



Delovna verzija 0.5



16.4.2018

ZGODOVINA DOKUMENTA

Datum	Verzija	Opis	Avtor
16.2.2017	0	Opis metapodatkovnega okolja in zakonskih podlag	Mag. Igor Tričkovič Rifelj
02/07/17	0.1	Dodana poglavja: Uvod, pristop, shema, objava	Marko Ambrož
04/08/17	0.1 a	Dopolnitve in redakcija poglavij, miselni vzorec pojmovnika	Dr. Miro Lozej
06/08/17	0.1 b	Dodani viri, besedni oblak	Marko Ambrož
16.04.18	0.2	Dodana poglavja glede podpore vnosa in vzdrževanja CNB	Marko Ambrož, Dr. Miro Lozej

Kazalo vsebine

STRATEGIJA UPRAVLJANJA SEMANTIČNE INTEROPERABILNOSTI	1
UVOD	3
PRISTOPI	4
Naloge upravljavca centralne infrastrukture na področju upravljanja metapodatkov	4
PLATFORMA ZA UPRAVLJANJE SEMANTIČNE INTEROPERABILNOSTI	6
Konceptualni nivo	6
Logični nivo.....	7
Fizični nivo.....	7
UVEDBA SISTEMA UPRAVLJANJA Z METAPODATKI.....	8
ZEUS, zakonska ureditev področja	8
Izvajanje obstoječih predpisov na področju evidentiranja podatkovnih zbirk.....	10
Razvoj orodja za pomoč pri upravljanju centralnega besednjaka CNB.....	10
Določanje upravljavcev in obdelovalcev podatkovnih zbirk	11
Javna objava metapodatkov in slovarjev	11
VIRI	11
PRILOGA A	13

UVOD

Razvoj digitalne uprave, uvajanje poslovne inteligence in obdelave podatkov velikega obsega, objava odprtih podatkov, sočasno zagotavljanje vgrajene zasebnosti, zagotavljanje skladnosti po GDPR in vgrajena informacijska varnost, so vse aktivnosti, ki temeljijo na podatkih. Vendar ni dovolj, da imamo podatke, pomembno je tudi pravilno enako razumevanje pomena pri vseh uporabnikih.

V preteklosti je bil dosežen velik napredek na nivoju tehnične interoperabilnosti. Izdelani so bili tehnični interoperabilnostni gradniki in horizontalne funkcije, s katerimi smo dobili nagrade svetovnega nivoja. Vendar bo v prihodnosti potrebno posvetiti tudi več inovativnih pristopov in vložka, na področju semantične interoperabilnosti.

Večino razvoja informacijskih rešitev se v državni upravi naroča pri zunanjih dobaviteljih, programerskih hišah. Posledično to pomeni, da v procesu razvoja nastopajo številne razvojne skupine z poznavanjem ozkih, silosnih vsebinskih področij. Nastale rešitve za posamezna področja ostajajo v okviru (domenskih) silosov.

S pristopi, ki jih opredeljujemo in uvajamo s pomočjo te strategije, želimo nastalo situacijo preseči, in ponuditi platforme in orodja za učinkovit pregled nad obstoječimi podatkovnimi viri (predvsem registri), njihovimi strukturami in enostavno gradnjo učinkovito povezanih podatkovnih modelov. Ob tem želimo hkrati tudi postaviti skupne standarde na celi vertikalni področja semantične interoperabilnosti. Vertikalo razumemo od ontologij na vrhu preko metapodatkovnih slovarjev, jedrnih modelov, logičnih domenskih modelov do najnižjega nivoja, ki ga predstavljajo fizični podatkovni modeli, sestavni deli informacijskih rešitev.

S pomočjo orodij in postopkov opredeljenih v strategiji prioritarno rešujemo:

- enkratni zapis (Once Only Principle),
- učinkovitejše in standardizirano načrtovanje modelov informacijskih rešitev (pri novih in reinžiniringu obstoječih),
- večjo stopnjo zanesljivosti in kvalitete podatkov, ki se izmenjujejo med sistemi (pravilo ene resnice – enako razumevanje pomena podatkov pri vseh uporabnikih),

Če smo v preteklosti dosegli velik napredek na nivoju tehnične interoperabilnosti, izdelali gradnike, s katerimi smo dobili nagrade svetovnega nivoja, bo v prihodnosti potrebno posvetiti bistveno več inovativnih pristopov in truda, na področju semantične interoperabilnosti.

Vendar aktivnosti omejene strogo na semantično interoperabilnost, še ne zagotavljajo vseh potrebnih podlag za doseganje ciljev digitalne uprave. Vzpostavitev semantične interoperabilnosti je potrebno izvesti na način, da ne bo sama sebi namen, temveč, da bo v pomoč, če ne osnovno orodje za razvoj novih sistemov in kot temelj v okviru objave odprtih podatkov in orodij, za širšo strokovno javnost.

Za izvedbo nalog opredeljenih v Strategiji predlagamo centralnega nosilca, skrbnika, Ministrstvo za javno upravo, v vlogi tehničnega upravljavca semantične podatkovne platforme. Hkrati koordinatorske dejavnosti med posameznimi upravljavci vsebinskih področij, informacijskih rešitev, registrov in evidenc.

Dokument Strategija upravljanja semantične interoperabilnosti je nastal v okviru projekta Tehnična in semantična prenova registrov in evidenc.

PRISTOPI

Glede na to, da so standardi in orodja za ožjo semantično interoperabilnost razviti in zreli za uporabo, je najpomembnejše vprašanje, ki ga Strategija za upravljanje metapodatki obravnava, kako vzpostaviti in vzdrževati metapodatkovne zbirke skozi čas. Kot vedno obstaja več možnih poti in pristopov, iščemo kombinacijo najučinkovitejših.

1. Osnovni pristop so zakonske podlage, ki bi upravljavcu centralne infrastrukture omogočile obvezo upravljavcev podatkovnih zbirk, da jih evidentirajo v centralne evidence, orodja.

Ta pristop je verjetno potreben, ni pa nujno zadosten, saj pogosto, če ne praviloma, nastane situacija, v kateri se pričakuje, da se bodo zakonske obveze izpolnile brez virov (človeških, finančnih).

2. Pristop, ki bi bil komplementaren prvemu, bi temeljil na razširitvi sektorskih in centralnih projektov na način, da bi se del analize in razvoja posvetil semantiki in vzdrževanju metapodatkovnega področja, ki ga posamezni projekt obravnava.
3. Tretji prav tako dopolnilni pristop pa se nanaša na mehanizme in postopke v okviru procesov gostovanja na centralni informacijski infrastrukturi s katerimi bi težili k čim bolj avtomatiziranemu polnjenju centralnih metapodatkovnih zbirk na podlagi standardiziranih opisov, bodisi v dokumentaciji ali izvorni kodi izdelkov.
4. Četrti ne nujno tudi pristop na zadnjem mestu je uporaba obstoječih mehanizmov in zakonskih podlag na področju obveznega evidentiranja zbirk osebnih podatkov ter zbirk odprtih podatkov.

Pri vsakem od naštetih pristopov bi lahko bila koristna in v pomoč orodja strojnega oziroma globokega učenja, pristopi obdelave podatkov (tekstualnih) velikega obsega.

Naloge upravljavca centralne infrastrukture na področju upravljanja metapodatkov

Ministrstvo za javno upravo v okviru nalog na področju nalog interoperabilnosti, uvajanja gradnikov in skupnih horizontalnih funkcij zagovarja princip tako imenovanega enkratnega zapisa podatkov (Once Only Principle – podatek se vpiše le enkrat). Da bi se lahko ta princip uveljavil, je potrebno omogočiti projektnim in razvojnim skupinam dostop do metapodatkovnih opisov obstoječih podatkovnih virov (registrov in evidenc). Za vzpostavitev skupnega repozitorija metapodatkovnih opisov, je potrebno zagotoviti sodelovanje organov oziroma obvezo registracije podatkovnih virov (preko različnih mehanizmov), ter seveda tudi z druge strani pri razvoju ali nadgradnjah informacijskih rešitev obvezo projektnim skupinam in razvojnim ekipam po upoštevanju skupnega metapodatkovnega repozitorija. Še posebej je potrebno nasloviti obveznost pripravljavcem, zakonodajnih besedil, da upoštevajo obstoječe podatkovne vire. S ciljem, da se uvajajo zajem novih podatkov samo v primerih, ko zanje še ne obstajajo podatkovni viri. in še posebej, da upoštevajo načelo sorazmernosti pri zahtevah po podatkih v različnih upravnih postopkih, bodisi na strani vlagateljev, bodisi na strani poročevalcev.

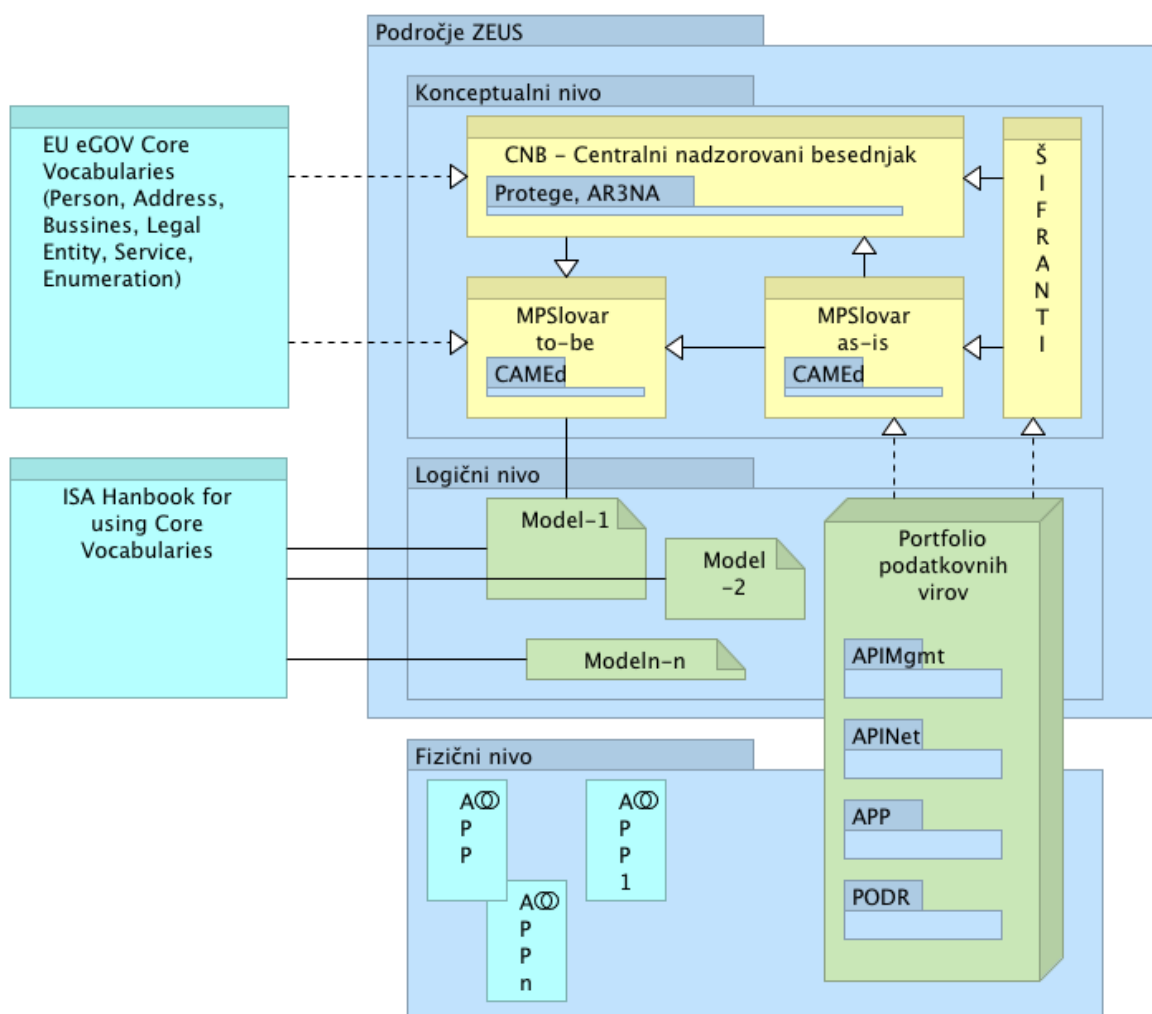
V predhodni zvezi je potrebno vzpostaviti tudi podlago, da se upravljavcem informacijskih sistemov, ki zagotavljajo podporo postopkom omogoči dostop do standardnega okvira podatkovnega nabora (jedrni – core podatkovni modeli, evropski in nacionalni), ki ga potrebujejo za izvajanje postopkov v okviru ZUP.

V nadaljevanju se za potrebe interoperabilnosti in zagotavljanju konsistentnosti digitalne uprave zahteva dosledno upoštevanje skupnih šifrantov in klasifikacij. Skupni šifranti in klasifikacije se

vključujejo v skupne gradnike kot na primer Razvid institucij Javnega Sektorja (zahteva nastopi v veljavo, ko bodo zagotovljene tehnične možnosti).

Na področju skupnih horizontalnih rešitev, kot na primer državni portal e-Uprava in eVEM je zelo pomembno, da se določijo upravljavci/lastniki ne samo podatkovnih zbirk temveč tudi podatkov znotraj horizontalnih rešitev, saj tipično ministrstva in organi v sestavi nastopajo s svojimi pod-zbirkami v skupni zbirki, katere tehnični skrbnik je sicer MJU. Lastniki in upravljavci pa v tej vlogi nastopajo samo v delnem naboru zapisov skupne tehnične zbirke, ki se nanašajo na njihove postopke, objave in obdelave. V ta namen je v pomoč pri določanju lastništva posameznih podatkovnih zapisov tako imenovana matrika razmejitve odgovornosti, ki na postopek natančno določa lastnika oziroma upravljavca pod-zbirke podatkov/zapisov/dokumentov. To vprašanje naslavlja ZEUS še posebej zaradi morebitnih nesporazumov oziroma poenostavljenega razumevanja. Tako je na primer potrebno odločiti, če je MJU dolžan posredovati tretjim osebam podatke, ki so hranjeni na njegovih diskovnih pogonih/nosilcih podatkov, čeprav je dejanski skrbnik/lastnik teh podatkov neka druga institucija.

PLATFORMA ZA UPRAVLJANJE SEMANTIČNE INTEROPERABILNOSTI



Konceptualni nivo

CNB - Centralni nadzorovan besednjak: Centralni nadzorovan besednjak je eksplicitno določen seznam pojmov, ki ga upravlja organ za registracijo (Ministrstvo za javno upravo). Vse besede v nadzorovanem besednjaku morajo imeti jasno, nedvoumno in ne-redundantno definicijo. Najmanj, kar mora centralni nadzorovan besednjak zagotavljati, je:

1. če se nek termin uporablja za različne koncepte in v različnih kontekstih, mora biti njegovo ime v določenem kontekstu eksplicitno kvalificirano,
2. če se uporablja več terminov, ki spadajo v isti koncept, mora biti eden izmed njih določen kot prednostni, preostali pa kot sinonimi, psevdonimi in podobno.

V nadzorovanem besednjaku so termini lahko organizirani tudi v hierarhično strukturo. Vsak termin je lahko v enem ali več odnosih nadrejenosti ali podrejenosti do drugih terminov. Dovoljena tudi mreža, kar pomeni, da ima lahko termin več nadrejenih terminov.

V nadzorovanem besednjaku lahko odnosi med koncepti vključuje tudi asociativne (nehierarhične) odnose. Beležene pa so tudi druge lastnosti (izvorno vsebinsko področje) odnosi (sinonimi, homonimi, šifriranje s šifranti...)

MPSlovar – Meta Podatkovni slovar: je v bistvu podatkovna shema, izdelana v standardu CAM (Content Assembly), ki vsebuje podatkovne opise virov, ki so na voljo, dostopni informacijskim rešitvam. V tem dokumentu predlagamo vzpostavitev dveh metapodatkovnih slovarjev, slovar „As-is“ z dejanskimi obstoječimi metapodatkovnimi opisi in slovar „To-be“ z metapodatkovnimi opisi, ki so standardizirani (Camel case).

Šifranti: Šifranti in klasifikacije so referenčni podatki, ki se vodijo centralno in so pomembni za več informacijskih rešitev.

Skupni šifranti imajo ključno vlogo pri digitalizaciji uprave, saj so tako pomenski kot vsebinski vezni element med različnimi sistemi, domenskimi področji.

Logični nivo

Sestoji iz posamičnih jedrnih podatkovnih modelov in Portfolija podatkovnih storitev.

Jedrni podatkovni model: Jedrni podatkovni model je ponovno uporabljen in razširljiv podatkovni model, ki zajema temeljne značilnosti entitet na kontekstualno nevtralen način. Organi javne uprave lahko uporabljajo in razširijo enotne jedrne podatkovne modele na naslednje načine:

- **Razvoj novih aplikacij:** jedrne podatkovne modele se lahko uporabijo kot privzeto izhodišče za oblikovanje konceptualnih in logičnih podatkovnih modelov novo razvitih informacijskih sistemov.
- **Izmenjava podatkov med sistemi:** Jedrni podatkovni modeli lahko predstavljajo temelj kontekstualno specifičnih podatkovnih modelov za izmenjavo podatkov med obstoječimi informacijskimi sistemi.
- **Podatkovna integracija:** Jedrni podatkovni modele se lahko uporablja za integracijo podatkov, ki prihajajo iz različnih heterogenih podatkovnih virov.
- **Objava odprtih podatkov:** Jedrni podatkovni modeli se lahko uporabljajo kot temelj za skupni format za izvoz podatkov iz podatkovnih zbirk.

Portfolio podatkovnih storitev: je sistem, ki vzpostavlja relacije med aplikacijami, ki uresničujejo, izpostavljajo podatkovne storitve, njihove definicije, lastnike in upravljavce, pogoje uporabe, sheme in vizualizacijo medsebojnih klicev.

Na podlagi podatkov sistema portfolija podatkovnih storitev in metapodatkovnega slovarja se gradi podatkovni repozitorij in njegov prikaz v obliki podatkovnega zemljevida.

Fizični nivo

Predstavljajo informacijske rešitve s svojimi fizičnimi podatkovnimi zbirkami: APP-1 ... APP-N. Pretvorba logičnega podatkovnega modela v fizični se privzeto izvaja preko programske storitve, ki logičnemu modelu doda režijske attribute in arhivske tabele, ter pripadajoče prožilce, ki omogočajo zapis podatkovnih sprememb (vpis, sprememba, izbris).

Če je logični podatkovni model neodvisen od tehnološke izvedbe podatkovne zbirke, se fizični podatkovni model pripravi na točno določeno tehnološko izvedbo podatkovne zbirke.

UVEDBA SISTEMA UPRAVLJANJA Z METAPODATKI

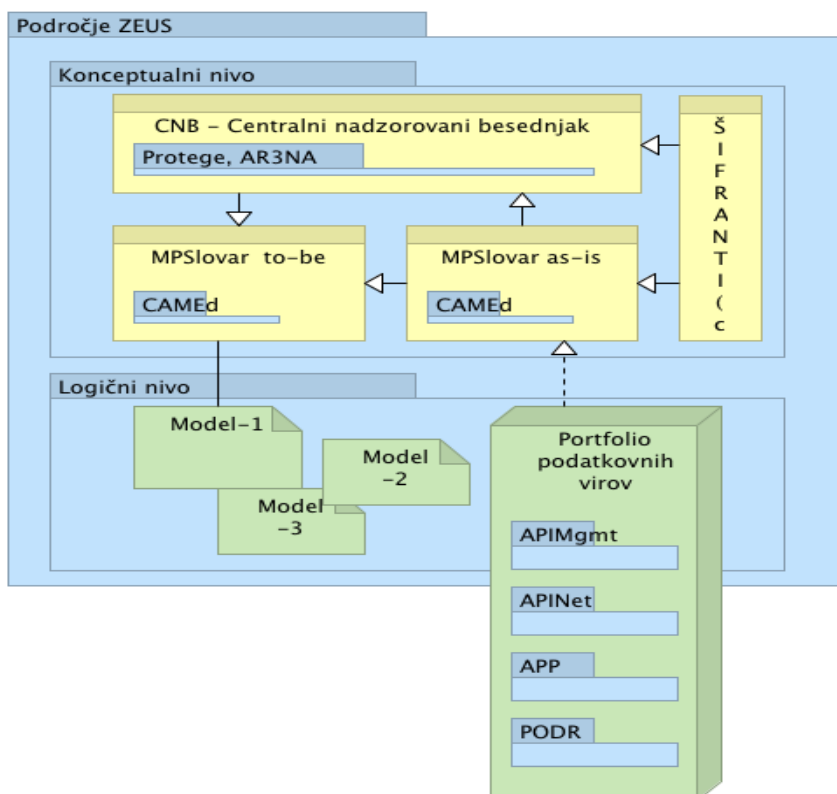
Poglavje obravnava vprašanja kako vzpostaviti in vzdrževati metapodatkovne zbirke in slovarje skozi čas. Kot vedno obstaja več možnih poti in pristopov, iščemo kombinacijo najuspešnejših.

ZEUS, zakonska ureditev področja

ZEUS je skupno delovno ime, pod katerim zbiramo potrebe po zakonskih spremembah, ne samo na področju upravljanja s podatki, temveč za širše področje digitalizacije.

Izhodišča za formalno ureditev področja se umeščajo pod kodno ime ZEUS. Zakonske podlage so prvi pristop k vzpostavitvi in upravljanju metapodatkov.

Na sliki je predstavljen obseg nivojev in komponent, ki naj bi jih pokrival zakon.



Zakon naj ureja:

1. Razvoj sektorsko specifičnih podatkovnih modelov mora, povsod kjer so ti na voljo, temeljiti na:
 - Enotnih pojmi in njihovih definicijah v centralnem nadzorovanem besednjaku javne uprave.
 - Enotnih jedrnih podatkovnih modelih (ang. core vocabulary application profile), v metapodatkovnem repozitoriju javne uprave.
 - Enotnih referenčnih podatkih (šifrantih) iz centralnega registra šifrantov.
2. Enotni jedrni podatkovni modeli morajo temeljiti na terminih, ki so definirani v centralnem nadzorovanem besednjaku.
3. Termini, ki so definirani v centralnem nadzorovanem besednjaku morajo biti zapisan v strojno berljivem formatu (npr. SKOS, RDF).
4. Termini, ki so definirani v centralnem nadzorovanem besednjaku so razdeljeni glede na specifična vsebinska področja.
5. Termine specifične domene definirajo in vzdržujejo organi, ki področje na katero se termini nanašajo, pokrivajo v skladu z zakonodajo.
6. Enotnih referenčne podatke (šifrante) definirajo in vzdržujejo organi, ki področje na katero se ti referenčni podatki nanašajo, pokrivajo v skladu z zakonodajo.
7. Enotne jedrne podatkovne modele definirajo in vzdržujejo organi, ki področje na katero se ti podatkovni modeli nanašajo, pokrivajo v skladu z zakonodajo.
8. Ministrstvo za javno upravo RS skrbi za:
 - a. administracijo ter medsebojno usklajevanje vsebinskih področij in pojmov v centralnem nadzorovanem besednjaku.
 - b. medsebojno usklajevanje enotnih jedrnih podatkovnih modelov.
 - c. medsebojno usklajevanje enotnih referenčnih podatkov (šifrantov).
 - d. Dostopnost in objavo meta podatkov
 - e. Organizacijo mreže vsebinskih skrbnikov po institucijah, ki so pristojne za posamezna vsebinska področja
 - f. uvajanje in izobraževanje vsebinskih skrbnikov meta podatkov

Zakon mora opredeliti tudi nekatera razmerja in obveznosti med posameznimi vlogami, v katerih nastopajo posamezne institucije ter ustrezne pristojnosti:

- a. Kdo je lastnik podatka (pravna ali fizična oseba, katere lastnost opisuje podatek)
- b. Kdo je skrbnik podatka (fizična oseba, ki podatke zbira in z njimi upravlja, da bi lahko izvajala dejavnost, za katero je po zakonu pristojna)
- c. Kdo je skrbnik systemske opreme, kjer so podatki hranjeni
- d. Kdo in kdaj je uporabnik podatkov in kdaj je uporabnik pooblaščen za vpogledovanje
- e. Kdo in kako podatke posreduje
- f. Kaj je računalniška izmenjava podatkov
- g. Kdo je zadolžen za vzpostavljanje podatkovnih storitev in to vzpostavitev financira (tisti, ki storitev potrebuje, ali tisti, ki je skrbnik aplikacije)
- h. Obveznost (ali ne) vodenje sprememb tako, da so možni vpogledi v stanje podatkov po preteklih časovnih presekih

Razširitev sektorskih in centralnih projektov s semantiko

Pristop, ki bi bil komplementaren zakonski ureditvi področja, bi temeljil na razširitvi sektorskih in centralnih projektov na način, da bi se del analize in razvoja posvetil semantiki in vzdrževanju sektorsko specifičnega metapodatkovnega področja, ki ga posamezni projekt obravnava.

Pristop z mehanizmi zajema metapodatkovnih opisov v okviru procesa gostovanja

Tretji prav tako dopolnilni pristop pa se nanaša na mehanizme in postopke v okviru procesov gostovanja na centralni informacijski infrastrukturi s katerimi bi uvedli avtomatizirano polnjenje centralnih metapodatkovnih zbirk na podlagi standardiziranih opisov, bodisi v dokumentaciji, funkcijah sistema (kot na primer objava WSDL in WADL) ali izvorni kodi izdelkov.

Izvajanje obstoječih predpisov na področju evidentiranja podatkovnih zbirk

Četrti ne nujno tudi pristop na zadnjem mestu je uporaba obstoječih mehanizmov in zakonskih podlag na področju obveznega evidentiranja zbirk osebnih podatkov v skladu z GDPR (direktivo: General Data Protection Regulation) ter evidentiranja zbirk odprtih podatkov na podlagi ZDIJZ. Določitev formalni upravljavcev podatkovnih zbirk je zavezujoča naloga najmanj po dveh smereh. GDPR zavezuje upravljavce in obdelovalce, da se zbirke osebnih podatkov zajamejo v evidencah, ZDIJZ pa zavezuje upravljavce zbirk, da se vse zbirke zajamejo v evidencah. Vendar oba predpisa ne predpisujeta vodenje centralnih zbirk temveč lokalne pri upravljavcih.

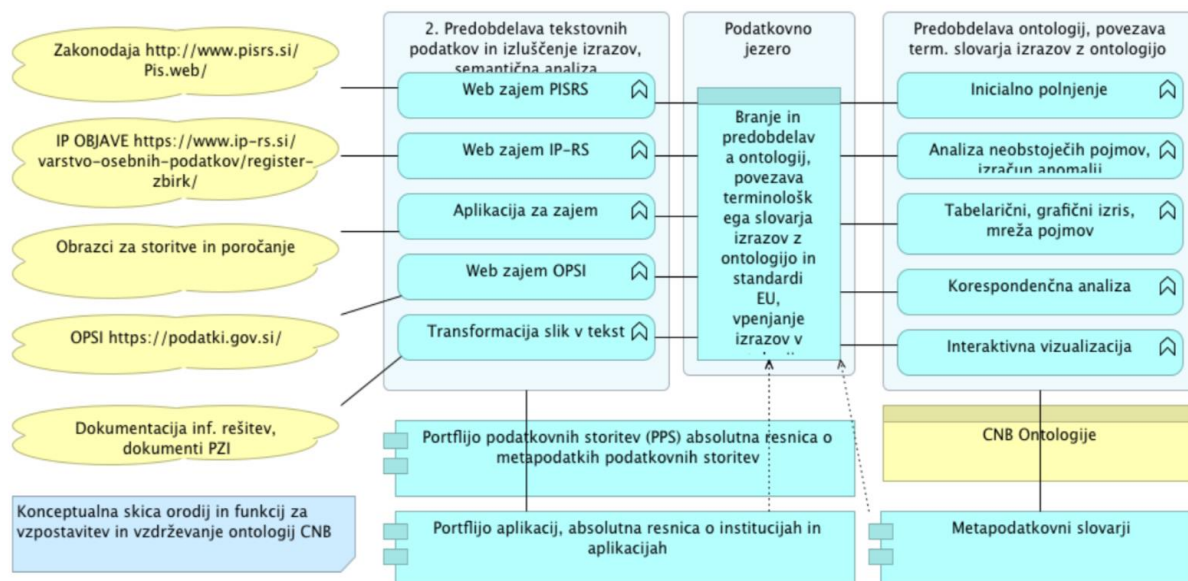
V tej situaciji se lahko v okviru upravljanja centralne infrastrukture vzpostavijo centralne zbirke, ki olajšajo, pomagajo zavezancem pri vzpostavitvi evidenc.

Kot enega primernih od centralnih sistemov se lahko izbere 42@EVA oziroma portfolijo aplikacij.

Razvoj orodja za pomoč pri upravljanju centralnega besednjaka CNB

Priprava in vzdrževanje centralnega besednjaka (CNB) je ena ključnih načrtovanih nalog, ki jih uvaja strategije upravljanja z meta podatki. Ker bo CNB pokrival širok nabor vsebinskih področij sistemov državne uprave, bomo zaupali opisovanje pojmov in njihovih definicij skrbnikom za posamezna vsebinska področja. Pričakujemo, da bo količina opisanih pojmov kmalu prevelika, da bi lahko različni vnašalci imeli vpogled v celoto in tako ugotavljali, ali je bil nek pojem že opisan, ali je njegov opis primeren ali pa gre za sinonim. Ker take relacije med pojmi bistveno olajšajo iskanja, je pomembno da jih ne spregledamo in da jih ustrezno zabeležimo v besednjaku.

Da bi se vzdrževanje vsebine CNB podprlo z ustreznimi orodji za konsistenten zajem podatkov v besednjak predlagamo razvoj orodja s področja analitike besedil, vizualizacije in strojnega učenja. Identificiranih je vsaj 5 virov vhodnih besedilnih, ki jih je potrebno preoblikovati in na primeren način kontekstualno ponuditi pri procesu polnjenja, vzdrževanja in preverjanja konsistentnosti CNB.



Slika prikazuje konceptualno shemo procesa podpore polnjenja in vzdrževanja CNB.

Določanje upravljavcev in obdelovalcev podatkovnih zbirk

Za vsak sistem, se v fazi projekta oziroma namestitvenih postopkov določijo atributi za vsak postopek in pripadajoči podatkovni sklop v matriki razmejitve odgovornosti. Matrika se informatizira in vodi v okviru informacijskega sistema „42@EVA - Portfolio aplikacij“.

Javna objava metapodatkov in slovarjev

Po vzoru Finske (<http://finto.fi/en/>, <http://iow.csc.fi/>), se portal odprtih podatkov (podatki.gov.si) dopolni z orodji za objavo metapodatkov o zbirkah, ki so objavljene na portalu v obliki odprtih podatkovnih setov.

VIRI

Finski core modeli in aplikacijski profili:

<http://iow.csc.fi/group?id=https:%2F%2Ftt.eduuni.fi%2Fsites%2Fscsc-iow%23JHS>

Finnish Public Sector Core model:

<http://iow.csc.fi/model/jhs/>

Application profile for systematic terminologies

<http://iow.csc.fi/model/st/>

Vocabulary and concept workbench (<https://sanasto.csc.fi/>)

Implementation of the Finnish Government Programme (<http://valtioneuvosto.fi/en/implementation-of-the-government-programme>)

Say, “S” (as) Semantics – and Mean it! Path to Semantically Interoperable Digital Research Services

(<http://www.sciencedirect.com/science/article/pii/S1877050917303009?via%3Dihub>)

EU. 2017. "EuroVoc, Večjezični Teazver Evropske Unije."
<http://eurovoc.europa.eu/drupal/?q=node/1547>.

Heath, Tom, and Christian Bizer. 2011. "Linked Data: Evolving the Web into a Global Data Space." *Synthesis Lectures on the Semantic Web: Theory and Technology* 1 (1). Morgan & Claypool Publishers: 1–136.

Interoperability Solutions for European Public Administration (ISA). 2013. "How Linked Data Is Transforming eGovernment." https://joinup.ec.europa.eu/sites/default/files/inline-files/D4.3.2_Case_Study_Linked_Data_eGov.pdf.
2016. "E-Government Core Vocabularies Handbook." https://ec.europa.eu/isa2/library/e-government-core-vocabularies-handbook_en.

PRILOGA A

Miselni vzorec pojmovnika

