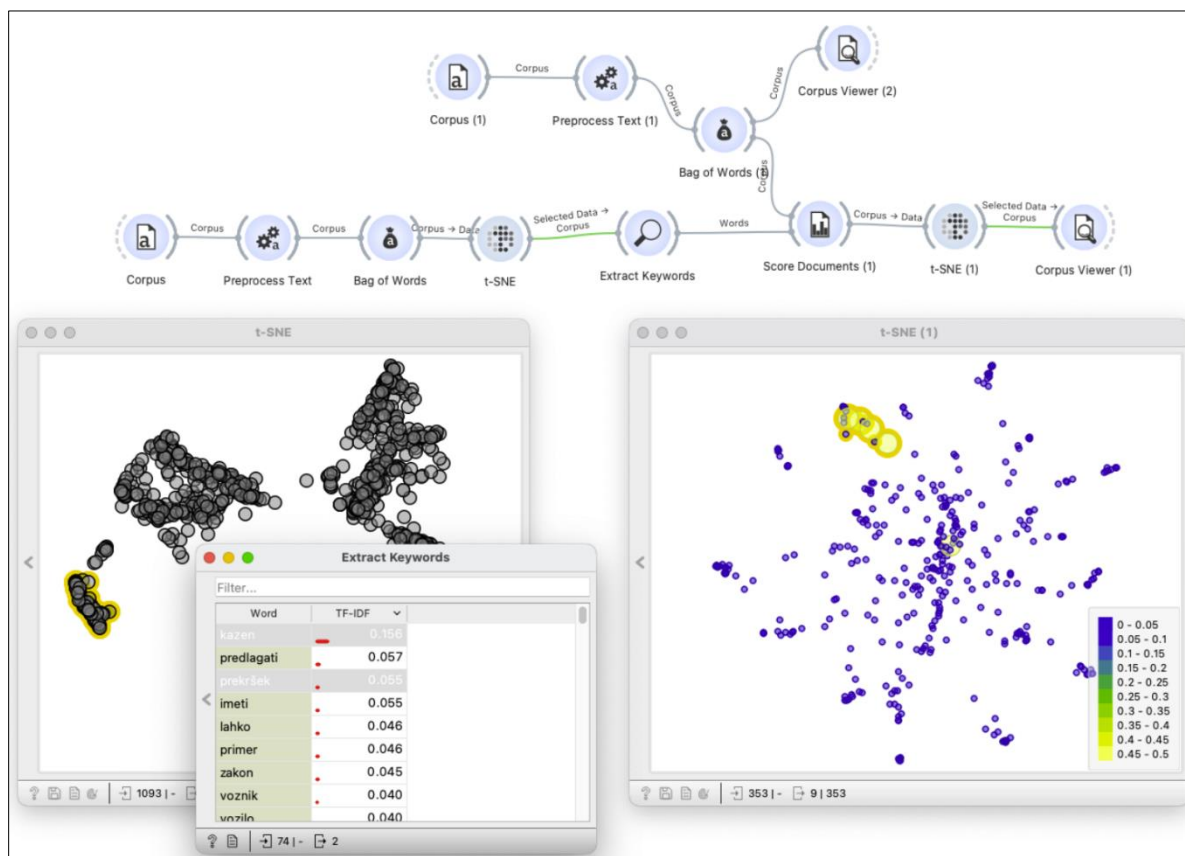
European Commission, Joint Research Centre (JRC) [Dataset] PID: <http://data.europa.eu/89h/jrc-10109-featureconcept>

4.6 Ugotovitve glede pomensko zasnovane generalizacije in povečane večkratne uporabe

Večkratna uporaba podatkov je mogoča, če so zagotovljeni pogoji, kot so povezljivost, odprtost, in kakovost podatkov (POK), ki sodi v okvir operacij EU za tehnično in semantično prenovo registrov in evidenc. V nadaljevanju predstavljamo ključna analitična orodja izvedena v slovenskem okviru navedenih operacij.

Semantični analizator NIO

Razvito je bilo pilotno odprtokodno orodje »semantični analizator« (Lavbič, Žitnik, 2023), ki za svoje delovanje uporablja tehnike umetne inteligence za obdelavo naravnega jezika, s katerim je mogoče analizirati slovenska besedila *po vsebini* (na primer zakonske dokumente), ugotavljati, *kateri pojmi so ključni za razumevanje vsebine* in kateri drugi dokumenti oziroma besedila so glede na semantično/vsebinsko analizo najbolj sorodni. Pojmi na podlagi semantične analize tako postanejo vezni člen med več dokumenti in lahko med njimi odkrivamo soodvisnosti. Tako lahko na primer poiščemo tiste zakonske dokumente, ki se sklicujejo na vir podatkov (register, evidenca, seznam (NIO, 2023).



Slika 37 Procesni diagram semantičnega analizatorja in izluščene pomembne besede iz zakona v okolju Protege Desktop 5.2.0

Orodje semantični analizator je treba vključiti v delovne procese ter načine povezovanja s portali ter namenskimi programi državne uprave.

Usmeritev za GURS: Na področju delovanja GURS uporabiti navedeno orodje, ki bo omogočilo analizo zakonskih dokumentov in s tem gradnjo strokovnih besednjakov na področjih delovanja geodetske uprave in sorodnih področij, s katerimi je stroka v strokovni in zakonodajni soodvisnosti.

Razprave skupnosti uporabnikov semantičnega analizatorja NIO

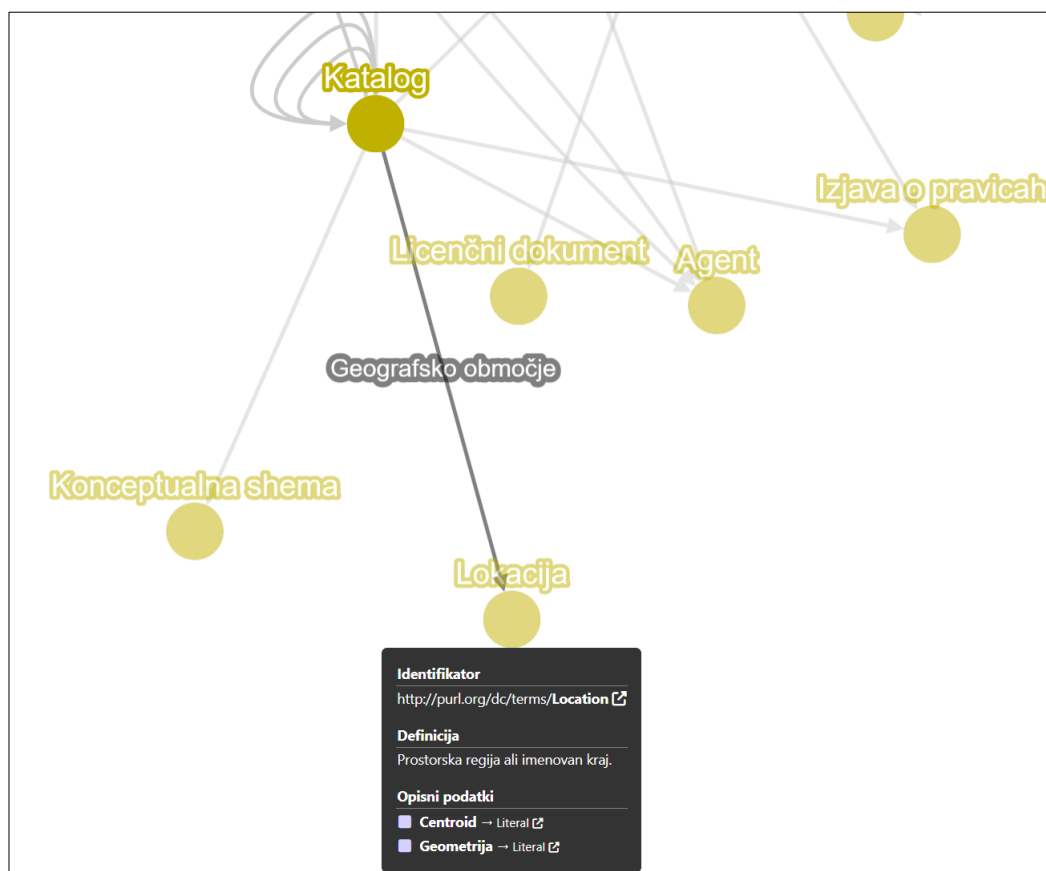
Za razpravljanje v skupnosti uporabnikov je v okviru analizatorja pripravljen poseben zavihek imenovan »issues«, kjer je mogoče pregledovati in sodelovati v obstoječih razpravah ali pa odpirati nove razprave. Ugotavljamo, da ni na voljo nobenih aktivnih razprav glede semantičnega analizatorja dokumentov.

Usmeritev za GURS: Razpravne skupnosti NIO se navezujejo na naš predlog o uvedbi portala za semantično usklajevanje, katerega zasnovo smo izvedli v Rezultati naloge 2.3: Izdelava idejne zasnove platforme za podporo procesu pomenskega usklajevanja. Predlagamo obogatitev NIO razpravne skupnosti z elementi iz rezultatov 2.3.

Podatkovni zemljevid NIO

Pregled nad vzorčnim izborom podatkov, ki se nahajajo na Portalu odprtih podatkov Slovenije je mogoč preko prototipa aplikacije »Podatkovni zemljevid« (The Slovene Data map ontology; Žitnik, Lavbič, 2023, <https://dejanl.github.io/PZ/pz/index-sl.html>), semantične sheme v obliki grafa povezanih podatkov v slovenščini. Shema podatkovnega zemljevida je bila izdana v septembru 2023; je slovar RDF, ki je zasnovan z namenom *opisovanja slovenskih portalov odprtih podatkov*. Uporaba sheme PZ za opisovanje zbirk podatkov, posameznih izvozov in atributov lahko upravljavcem podatkov poveča vidnost njihovih podatkov in omogoči končnim uporabnikom, da najdejo ustrezen API za samodejno pridobivanje le njih.

Podatkovni zemljevid vsebuje tudi priporočila pri objavi podatkov ter pregled priporočenih orodij (<https://fri-mdp.github.io/Podatkovni-zemljevid-2023/pz-viewer/>). Podatkovni zemljevid uporablja 28 imenskih prostorov oziroma obstoječih ontologij. Shema podatkovnega zemljevida sestavljajo razredi, objektne lastnosti, podatkovne lastnosti in poimenovani primerki.



Slika 38 Primer raziskovanja sheme povezanih podatkov iz kataloga OPSI za termina *lokacija* in *geografsko območje*.

S prostorskega vidika so v shemo podatkovnega zemljevida vključeni naslednji geoprostorski termini in relacije, pomembni za povezovanje geoprostorskih zbirke podatkov:

- geografsko območje – prostorska značilnost vira,
- geometrija - ima domeno lokacija: iri: <http://w3.org/ns/locn#geometry>,
- lokacija - prostorska regija ali imenovan kraj,
- centroid - je tudi opredeljen kot »object property«
- klasifikacija - sistem organizacije znanja (kos), ki se uporablja za razvrščanje naborov podatkov kataloga.

Podatkovne lastnosti:

- prostorska resolucija (v metrih),
- ali je podatkovni vir visoke vrednosti - lastnost, ki definira, ali je podatkovni vir visoke vrednosti glede na evropski pravni red, ki definira seznam podatkovnih virov visoke vrednosti,
- ali vsebuje občinske podatke.

Usmeritev za GURS: Ugotavljamo, da bo v tako imenovanem podatkovnem zemljevidu NIO potrebna ustrezna razširitev nabora razredov za domene prostorskih podatkov, ki jih upravlja GURS, oziroma sorodni deležniki, ki so skrbniki prostorskih podatkov in metapodatkov. Gre za pojme, ki so geoinformacijsko in geodetsko specializirani.

Platforma Evidenca2Go je bila razvita v okviru aktivnosti Tehnična in semantična prenova registrov in evidenc, za reševanje problematike preglednic v Excelu v državni upravi tako, da lahko zaposleni v javni upravi enostavno razvijajo lastne aplikacije brez poznavanja programiranja.

Platforma Klasifikacije2Go

Centralno vodenje klasifikacij (šifrantov) je po novem zagotovljena preko platforme Klasifikacije2Go, s čimer se lahko izognemo ponavljanju in vzdrževanju iste klasifikacije (šifrante) v različnih aplikacijah in na različnih mestih javne uprave. To pomeni, da se lahko ista klasifikacija (šifrant) uporablja v več aplikacijah, omogočen je tudi uvoz obstoječih postavk klasifikacij preko datotek CSV in spremljanje sprememb šifrantov. Mogoča je vzpostavitev novih klasifikacij (šifrantov) in migracija ter posodobitev obstoječih klasifikacij.

Usmeritev za GURS: v platformi Klasifikacije2Go (NIO) je treba preveriti celovitost vključenosti šifrantov GURS v to platformo in posodobiti ali na novo vzpostaviti potrebne šifrante.

Vidik uporabnikov geoprostorskih podatkov in metapodatkov

Trenutna uporaba geoprostorskih informacij v strukturi SDI pogosto ne dosega prvotnega namena projekta (Tim Trainor, 2023 ArcNews). Pri kreiranju poslovnih rešitev za premostitev vzeli med problemi, potrebami uporabnika (trga) in poslovnimi rešitvami, o geoprostorskih izdelkih izvajalca je pri razmišljanju potrebna predvsem osredotočenost na *potrebe uporabnika*.

Prepoznavanje *potreb* uporabnikov in njihovih poti je ključno. Prihaja do trenda »širjenje baze strank«, kar pomeni povečanja vrst uporabnikov geoprostorskih podatkov in storitev; tako humanih kot strojnih uporabnikov, in sicer so to: pomočniki AI, digitalni dvojčki / pametna mesta, senzorji, sistemi / podatkovni prostori / ekosistemi, podatkovni posredniki / odprti podatki / prostorska podatkovna infrastruktura, razvijalci, GIS analitiki, strokovnjaki sorodnih strok (na primer Building Information Model - BIM), kartografi, odločevalci in angažirana javnost.

Za vse te raznovrstne namene, potrebe in deležnike je treba zagotoviti podatke, ki so neposredno uporabni za *odkrivanje in ponovno uporabo*. To pomeni, da mora imeti *podatek naslednje lastnosti*:

- *Ustrezno pripravljeni podatki*, pripravljeni za analizo, ki omogočajo dostop do funkcij kakovosti, z dobro dokumentiranimi atributi in ključnimi polji za združevanje z drugimi podatki.
- *Dostopen v različnih oblikah* (formatih), ki izpolnjujejo zahteve širokega občinstva, vključno z ljudmi in stroji, prek API-jev, prenosov in uporabniku prijaznih aplikacij za raziskovanje netehničnih uporabnikov.
- *Podatek enostaven za razumevanje*, simboliziran in konfiguriran za hitro vizualizacijo.
- *Podatek enostaven za uporabo, samo opisljiv* (opremljen z metapodatki), interoperabilen in licenciran za ponovno uporabo.
- *Zanesljiv*, obstojen in optimiziran za razširljivost in zmogljivost.
- *Preprost za iskanje*, optimiziran za iskalnik in odkrit v samopostrežnem globalnem ekosistemu.

Usmeritev za GURS: Ugotavljamo, da je obstoječa dokumentacija preveč tehnična

- za potencialne nove uporabnike,
- za raziskovanje
- za »semantično prepoznavnost, vidnost in razumljivost« na spletu.

Za zbirke podatkov je treba odgovoriti na vprašanja o storitvah s podatki:

- *Ali ima podatek prej naštete lastnosti?*
- *Kateri podatki so na voljo?*
- *Kje najdem podatke?*
- *Kako zanesljivi so podatki?*
- *Kakšna je kakovost podatkov?*
- *Za kaj se lahko uporabljajo podatki?*

Najboljše prakse za pooblašcene ponudnike geoprostorskih podatkov se vežejo na zagotavljanje največje uporabnosti, »užitnosti« porazdeljenih geoprostorskih vsebin. V ta namen je treba zagotavljati *stalno izboljševanje v sintaktičnem in semantičnem smislu*, ki je osredotočeno na uporabnike, ki jim je treba zagotoviti celovito kakovost uporabniške izkušnje. To pomeni, da je treba zagotoviti *največjo mogočo stopnjo odkrivanja in ponovne uporabe* geoprostorskih in z njimi povezanih podatkov z zagotavljanjem možnosti za vključenost v sodelovalne sisteme porazdeljevanja podatkov na dinamične načine, ki temeljijo na odprtih standardih in najboljših praksah na svetovnega spleta.

Relativno nova (2003) »Direktiva o odprtih podatkih« omogoča evropskim ponudnikom in uporabnikom, da poiščejo rešitve in sprejmejo politiko, ki temelji na takšnih podatkih, ki omogočajo, da se spoprimejo z izzivi, od lokalnih do globalnih. Pri pripravi ponudbe *nove generacije geoprostorskih podatkov* in storitev je treba olajšati ustvarjanje *zelo kakovostnih metapodatkov* z ustreznimi razvojnimi orodji.

4.6.1 Orodja za semantično obravnavo geoprostorskih podatkov

V tem razdelku navajamo nekaj najbolj uveljavljenih orodij za semantično obravnavo geoprostorskih podatkov.

GeoNetwork

GeoNetwork je *Katalog za upravljanje prostorsko referenciranih virov*, ki ponuja zmogljive funkcije za urejanje metapodatkov in iskanje ter interaktivni pregledovalnik spletnih zemljevidov. Trenutno se uporablja v številnih pobudah za infrastrukturo prostorskih podatkov po vsem svetu. Je odprtokodno integracijsko orodje na Githubu za domeno prostorske (in odprte) podatkovne infrastrukture, ki vsebuje obsežno in preverjeno programsko opremo za urejanje in upravljanje katalogov metapodatkov. Uporablja jo približno 85% trenutnih nacionalnih končnih točk INSPIRE.

Obiralec (harvester)

Med drugimi operacijami GeoNetwork omogoča tudi pridobivanje »obiranje ali žetev« (harvesting) s tem, da ustvarite/posodobite metapodatke tako, da od strežnika zahtevate, da metapodatke 'povleče' od nekod. To orodje uporabljajo tudi v okviru JRC INSPRE projekta Portal, za katerega so razvili nova orodja za poročanje (december 2023). Metapodatki, ki so bili »nabrani«, se opremijo s statističnimi podatki o datumu nabiranja, številu nabranih zapisov metapodatkov (po bazah, po storitvah), o prenosljivih (downlad) podatkovnih nizih in nizih, ki so na ogled (View).

Razvijalci (28) GeoNetworka pri podjetju GEOCAT so rangirali prihodnji razvoj spletne tehnologije za semantično integracijo z njihovega vidika. Največje poudarke so dobili podpora za *OGC-API Records*, podpora za DCAT in GeoDCAT, GeoNetwork uporabniške vmesnike, izboljšana dokumentacija v splošnem ter posodabljanje tistih knjižnic, ki so dosegle konec delovanja. Sledili pa so podpora za *OGC-API Features*, Optimizacija načina, kot Google indeksira vaš katalog metapodatkov, nadaljnja izboljšava obstoječih uporabniških vmesnikov, posodobitev Elasticsearch na različico 8 ter integracija ChatGPT v GeoNetwork iskalnik.

4.7 Usmeritve za nadaljnje raziskave in razvoj na področju semantične paradigme

Izboljšanje skupne rabe in ponovne uporabe podatkov je vroča tema za izkoriščanje vse večje količine podatkovnih virov in napredka v integracijskih tehnologijah za informiranje o trajnostnem razvoju naših ozemelj.

Usmeritev za Geodetsko upravo: splošno vprašanje izmenjave in ponovne uporabe podatkov ustreza različnim bolj specifičnim deležem aktivnosti:

- upravljanje metapodatkov,
- uspešnost odkrivanja lastnih (GURS) izdelkov s strani potencialnih uporabnikov,
- spremljanje učinkovitosti uporabe podatkov pri uporabnikih,
- odkrivanje podatkov za lastno dejavnost: interno ali iz zunanjih virov.

Tehnologije semantičnega spleta (semantic web - SW), zlasti povezani podatki in grafi znanja (KG), predstavljajo priložnosti za doseganje teh deležev, zato so semantične tehnologije so že operativno sprejete v nekaterih projektih.

Nekateri pristopi se osredotočajo na oblikovanje povezanih metapodatkov in njihovo medsebojno povezavo z domeno grafov znanja KG, za izdelavo med domenskih katalogov ali za razvoj zmogljivosti odgovaranja na poizvedbe. Drugi pristopi uporabljajo tehnologije semantičnega spleta na ravni podatkov, bodisi za ustvarjanje naborov povezav bodisi za pretvorbo podatkov v povezane podatke.

Na evropski ravni se razvoj osredotoča na »graf znanja EuroSDR«, konsolidirati in deliti strokovno znanje in izkušnje v zvezi z izdelki geopodatkov v Evropi. Ti vključujejo izdelke javno pooblaščenih organov ali drugih zainteresiranih strani in tudi ad hoc izdelke, izpeljane za posebne namenske programe. Graf znanja bo podpiral objavo, skupno rabo in ponovno uporabo povratnih informacij geoprostorskih uporabnikov, ki jih izrazijo »uporabniki« zunaj geodetsko-kartografskih agencij ali »podatkovni strokovnjaki« znotraj proizvodnih linij.

Namen prototipa grafa znanja EuroSDR je prikazati pričakovano korist od KG na izdelkih geopodatkov kot tudi stroške za njeno doseganje. Služiti mora kot odskočna deska k večjemu in bolj ambicioznemu projektu. Posledično bo prototip oblikovan na podlagi dejanskih potreb, in sicer:

- konsolidacije opozoril uporabnikov na geoportalih geodetskih uprav (NMCA);
- izmenjava izkušenj v zvezi z zbiranjem fragmentov metapodatkov, pomembnih za zadevno aplikacijo, na primer analiza časovnega razvoja stavbe na podlagi topografskih podatkov o zgradbi;
- izmenjava izkušenj v zvezi z integracijo različnih virov za ustvarjanje grafa znanja na ravni podatkov;
- izmenjava izkušenj v zvezi z integracijo topografskih in statističnih podatkov.

Usmeritev za GURS: sodelovati pri nastajanju grafa znanja EuroSDR

Gre za graf znanja (KG) na ravni metapodatkov, ki je lahko koristen za informiranje izdelave KG na ravni podatkov pri nadaljnjem delu. Platforma na github: <https://github.com/EuroSDR-LDG> je prilagojena arhitekturnim zahtevam za deljenje besednjakov in odlomkov podatkov in kod.

Metapodatki po standardu ISO19139, ki opisujejo nekatere obstoječe podatkovne produkte znotraj kartografskih in katastrskih agencij, so bili preoblikovani v metapodatke RDF. Povezave med vozlišči podatkovnih izdelkov bi bile ustvarjene na podlagi kategorij tem in ključnih besed, povezanih s temi vozlišči. To bi bil osnutek grafa znanja o geo-digitalnih sredstvih.

Obstajajo tudi bogatejša povezave med izdelki od tistih, ustvarjenih z iskanjem po običajnih ključnih besedah. To so lahko na primer metode za povezovanje podatkov dveh različnih produktov. Pri tem je treba upoštevati tudi Večjezični slovar za področje geoinformatike (Lisec et al, 2023).

Vendar je na voljo več podrobnosti o teh podatkovnih izdelkih, kot je kodirano v obveznih metapodatkovnih elementih ISO19139: na primer podrobne *informacije o poreklu in specifikacije izdelka*. Te elemente klasično dokumentirajo ponudniki geopodatkov. Zdaj je jasno, da uporabniki, ki pridobijo tudi strokovnost in znanje o geopodatkovnih izdelkih, ki jih uporabljajo, lahko *ustvarijo povratne informacije*. Torej uporabniki podatkovnih izdelkov na tako imenovanih »policah izdelkov« (OTS) pridobijo strokovno znanje o teh izdelkih, ki ga lahko delijo z ustvarjanjem eksplicitnih povratnih informacij.

Usmeritev za GURS: Organizirati aktivnosti, da bi graf znanja o geopodatkovnih izdelkih združil strokovno znanje in izkušnje, povezane s podatkovnimi izdelki iz mrež ponudnikov podatkov preko EuroSDR, s podpiranjem objave, izmenjave in ponovne uporabe povratnih informacij uporabnikov geoprostora (GUF), ki jih izrazijo uporabniki zunaj kartografskih agencij, ter morebitnih izjav ponudnikov znotraj proizvodnih linij, vključno z izražanjem povezav med izdelki.

5 LITERATURA

Ariza-López, F. J., Rodríguez-Pascual, A., Lopez-Pellicer, F. J., Vilches-Blázquez, L. M., Villar-Iglesias, A., Masó, J., Díaz-Díaz, E., Ureña-Cámara, M. A., González-Yanes, A. (2021). An Analysis of Existing Production Frameworks for Statistical and Geographic Information: Synergies, Gaps and Integration. ISPRS International Journal of Geo-Information, 10(6), 374. DOI: <https://doi.org/10.3390/ijgi10060374>

- Attard, J., Orlandi, F., Auer, S. (2017). ExConQuer: Lowering barriers to RDF and Linked Data reuse. https://www.researchgate.net/publication/312889722_ExConQuer_Lowering_barriers_to_RDF_and_Linked_Data_re-use, accessed 20-May-23.
- Beek, W., Folmer, E. (2017). AN INTEGRATED APPROACH FOR LINKED DATA BROWSING. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W2*, 35–38. DOI: <https://doi.org/10.5194/isprs-archives-XLII-4-W2-35-2017>
- Berners-Lee, T. (1996). The world wide web - Past, present and future. *Journal of Digital Information*, 1(1).
- Booth, A., Noyes, J., Flemming, K., Moore, G., Tunçalp, Ö., Shakibazadeh, E. (2019). Formulating questions to explore complex interventions within qualitative evidence synthesis. *BMJ Global Health*, 4(Suppl 1), e001107. DOI: <https://doi.org/10.1136/bmjgh-2018-001107>
- Bucher, B., Folmer, Erwin, Brennan, R., Beek, Wouter, Hbeich, E., Würriehausen, F., Rowland, L., Maturana, R. A., Alvarado, E., Buyle, R., Di Donato, P. (2020). Spatial Linked Data in Europe: Report from Spatial Linked Data Sessions at Knowledge Graph in Action. https://www.euroedr.net/sites/default/files/uploaded_files/euroedr_publication_ndeg_73.pdf, accessed 16-May-23.
- Bénédicte Bucher, Pekka Latvala, Eva Bookjans, Nicolas Bus, Audrey Bouet, Erwin Folmer, Marjan Ceh, David Portoles, Luis Martinez Lorenzo, Dimitris Kotzinos, Hara Papadaki, Pasquale Di Donato. (2024). EuroSDR Knowledge Graph about Geodata Products in Europe. EuroSDR Linked Data Project Mid-Term Report. Official Publication No 76 of European Spatial Data Research.
- Çağdaş, V., Stubkjær, E. (2015a). Core immovable property vocabulary for European linked land administration. *Survey Review*, 47(340), 49–60. DOI: <https://doi.org/10.1179/1752270614Y.0000000093>
- Çağdaş, V., Stubkjær, E. (2015b). A SKOS vocabulary for Linked Land Administration: Cadastre and Land Administration Thesaurus. *Land Use Policy*, 49, 668–679. DOI: <https://doi.org/10.1016/j.landusepol.2014.12.017>
- Coumans, F. (2017). National Mapping as a Service: Ordnance Survey Ireland Shows New Perspectives. <https://www.gim-international.com/content/article/national-mapping-as-a-service>, accessed 16.3.2023.
- Crompvoets, J., Streilein, A., Masser, I. (2016). *How should Geodetske upraves adapt to alternative sources for Geodetske uprave data?* (European Spatial Data Research No. Official Publication No 65).
- Crompvoets, J., Vancauwenberghe, G. (2018). Governance of national spatial data infrastructures in Europe. *International Journal of Spatial Data Infrastructure Research*, 253–285. DOI: <https://doi.org/10.2902/1725-0463.2018.13.art16>
- Cruzes, D. S., Dybå, T. (2010). Synthesising evidence in software engineering research. In *International Symposium on Empirical Software Engineering and Measurement*.
- Debruyne, C., Clinton, É., Mcnerney, L., Nautiyal, A., O'Sullivan, D. (2016). *Serving Ireland's Geospatial Information as Linked Data*.
- The Directive on open data and the reuse of public sector data (European Commission 2019).
- Folmer, E., Beek, W., Rietveld, L., Namangaya A.H., B. M. (2018). Linked data viewing as part of the spatial data platform of the future. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W8). DOI: <https://doi.org/10.5194/isprs-archives-XLII-4-W8-49-2018>
- Folmer, E., Beek, W., Rietveld, L., Ronzhin, S., Geerling, R., Haan, D. den, Bui T.X. (2019). Enhancing the usefulness of open governmental data with linked data viewing techniques. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2019-January*. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083956109&partnerID=40&md5=de9a48645a167b7040a65824bc343194>.
- Folmer, E., Ronzhin, S., van Hillegersberg, J., Beek, W., Lemmens, R. (2020). Business rationale for linked data at governments: A case study at the Netherlands kadaster data platform. *IEEE Access*, 8, 70822–70835. DOI: <https://doi.org/10.1109/ACCESS.2020.2984691>
- Goodwin, J., Dolbear, C., Hart, G. (2008). Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. *Transactions in GIS*, 12, 19–30. DOI: <https://doi.org/10.1111/j.1467-9671.2008.01133.x>

- Hietanen, E., Lehto, L., Latvala, P. (2016). PROVIDING GEOGRAPHIC DATASETS AS LINKED DATA IN SDI. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B2*, 583–586. DOI: <https://doi.org/10.5194/isprs-archives-XLI-B2-583-2016>
- Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., Welch, V. A. (2022). Cochrane Handbook for Systematic Reviews of Interventions. <https://training.cochrane.org/handbook>, accessed 20-May-23.
- Huang, W., Raza, S. A., Mirzov, O., Harrie, L. (2019). Assessment and Benchmarking of Spatially Enabled RDF Stores for the Next Generation of Spatial Data Infrastructure. *ISPRS International Journal of Geo-Information*, 8(7), 310. DOI: <https://doi.org/10.3390/ijgi8070310>
- Huerta, Schade, Granell (Ed.) 2014. *Cadastral data integration through Linked Data agile2014_153*, accessed 17.3.2023.
- Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E. (2014). Linked Data Finland: A 7-star Model and Platform for Publishing and Reusing Linked Datasets. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, & A. Tordai (Eds.), *Lecture Notes in Computer Science. Semantic Web: ESWC 2014 Satellite Events* (Vol. 8798, pp. 226–230). Cham: Springer Nature. DOI: https://doi.org/10.1007/978-3-319-11955-7_24
- Infrastructure for Spatial Information in the European Community (INSPIRE). Directive 2007/2/EC of the European Parliament and the Council. (2007). <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32007L0002>, accessed 23-May-23.
- Jalali, S., Wohlin, C. (2012). Systematic literature studies: Database searches vs. backward snowballing. In *Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (pp. 29–38). DOI: <https://doi.org/10.1145/2372251.2372257>
- Jovanovik, M., Homburg, T., Spasić, M. (2021). A GeoSPARQL Compliance Benchmark. *ISPRS International Journal of Geo-Information*, 10(7), 487. DOI: <https://doi.org/10.3390/ijgi10070487>
- Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. <https://www.semanticscholar.org/paper/Procedures-for-Performing-Systematic-Reviews-Kitchenham/29890a936639862f45cb9a987dd599dce9759bf5>.
- Lanucara, S., Praticò, S., Modica, G. (2019). Harmonisation and Interoperable Sharing of Multi-temporal Geospatial Data of Rural Landscapes. In F. Calabrò, L. Della Spina, & C. Bevilacqua (Eds.), *Smart Innovation, Systems and Technologies: Vol. 100. New Metropolitan Perspectives: Local Knowledge and Innovation Dynamics Towards Territory Attractiveness Through the Implementation of Horizon/E2020/Agenda2030 - Volume 1* (Vol. 100, pp. 51–59). Cham: Springer International Publishing; Imprint: Springer. DOI: https://doi.org/10.1007/978-3-319-92099-3_7
- Ma, X. (2017). Linked Geoscience Data in practice: where W3C standards meet domain knowledge, data visualisation and OGC standards. *Earth Science Informatics*, 10(4), 429–441. DOI: <https://doi.org/10.1007/s12145-017-0304-8>
- Mai, G., Janowicz, K., Yan, B., Scheider, S. (2019). Deeply integrating Linked Data with Geographic Information Systems. *Transactions in GIS*, 23(3), 579–600. DOI: <https://doi.org/10.1111/tgis.12538>
- Mouzakitis, S., Papaspyros, D., Petychakis, M., Koussouris, S., Zafeiropoulos, A., Fotopoulou, E., Farid, L., Orlandi, Fabrizio, Attard, Judie, Psarras, J. (2017). Challenges and opportunities in renovating public sector information by enabling linked data and analytics. *Information Systems Frontiers*, 19(2), 321–336. DOI: <https://doi.org/10.1007/s10796-016-9687-1>
- Nguyen, T., Jump, A., Casey, D. (2022). Emerging Tech Impact Radar: 2023: Gartner Research Excerpt. <https://emtemp.gcom.cloud/ngw/globalassets/en/doc/documents/767259-emerging-tech-impact-radar-2023-excerpt.pdf>, accessed 10-May-23.
- O'Leary, D. E. (2008). Gartner's hype cycle and information system research issues. *International Journal of Accounting Information Systems*, 9(4), 240–252. DOI: <https://doi.org/10.1016/j.accinf.2008.09.001>
- Octaviano, F., Felizardo, K., Maldonado, J., Fabbri, S. (2015). Semi-automatic selection of primary studies in systematic literature reviews: is it reasonable? *Empirical Software Engineering*, 20, 1898–1917.
- Owusu-Banahene, W. (2018). Making attributes from the Linked Open Data (LOD) cloud a part of Spatial Data Infrastructures (SDIS) for thematic mapping. *AFRREV STECH: an International Journal of Science and Technology*, 7(1), 20–33. DOI: <https://doi.org/10.4314/stech.v7i1.3>

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, Tianjing, Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, James, Tricco, A. C., Welch, V. A., Whiting, P., Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical Research Ed.)*, 372, n71. DOI: <https://doi.org/10.1136/bmj.n71>
- Person, T., Ruiz-Rube, I., Mota, J. M., Cobo, M. J., Tselykh, A., Doderio, J. M. (2021). A Comprehensive Framework to Reinforce Evidence Synthesis Features in Cloud-Based Systematic Review Tools. *Applied Sciences*, 11(12), 5527. DOI: <https://doi.org/10.3390/app11125527>
- Polleres, A., Kamdar, M. R., Fernández, J. D., Tudorache, T., Musen, M. A. (2020). A more decentralised vision for Linked Data. *Semantic Web*, 11(1), 101–113. DOI: <https://doi.org/10.3233/SW-190380>
- Richardson, W. S., Wilson, M. C., Nishikawa, J., Hayward, R. S. (1995). The well-built clinical question: a key to evidence-based decisions. *ACP Journal Club*, 123(3), A12. DOI: <https://doi.org/10.7326/ACPJC-1995-123-3-A12>
- Ronzhin, S., Folmer, E., Maria, P., Brattinga, M., Beek, W., Lemmens, R., van't Veer, R. (2019). Kadaster knowledge graph: Beyond the fifth star of open data. *Information (Switzerland)*, 10(10). DOI: <https://doi.org/10.3390/info10100310>
- Rowland, A., Folmer, E., Beek, W. (2020). Towards Self-Service GIS—Combining the Best of the Semantic Web and Web GIS. *ISPRS International Journal of Geo-Information*, 9(12), 753. DOI: <https://doi.org/10.3390/ijgi9120753>
- Rowland, A., Overeem, J., Folmer, E., & Yaman B., Sherif M.A., Ngomo A.-C.N., Haller A. (Eds.) 2021. *Demo: GeoDataWizard for linked spatial data creation*. : Vol. 2977: CEUR-WS. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117757990&partnerID=40&md5=7acbf3e48131d15b8c0bb13785e6ded9>.
- Rowland, Alexandra, Folmer, Erwin, Beek, Wouter, Wenneker, R. (2022). Interoperability and Integration: An Updated Approach to Linked Data Publication at the Dutch Land Registry. *ISPRS International Journal of Geo-Information*, 11(1), 51. DOI: <https://doi.org/10.3390/ijgi11010051>
- Saavedra, J., Vilches-Blázquez, L. M., & Boada, A. (Eds.) 2014. *Cadastral data integration through Linked Data*.
- Schleidt, K., O'Grady, M., Grellet, S., Feliachi, A., van der Schaaf, H. (2020). ELFIE - The OGC Environmental Linked Features Interoperability Experiment. In Birukou & Athanasiadis (Eds.), *IFIP Advances in Information and Communication Technology. Environmental Software Systems. Data Science in Action* (Vol. 554, pp. 188–193). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-39815-6_18
- Schmidt, L., Olorisade, B. K., McGuinness, L. A., Thomas, James, Higgins, Julian. (2021). Data extraction methods for systematic review (semi)automation: A living systematic review. *F1000Research*, 10, 401. DOI: <https://doi.org/10.12688/f1000research.51117.1>
- Tim Berners- Lee. (2010). 5-star Open Data. <https://5stardata.info/en/>, accessed 15-May-23.
- United Nations Expert Group on Land Administration, Management. (2019). Framework for Effective Land Administration: A reference for developing, reforming, renewing, strengthening or modernising land administration and management systems.
- Van den Brink, L., Barnaghi, P., Tandy, J., Atemez, G., Atkinson, R., Cochrane, B., Fathy, Y., García Castro, R., Haller, A., Harth, A., Janowicz, K., Kolozali, Ş., van Leeuwen, B., Lefrançois, M., Lieberman, J., Perego, A., Le-Phuoc, D., Roberts, B., Taylor, K., Troncy, Raphaël. (2018). Best practices for publishing, retrieving, and using spatial data on the Web. *Semantic Web*, 10(1), 95–114. DOI: <https://doi.org/10.3233/SW-180305>
- Velitchkov, I. (2021). Endorse 5 star deployment scheme for Linked Data: Barriers to the uptake of Linked Data. Endorse Conference <https://op.europa.eu/en/web/endorse-2021/programme>. https://op.europa.eu/documents/10120270/10133951/VELITCHKOV_presentation_Endorse+5+stars+deployment+scheme+for+Linked+Data.pdf/74ecc573-d874-37f3-f133-13968b1cc4eb?t=1616618192854, accessed 22-May-23.

- Vilches-Blázquez, L. M., Saavedra, J. (2019). A framework for connecting two interoperability universes: OGC Web Feature Services and Linked Data. *Transactions in GIS*, 23(1), 22–47. DOI: <https://doi.org/10.1111/tgis.12496>
- Wiemann, S., Bernard, L. (2016). Spatial data fusion in Spatial Data Infrastructures using Linked Data. *International Journal of Geographical Information Science*, 30(4), 613–636. DOI: <https://doi.org/10.1080/13658816.2015.1084420>
- Wohlin, C., Kalinowski, M., Romero Felizardo, K., Mendes, E. (2022). Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. *Information and Software Technology*, 147, 106908. DOI: <https://doi.org/10.1016/j.infsof.2022.106908>
- Jesenko, M., Lozej, M., Kern Pipan, K., Godec, P., Tanko, V., Žagar, L., Pretnar Žagar, A., Đukić, N. in Zupan, B. 2022. Semantični analizator – razvoj programskega okolja za algoritmično obdelavo slovenskih besedil. *Uporabna informatika*. 30, 2 (jun. 2022). DOI:<https://doi.org/10.31449/upinf.156> Zagorje ponudba
- Gajšek, N. (2022). Terminološka načela in praksa njihovega upoštevanja v terminografiji - predstavitev za Terminološko delovno skupino. Zapis sestanka 28. 2. 2022
- Jeraj, J., Puhar, M., Fuat, F., 2022. Podatkovni prostori, pogled iz Slovenije. Gaia-x Hub Slovenia
- Ross, R., G., 2020. Enabling Your Data to Speak the Language of the Business, 2nd ed.
- Brooks, P., Ellams, L., Kilburn, T., Roussell, R., Hughes, T., “5 best practices for data enhancement and data cleansing”.
- Balijs, D., Lavbič, D., 2015. Obogatitev podatkov v relacijski podatkovni bazi z Linked Data slovarji. Magistrska naloga, Fakulteta za računalništvo in informatiko, Univeza v Ljubljani.
- Repnik, L., Ojstršek, M., Implementacija infrastrukture za podporo pomenskemu rudarjenju v poslovnih informacijskih sistemih. Magistrsko delo, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru.
- Tričkovič, I., Uvajanje semantične interoperabilnosti v javni upravi. 2019. *Uporabna informatika letnik XXVII*, št. 2.
- Lisec, A., et all. 2023. Večjezični slovar za področje geoinformatike. Končno poročilo projekta CRP V2-2133, februar 2023.

<https://docs.ogc.org/is/17-003r2/17-003r2.html>

<https://www.techtarget.com/searchapparchitecture/definition/Resource-Description-Framework-RDF>

<https://graphql.org/learn/thinking-in-graphs/>

https://golden.com/wiki/Apollo_Graphql-VWJ5DJG

[https://chengzhizhao.com/here-is-what-i-learned-using-apache-airflow-over-6-years/#:~:text=Apache%20Airflow%20is%20undoubtedly%20the,ETL\)%20is%20a%20milestone.](https://chengzhizhao.com/here-is-what-i-learned-using-apache-airflow-over-6-years/#:~:text=Apache%20Airflow%20is%20undoubtedly%20the,ETL)%20is%20a%20milestone.)

6 PRILOGE

6.1 Vprašalnik za deležnike semantične analize

26. 5.	Geo vprašalnik (za STN)	Ver: 0.5
0 Okolje zajema in hrambe izvirnega podatkovnega seta		
0.1 Opis okolja		
0.1.1	Opis skupnega podatkovnega okolja	
	Opišite, kateri podatkovni seti so bili shranjeni v skupnem okolju	
	Kateri so bili skupni grafični elementi/podatki , ki so si jih seti delili	
	Ali obstajajo kakšni povezovalni elementi (ključi) za vzpostavljanje vsebinskih povezav med seti (relacije med tabelami)	
	Ali obstajajo vsebinski atributi, ki so bili prepisani iz druge osnove (sloja, tabele)	razlege, ki se uporabljajo v praksi (npor Šarlah priročnik)
	Ali obstaja dokumentacija o tem okolju , kdo z njo razpolaga, ali je morda evidentirana kot dokumentarno gradivo	
0.1.2	Opis skupnega tehničnega okolja	
	Opišite, osnovne značilnosti tehničnega okolja (baza, vmesnik, ...), vključno s podatki o dobaviteljih oz. razvijalcih in vzdrževalcih	
	Ali obstaja dokumentacija o tem okolju, kdo z njo razpolaga, ali je morda evidentirana kot dokumentarno gradivo	dok o tehničnem okolju
0.1.3	Opis dokumentacije o celotnem okolju	
	Ali obstaja dokumentacija, ki se nanaša na vse podatkovne sete oz. na celotno okolje	širša slika okoliščin za sklepanje za samodejno strojno sklepanje
	pravne podlage	
	priročniki	
	interna navodila	
	študije	
	drugo ...	
1 Izvirni podatkovni set		
1.1 Opis konceptualnega dela podatkovnega seta		
1.1.1	Namen podatkov in način njihove uporabe	
	Opišite namen podatkov, načine njihove uporabe	
	V kakšnih procesih so se ti podatki uporabljali ? Na kakšen način?	kam bi lahko pomensko povezali....
1.1.2	Pravni okvir	
	Na podlagi katerega predpisa /seznama predpisov/ je bila ta evidenca ustvarjena	Temeljni predpis, dodatni predpisi (javni/interni) Kje in na kak način so dostopni.
	Kateri predpisi urejajo urejanje in uporabo te evidence	opredelitve pravnih pojmov na področju
	Ali obstajajo odstopanja od predpisov v praksi (oz. ostale posebnosti)	okoliščine osnovnega primera uporabe

	Ali se je v zvezi z setom spreminjal temeljni predpis? Kdaj? V čem bistvene spremembe? Posledice sprememb na način upravljanja, na podatke?	spremljanje sprememb (metodolške razlike v vnašanju podatkov) snemalna pravila
	Ali so se v zvezi z setom spreminjali drugi predpisi (interni, navodila)? Kdaj? V čem bistvene spremembe? Posledice sprememb na način upravljanja, na podatke?	spmemba RKS, ponovljivost, reverzibilnost,
1.1.3.	Zgodovinski okvir	
	Kje in zakaj je prišlo do potrebe za nastanek/spremembo te evidence.	vzrok sprememb
	Ali lahko opišete okoliščine, pred sprejemom zakonodaje, ki so vplivale na prvo vzpostavitev.	okoliščine porve vzpostavitve
	Ali obstajajo podatkovni seti/evidence, ki jih je obravnavan set/evidenca zamenjala, nadomestila?	pretekli podatkovni seti (primer zajem s starih kart za novo bazo)
1.1.4.	Tematika	
	Pod katero INSPIRE tematiko bi spadal ta sloj	klasifikacija izvedena
	Kakšna je vsebinska ali fizična povezljivost tega sloja z drugimi podatkovnimi sloji podobne tematike	povezljivost (npr na stavbe, parcele), položajna povezljivost
	Kakšna je vloga podatkovnega sloja na nivoju države	vloga za državo
	Katerim sorodnim evidencam iz EU/v svetu bi lahko bila podobna ta evidenca	mednarodna podobnosti za sloj
1.1.5.	Časovni okvir	
	Kdaj je bila evidenca ustvarjena	
	Kdaj se je vzpostavitev začela in kdaj je bila vzpostavljena prva "uradna" različica? Kdaj je bila vzpostavljena prva popolna različica, ki je obsegala celotno predvideno ozemlje z vsemi predvidenimi podatki?	začetek, popolnost, datum zajema, datum veljavnosti, obdobje zajema
	Ali in kdaj je bila zaključena	prenehanje aktivne uporabe - prehod v arhivsko rabo
	Obdobje vodenja evidence je torej ...	
1.1.6	Vzpostavitev in vzdrževanje evidence	
1.1.6.1	Vzpostavitev evidence	
	Kateri viri podatkov so bili uporabljeni za vzpostavitev podatkovnega seta	
	Tehnična metodologija (tehnike) zajema podatkov za vzpostavitev - na kakšen način so podatki nastali (izvor, časovni okvir, uporabljeni viri za izdelavo podatka)	vhodni viri (anlogni?), tehnika/tehnologija, snemalna pravila
1.1.6.2	Vzdrževanje evidence	
	Kaj so (bili) razlogi za vnos sprememb v podatkovni set oz. evidenco	Osnova za odločitev za zajem - sprožitev postopka
	Kateri viri so bili uporabljeni za vnos sprememb?	
	Kako so (bile) spremembe dokumentirane (v sami zbirki podatkov in morebiti tudi v dokumentnem sistemu)?	dokumentiranje postopka zajema
	Opišite postopke vzdrževanja in spreminjanja evidence	posodabljanje podatkov

	Ekranske slike programske opreme / Video postopka vzdrževanja	primeri starega sistema - posnetki-dokumentiranja za arhiv (postopkov, funkcionalnosti sistema)
	Pogostost vzdrževanja / vnosa sprememb	frekvenca sposodabljanja
1.1.6. Najpogostejše poizvedbe		
	Kdo so tipični uporabniki podatkov seta/evidence	opisovanje za razumevanje
	Poizvedbe Končnih uporabnikov	
	Primer tipičnih izpisov iz te evidence	Vseh, ki so pomembni ali celo predpisani? (slednje zabeležiti, če obstaja)
	Ali obstajajo predloge za izpis evidence	
	obstoječi izpisi	
	Ekranske slike programske opreme / Video postopka poizvedovanja	
	Ali obstajajo SQL stavki, ki omogočajo ponovno izdelavo izpisov oz. Logični potek izvajanja poizvedbe (UML shema)	
	Poizvedbe notranjih uporabnikov / vzdrževalcev	
1.1.7 Druga Opisna Dokumentacija		
	Opišite dokumentacijo, ki opisuje prostorski podatkovni set oz. je specifična zanj (npr. Uporabniški priročniki, podobne prakse, metodološka pravila, znanstveni članki, objave...)	To je dopolnilo enemu od uvodnih vprašanj
1.2 Opis temeljnih lastnosti izvirnega podatkovnega seta		
1.2.1 Geometrija prostorskih podatkov		Ustvarjalec ali imetnik podatkovnega seta opiše geometrijske lastnosti
	V katerem koordinatnem sistemu se upravlja ta set prostorskih podatkov	izvirnega prostorskega podatkovnega seta.
	Ali je v topološki odvisnosti od drugih slojev (kakšni)	
1.2.2 Struktura seta prostorskih podatkov		
	Ali nam lahko pokažete objektni katalog prostorskega podatkovnega seta	
	Opišite in z diagrami predstavite relacije med tabelami prostorskega podatkovnega seta	
	Opišite logično strukturo atributov in njihovo zalogo vrednosti	kako so se te vsebine zajemale, iz katerih virov. Ali so evidence celovite, ali obstajajo napake, netočnosti...
1.2.3 Metapodatki		Ustvarjalec ali imetnik podatkovnega seta pripravi metapodatkovni opis
	Ali imate metapodatke v obliki INSPIRE	skladen z metapodatkovnimi standardi ali v posebnih primerih metapodatke zapiše v že definirani obliki.
	Če ne, ali imate metapodatke v ISO 19115 obliki	
	če ne, ali imate metapodatke v kakršni koli drugi obliki?	
1.2.4 Vizualizacija in kartografski znakovni katalog		
	Ali obstaja za ta podatkovni set določena simbologija oz. priporočena simbologija (kartografski znakovni katalog)	

	Ali je potrebno še kaj drugega za ustrezno vizualizacijo podatkovnega seta?	
	Če obstaja ali jo je možno pretvoriti v SLT obliko?	
1.3	Opis fizičnega zapisa	
1.3.1	Izvorni formati & izvorni medij	
	V kakšnem formatu se podatkovni set hrani v produkcijskem okolju?	
	V kakšnem formatu se podatkovni set izdaja v distribucijskem okolju?	