Výstup č. 1

Centrálny dátový model verejnej správy a plán jeho realizácie

Zmluva o dielo č. 302/2018

*Projekt:*

Zlepšenie využívania údajov vo verejnej správe

*ITMS kód projektu:*

314011S979

Obsah

[1 Účel dokumentu 9](#_Toc9682350)

[1.1 Použité skratky 9](#_Toc9682351)

[2 Manažérske zhrnutie 11](#_Toc9682352)

[2.1 Sémantický web – Web 3.0 11](#_Toc9682353)

[2.2 Centrálny model údajov 11](#_Toc9682354)

[2.3 Oblasti použitia 13](#_Toc9682355)

[2.4 Potrebné kroky 13](#_Toc9682356)

[2.5 Záver 14](#_Toc9682357)

[3 Súčasný stav 14](#_Toc9682358)

[3.1 Strategické dokumenty 14](#_Toc9682359)

[3.2 Prieskumy a požiadavky 14](#_Toc9682360)

[3.3 Štandardizácia 15](#_Toc9682361)

[3.4 Centrálny metainformačný systém 15](#_Toc9682362)

[3.5 Metodika dátového modelovania 16](#_Toc9682363)

[4 Predstavenie centrálneho modelu údajov 16](#_Toc9682364)

[4.1 Koncept (FAQ) 16](#_Toc9682365)

[4.1.1 Sémantické dátové štandardy 16](#_Toc9682366)

[4.1.2 Triplet - Trojica 16](#_Toc9682367)

[4.1.3 Ontológie 17](#_Toc9682368)

[4.1.4 Linkované údaje 18](#_Toc9682369)

[4.1.5 SPARQL 18](#_Toc9682370)

[4.1.6 Čo to je Centrálny model údajov verejnej správy 18](#_Toc9682371)

[4.1.7 Referencovateľný identifikátor 19](#_Toc9682372)

[4.1.8 Stotožňovanie entít 20](#_Toc9682373)

[4.1.9 SHACL 21](#_Toc9682374)

[4.1.10 Formáty serializácie sémantických dátových štandardov 21](#_Toc9682375)

[4.2 Otvorené údaje: 4 vs 5 hviezdičkové data 23](#_Toc9682376)

[4.2.1 Definícia 23](#_Toc9682377)

[4.2.2 Rozdiel medzi 4 a 5 hviezdičkovými dátami : 23](#_Toc9682378)

[4.3 Štandardizované prvky 24](#_Toc9682379)

[4.3.1 Ontológia fyzickej osoby 24](#_Toc9682380)

[4.3.2 Ontológia právneho subjektu 25](#_Toc9682381)

[4.3.3 Ontológia finančných entít 26](#_Toc9682382)

[4.3.4 Ontológia lokácie 27](#_Toc9682383)

[4.3.5 Ontológia egovernment entít 28](#_Toc9682384)

[4.4 Medzinárodné ontológie 29](#_Toc9682385)

[4.5 Katalóg dátových prvkov 29](#_Toc9682386)

[4.6 Metaúdaje katalógu, datasetu, distribúcie otvorených údajov 30](#_Toc9682387)

[5 Aplikácia Centrálneho modelu údajov v eGov 31](#_Toc9682388)

[5.1 Kompetencie 31](#_Toc9682389)

[5.2 Číselníky 32](#_Toc9682390)

[5.3 CSRÚ a otvorené údaje 34](#_Toc9682391)

[5.4 Dereferenciácia 35](#_Toc9682392)

[5.4.1 Predstavenie 35](#_Toc9682393)

[5.4.2 Koncept realizácie 36](#_Toc9682394)

[5.4.3 Rozšírené parametre dereferenciácie 38](#_Toc9682395)

[5.4.4 Dereferenciácia súborov 39](#_Toc9682396)

[5.5 Elektronické formuláre 40](#_Toc9682397)

[5.5.1 Referencovatený identifikátor XML menného priestoru a elektronického formulára 40](#_Toc9682398)

[5.5.2 Referencovateľný identifikátor pre súčasti elektronického formulára 41](#_Toc9682399)

[5.5.3 Metaúdaje elektronického formulára 44](#_Toc9682400)

[5.6 Referenčné údaje 44](#_Toc9682401)

[6 Procesy 46](#_Toc9682402)

[6.1 Priradenie jednotného referencovateľného identifikátora 46](#_Toc9682403)

[6.1.1 Popis 46](#_Toc9682404)

[6.1.2 Príklad 47](#_Toc9682405)

[7 Prílohy 48](#_Toc9682406)

[7.1 Príloha 1 – Výstupy aktivity 1.2 a 1.3 – zoznam dátových entít a číselníkov 48](#_Toc9682407)

[7.2 Príloha 2 – Modely dátových entít 49](#_Toc9682408)

[7.2.1 Finančná správa 50](#_Toc9682410)

[7.2.2 Register úpadcov 57](#_Toc9682412)

[7.2.3 Protimonopolný úrad 58](#_Toc9682414)

[7.2.4 Register adries 64](#_Toc9682416)

[7.2.5 Register fyzických osôb 67](#_Toc9682418)

[7.2.6 Register právnických osôb 70](#_Toc9682420)

[7.2.7 Sociálna poisťovňa 73](#_Toc9682422)

[7.2.8 UPSVaR 74](#_Toc9682424)

[7.2.9 Zdravotná poisťovňa 77](#_Toc9682426)

[7.3 Príloha 3 - Príklad serializácie dátového prvku 81](#_Toc9682428)

[7.3.1 JSON-LD 81](#_Toc9682429)

[7.3.2 RDF/XML 82](#_Toc9682430)

[7.3.3 N-Triples 82](#_Toc9682431)

[7.4 Príloha 4 – Metaúdaje otvorených údajov 83](#_Toc9682432)

[7.4.1 Katalóg 83](#_Toc9682433)

[7.4.2 Dataset 84](#_Toc9682434)

[7.4.3 Distribúcia 87](#_Toc9682435)

**Zoznam obrázkov**

[Obrázok 1: Príklad stotožnenia prostredníctvom sémantických väzieb 21](#_Toc9682436)

[Obrázok 2: Ukážka výsledku stotožnenia sémantickými vzťahmi 21](#_Toc9682437)

[Obrázok 3: Dátová schéma fyzickej osoby 25](#_Toc9682438)

[Obrázok 4: Dátová schéma právneho subjektu 26](#_Toc9682439)

[Obrázok 5: Dátová schéma finančných entít 27](#_Toc9682440)

[Obrázok 6: Dátová schéma lokácie 28](#_Toc9682441)

[Obrázok 7: Dátová schéma eGovernment entít 29](#_Toc9682442)

[Obrázok 8: Dátová schéma metaúdajov otvorených údajov 30](#_Toc9682443)

[Obrázok 9: Koncepcia využitia Centrálneho modelu údajov 31](#_Toc9682444)

[Obrázok 10: Komunikačná schéma kooperácie platformy otvorených údajov a CSRÚ 35](#_Toc9682445)

[Obrázok 11: Základná komunikačná schéma dereferenciácie 37](#_Toc9682446)

[Obrázok 12: Dátová schéma záznamu o referečnom údaji 45](#_Toc9682447)

[Obrázok 13: OE\_FSSR\_DanovePriznaniaFOB\_Asynch\_v003 50](#_Toc9682448)

[Obrázok 14: OE\_FSSR\_DanovePriznaniaFOB\_Synch\_v003 51](#_Toc9682449)

[Obrázok 15: OE\_FSSR\_DanovePriznaniaPO\_Asynch\_v003 51](#_Toc9682450)

[Obrázok 16: OE\_FSSR\_DanovePriznaniaPO\_Synch\_v003 52](#_Toc9682451)

[Obrázok 17: OE\_FSSR\_DanoveSubjekty\_Asynch\_v002 53](#_Toc9682452)

[Obrázok 18: OE\_FSSR\_DanoveSubjekty\_Synch\_v002 54](#_Toc9682453)

[Obrázok 19: OE\_FSSR\_DanoveSubjektyRegisterDPH\_Asynch\_v003 55](#_Toc9682454)

[Obrázok 20: OE\_FSSR\_DanoveSubjektyRegisterDPH\_Synch\_v003 55](#_Toc9682455)

[Obrázok 21: OE\_FSSR\_Nedoplatky\_Asynch\_v007 56](#_Toc9682456)

[Obrázok 22: OE\_FSSR\_Nedoplatky\_Synch\_v007 56](#_Toc9682457)

[Obrázok 23: OE\_MinSprav\_RegisterUpadcov\_Sync\_v001 57](#_Toc9682458)

[Obrázok 24: OE\_SEMP\_REGISTER\_POSKYTOVATELIA\_VYKONAVATELIA\_v001 58](#_Toc9682459)

[Obrázok 25: OE\_SEMP\_REGISTER\_PRIJEMCOV\_POMOCI\_ASYNC\_v001 59](#_Toc9682460)

[Obrázok 26: OE\_SEMP\_REGISTER\_PRIJEMCOV\_POMOCI\_SYNC\_v001 60](#_Toc9682461)

[Obrázok 27: OE\_SEMP\_REGISTER\_SCHEMY\_MP\_v001 62](#_Toc9682462)

[Obrázok 28: OE\_SEMP\_REGISTER\_SCHEMY\_SP\_v001 63](#_Toc9682463)

[Obrázok 29: RA\_AddressWithAddressPointByBuildingIndex\_v20 64](#_Toc9682464)

[Obrázok 30: RA\_AddressesWithAddressPointsByAttributes\_v20 65](#_Toc9682465)

[Obrázok 31: RA\_AddressesWithAddressPointsByBuildingIndexes\_v20 66](#_Toc9682466)

[Obrázok 32: PoskytnutieUdajovIFOOnlineWS-v1.0 67](#_Toc9682467)

[Obrázok 33: PoskytnutieUdajovIFOPodlaSpecifikacieWS-v1.0 68](#_Toc9682468)

[Obrázok 34: PoskytnutieZoznamuIFOPodlaVyhladavacichKriteriiWS-v1.0 69](#_Toc9682469)

[Obrázok 35: Poskytnutie\_IPO\_out-v2.4 70](#_Toc9682470)

[Obrázok 36: Poskytnutie\_suborov\_zmien\_PO\_out-v2.4 71](#_Toc9682471)

[Obrázok 37: Poskytnutie\_udajov\_PO\_out-v2.4 72](#_Toc9682472)

[Obrázok 38: OE\_SocPoist\_OdvodovePovinnosti\_v002 73](#_Toc9682473)

[Obrázok 39: OE\_UPSVaR\_Ciselniky\_v001 74](#_Toc9682474)

[Obrázok 40: OE\_UPSVaR\_EUoZ\_v013 75](#_Toc9682475)

[Obrázok 41: OE\_UPSVaR\_KartaUcastnika\_v012 76](#_Toc9682476)

[Obrázok 42: OE\_ZP\_Odvodova\_Povinnost\_v005 77](#_Toc9682477)

[Obrázok 43: OE\_ZP\_UdajeSZCO\_v002 78](#_Toc9682478)

[Obrázok 44: OE\_ZP\_ZamestnanciZamestnavatela\_v001 79](#_Toc9682479)

[Obrázok 45: OE\_ZP\_ZamestnaniaZamestnanca\_v001 80](#_Toc9682480)

# Účel dokumentu

Predkladaný dokument je súčasťou Výstupu 1 Zmluvy o dielo č. 302/2018 predmetom ktorej je “Centrálny dátový model verejnej správy a plán jeho realizácie“ (https://www.crz.gov.sk/index.php?ID=3816799&l=sk). Zmluva bola uzavretá s cieľom naplniť aktivity a výstupy definované v národnom projekte „Zlepšenie využívania údajov vo verejnej správe“ realizovanom z prostriedkov Európskeho sociálneho fondu prostredníctvom Operačného programu Efektívna verejná správa prioritná os 1 – Posilnené inštitucionálne kapacity a efektívna verejná správa v rámci Špecifického cieľa 1.1. Skvalitnené systémy a optimalizované procesy VS s realizáciou pre celé územie SR.

Dokument je primárne zameraný na predstavenie Centrálneho dátového modelu s cieľom prehľadným spôsobom oboznámiť s aktuálnym stavom v tejto oblasti ako aj definovať základné procesy týkajúce sa rozvoja Centrálneho dátového modelu. Dokument do značnej miery vychádza zo schválených materiálov zverejnených na wiki stránke[[1]](#footnote-1), ktoré boli vypracované v priebehu štandardizácie Centrálneho dátového modelu ako aj sémantických dátových štandardov ako celku.

Súčasťou dokumentu sú dátové modely všetkých dátových entít, ktoré je dnes možné referencovať cez centrálnu integračnú platformu (CSRÚ).

Kľúčovou prílohou tohto dokumentu je zoznam dátových objektov, ktorý sumarizuje súčasné a predovšetkým identifikuje potencionálne referenčné údaje a základné číselníky, ktoré sú prepoužívané v rôznych oblastiach štátnej správy. Tento zoznam predstavuje roadmapu z pohľadu dátových objektov, ktorých referencovanie a zdieľanie sú a/alebo budú centrálne adresované a je východiskom pre plánovanie ďalších implementačných aktivít. Tento zoznam predstavuje živý dokument, ktorý bude, za základe identifikovaných požiadaviek, pravidelne aktualizovaný Centrálnou dátovou kanceláriou verejnej správy.

## Použité skratky

| Skratka / Pojem | Vysvetlenie |
| --- | --- |
| **ArchiMate** | Jazyk na modelovanie enterprise architektúry spravovaný zoskupením The Open Group: http://www.opengroup.org/subjectareas/enterprise/archimate-overview. |
| **CSRÚ** | Centrálny systém referenčných údajov (dáta verejnej správy). |
| **Metadáta / Metaúdaje** | Na základe Výnosu Ministerstva financií Slovenskej republiky o štandardoch pre informačné systémy verejnej správy metaúdajmi sú štruktúrované údaje obsahujúce informácie o primárnych údajoch, pričom primárne údaje spravidla reprezentujú určitý hmotný objekt alebo nehmotný objekt. Metaúdaje sú určené najmä na vyhľadávanie, katalogizáciu a využívanie primárnych údajov. |
| **NKIVS** | Národná koncepcia informatizácie verejnej správy. |
| **Objekt evidencie** | Objektom evidencie je množina údajov o subjekte evidencie, ktorá je predmetom evidovania orgánom verejnej moci v rámci jeho pôsobnosti podľa osobitných predpisov a ktorá je jednoznačne identifikovaná identifikátorom objektu evidencie (akýkoľvek identifikátor tj. číselný ako aj referencovateľný) § 49 písm. b) zákona č. 305/2013 Z. z. Zákon o elektronickej podobe výkonu pôsobnosti orgánov verejnej moci a o zmene a doplnení niektorých zákonov (zákon o e-Governmente) (ISO 15489-1). |
| **Ontológia** | Ontológie sú formálnym spôsobom, ako popísať taxonomické a klasifikačné siete, ktoré v podstate definujú štruktúru vedomostí pre rôzne oblasti: podstatné mená predstavujúce triedy objektov a slovesá reprezentujúce vzťah medzi objektmi. Ontológiu si je možné predstaviť aj ako databázovú schému pre grafové databázy(sémantické databázy). |
| **OPII** | Operačný program Integrovaná Infraštruktúra. |
| **OpenData** | Otvorené údaje. |
| **OVM** | Orgán verejnej moci. |
| **RA** | Register adries. |
| **RDF** | RDF (Resource Description Framework) predstavuje štandard výmeny dát na webe. |
| **PS1** | Pracovná skupina na úrade podpredsedu pre investície a informatizáciu, ktorá sa zaoberá národnými dátovými štandardmi |
| **RÚ** | Referenčné údaje. Referenčným údajom je údaj objektu evidencie, ktorý je uvedený v zozname referenčných registrov schvaľovaných ÚPVII.  V zmysle zákona č. 305/2013 Z. z. Zákon o elektronickej podobe výkonu pôsobnosti orgánov verejnej moci a o zmene a doplnení niektorých zákonov (zákon o e-Governmente): „Zoznam referenčných registrov obsahuje ako referenčné údaje také údaje objektu evidencie, ktoré sú vo vzťahu k subjektu evidencie jedinečné, a taký register ako referenčný, v ktorom je podľa osobitných predpisov k týmto údajom ustanovená domnienka správnosti. V zozname referenčných registrov musí byť každý referenčný údaj priradený k referenčnému registru a každý referenčný register musí byť priradený aspoň k jednému referenčnému údaju.“ |
| **RFO** | Register Fyzických Osôb – referenčný register. |
| **RPO** | Register Právnických Osôb – referenčný register. |
| **RR** | Referenčný register verejnej správy (napr. RPO, RFO a pod.) |
| **ÚPVII** | Úrad podpredsedu vlády SR pre investície a informatizáciu. |
| **URI** | Jednotný identifikátor prostriedku je kompaktný reťazec znakov používaný na identifikáciu alebo pomenovanie zdroja. |
| **IRI** | Rozšírenie špecifikácie URI – povolenie použitia Unicode znakov (pri URI len ASCII znaky) |
| **CMU** | Centrálny model údajov (niekedy tiež označovaný ako Centrálny dátový model - synonymum) |
| **ISA2** | Organizácia pod záštitou Európskej komisie, ktorá sa zaoberá návrhmi a tvorbou odporúčaní a interoperabilných štandardov na úrovni EÚ. |
| **JRI** | Jednotný referencovateľný identifikátor |
| **Úseky a agendy verejnej správy** | Zoznam agend verejnej správy je ustanovený výnosom Ministerstva financií SR č. 478/2010 Z. z. o základnom číselníku úsekov verejnej správy a agend verejnej správy. |
| **WS-SOAP** | Webová služba používajúca SOAP protokol. |
| **Dátová entita** | Objekt , z reálneho sveta, určitého typu (napríklad fyzická osoba, právnická osoba) |
| **Dátový prvok** | Všeobecný názov pre dátovú entitu alebo dátovú vlastnosť |
| **Sémantická dátová schéma** | Viď ontológia |
| **XML** | XML znamená eXtensible Markup Language, v preklade rozšíriteľný značkovací jazyk, ktorý bol vyvinutý a štandardizovaný konzorciom W3C (World Wide Web Consortium) ako pokračovanie jazyka SGML a zovšeobecnenie jazyka HTML. Umožňuje jednoduché vytváranie konkrétnych značkovacích jazykov na rôzne účely a široké spektrum rôznych typov údajov. |

# Manažérske zhrnutie

[D](https://wiki.finance.gov.sk/pages/viewpage.action?pageId=23691376)áta patria medzi najdôležitejšie aktívum štátu. Na základe nich vedia štátne inštitúcie správne analyzovať situáciu a rozhodnúť tak, aby to bolo čo najoptimálnejšie vzhľadom na občana. Problematika dát je značne rozsiahla a je potrebné sa na ňu pozerať z rôznych pohľadov. Už v histórii sa hľadali mnohé riešenia, ktoré mali prostredníctvom technických prostriedkov vyriešiť problémy náročných integrácií. Výsledkom mnohoročného snaženia v rámci štandardizačných komisií je vytvorenie Centrálneho modelu údajov verejnej správy, ktorý prostredníctvom technologicky agnostického prístupu rieši obrovskú časť problematiky dát, dátových integrácií ako aj dátovej konzistencie či už na úrovni syntaktickej ako aj sémantickej.

## Sémantický web – Web 3.0

Základným kameňom Centrálneho modelu údajov je využitie konceptu Web 3.0, ktorý pred viac ako dekádou definoval Tim Berners Lee, zakladateľ internetu. Jeho základom je, že každý vzťah a entita je jednoznačne identifikovaná v definovanom priestore (v našom prípade hovoríme o celom priestore eGov). Na tento účel sa používajú URI/IRI identifikátory, ktoré sú jednotlivým objektom a väzbám priraďované. Prostredníctvom týchto identifikátorov vieme akýkoľvek dátový prvok jednoznačne identifikovať v rámci celej SR. Samotným identifikátorom je potrebné priradiť význam tzv. sémantickú dátovú schému. Túto schému pomenovávame Centrálny model údajov verejnej správy.

Zavedením jednotného popisu dát (centrálnej dátovej schémy) vzniká univerzálny slovník medzi strojmi. Vďaka CMÚ dokážu stroje pochopiť významu dát, ktoré spracovávajú. Dôležitým elementom pri sémantickom popise údajov je aplikovanie logických vzťahov medzi entitami ako dedenie tried a vzťahov, zjednotenia, prieniky, definovanie totožnosti a rozdielnosti medzi entitami a iné.

Popisovanie dát prostredníctvom sémantických väzieb bolo pretavené do vytvorenie sémantických dátových štandardov, ktorých súčasťou je aj Centrálny model údajov. I keď sa môže zdať, že Centrálny dátový model je viazaný len na prostredie slovenského eGovu, nie je to tak. Model do značnej miery vychádza z odporúčaní ako aj nadnárodných EÚ modelov, ktoré sú štandardizované v rámci ISA2.

## Centrálny model údajov

Centrálny model údajov je možné chápať ako dátový model kľúčových dátových entít a atribútov v eGov priestore (dátové entity s vysokým integračným potenciálom pre tretie strany). Vzniká tým priestor na unifikáciu slovníka a spôsobu popisu jednotlivých údajov. Jednotlivé časti boli preto už štandardizované a zároveň boli vykonané aj novely výnosu o štandardoch 55/2014. Jeho rozšírenie bolo posilnené začlenením Centrálneho modelu údajov do Strategických priorít OPII ako sú Manažment údajov a Otvorené dáta.

Dôležitým aspektom Centrálneho modelu údajov je, že je úzko napojený na nadnárodné štandardy. Na európskej úrovni prebieha prostredníctvom programu ISA2 štandardizácia nadnárodných dátových modelov. Centrálny model údajov tieto medzinárodné štandardy v plnej miere integruje a prepoužíva. Tým je zabezpečená dátová kompatibilita v rámci celej EÚ. Nakoľko má každá krajina svoje špecifiká, Centrálny model údajov rozširuje EÚ štandardy o potreby Slovenského eGov priestoru.

Využitie URI identifikátorov (jednotný referencovateľný identifikátor) výrazným spôsobom prispieva k naplneniu princípu jeden krát a dosť na dátovej úrovni. Keďže všetky objekty a väzby majú globálny identifikátor unikátny v rámci celej SR, je možné vždy jasne ukázať o akom objekte sa hovorí čím zavádza do IT priestoru toľko potrebnú jednoznačnosť. Ako príklad je možné uviesť reťazec 50349287, ktorý nenesie v sebe žiadnu sémantickú hodnotu a nie je jasné čo reprezentuje. Pri použití jednotného referencovateľného identifikátora https://data.gov.sk/id/legal-subject/50349287 je možné automaticky rozpoznať, že ide o entitu právneho subjektu.

Rovnakým spôsobom pristupujeme aj k popisu samotnej schémy tj. ontológie resp. Centrálnemu dátovému modelu. Keď majú elementy iba nejednoznačné názvy, je veľmi náročné na technickej úrovni zabezpečiť ich jednoznačnosť. Príkladom je prvok “meno”. Keď sú elementy pomenovávané na všeobecnej úrovni, stroj im nedokáže rozumieť. Zároveň nie je možné daný dátový prvok ani referencovať nakoľko ide o lokálny názov. Centrálny model údajov popisuje jednotlivé dátové elementy globálnymi identifikátormi, napr. https://data.gov.sk/def/ontology/legal-subject/name, ktorý jednoznačne definuje, že „meno“ vyjadruje názov firmy.

Oblastí využitia takto popísaných dát je nespočetne mnoho. Jedným z priamočiarych realizácií je tvorba doložky vplyvov v prípade zmien v legislatíve. Každá tabuľka a jej atribúty môžu byť popísané na základe dátových entít CMU tj. každý stĺpec každej tabuľky môže referencovať dátovú entitu CMU. Rovnaký prístup aplikujeme aj na integračné služby, kde vstupno-výstupné rozhrania služieb môžu mať svoju dátovú časť mapovanú na dátové prvky CMU (Pozn.: táto funkcionalita je plánovaná vrámci rozvoja Metainformačného systému, ktorý poskytne rozhranie na mapovanie dátových prvkov CMU na rozhrania služieb.). Tým, že budú rovnakým spôsobom popísané dátové schémy (služieb, formulárov, databázových tabuliek) a budú uchované v Metainformačnom systéme, bude možné robiť jednoduché dopyty nad týmito dátami – tj. kde sa aký dátový prvok nachádza, koľko krát a ako je používaný i mnohé iné. Potom v prípade zavádzania napríklad bezvýznamového identifikátora fyzickej osoby budeme vedieť, aké všetky systémy a oblasti budú tohto zmenou dotknuté a koľko bude táto zmena v konečnom dôsledku stáť. Tým sa zmení spôsob písania doložky vplyvov na rozpočet, ktoré budú podložené tvrdými dátami.

V rámci SR spravuje Centrálny model údajov UPVII. Túto agendu má na starosti Dátová kancelária, ktorá poskytuje odbornú, konzultačnú ako aj metodickú podporu pri zavádzaní Centrálneho modelu údajov do praxe. Pri správe do veľkej miery spolupracuje s pracovnou skupinou Dátové štandardy PS1, ktorá je takisto pod hlavičkou UPVII. Dátová kancelária na základe svojich zistení predpripravuje materiály do pracovnej skupiny PS1 v oblasti zmien a rozširovania CMÚ, ktoré sú na zasadnutí PS1 prerokovávané a schvaľované.

## Oblasti použitia

Aktuálne sú sémantické dátové štandardy schválené a ďalej sa pracuje na ich rozvoji. Svojim univerzálnym a technologicky agnostickým prístupom bolo možné ich zapracovať do mnohých nosných častí eGov architektúry.

**Živé identifikátory -** prepojenie IT sveta so svetom bežných občanov prostredníctvom dereferenciácie. Občania ako aj stroje sa na základe technických prostriedkov budú vždy vedieť jednoznačným spôsobom odkazovať

1. na entity bežného života ako identifikátor firmy, ktorý je zároveň odkaz na webovú stránku (napr. orsr.sk)
2. na entity, ktoré sú používané pre OpenAPI tj. dopytovanie sa na identifikátor entity vráti jeho štruktúrovanú informáciu vo formáte JSON alebo XML
3. dátové prvky sú jednoznačne referencovateľné a tým pádom je možné zaviesť globálnu syntaktickú a sémantickú validáciu na aplikačné služby
4. úložiska súborov

**Elektronické formuláre –** zjednotenie dátovej politiky pre použitie CMÚ a sémantických dátových štandardov v elektronických formulároch.

**Stotožňovanie entít, modelov a vzťahov –** na základe využitia sémantických vzťahov je možné riešiť problémy postupného nasadzovania referenčných identifikátorov ako aj postupný prechod na jednotné referencovateľné identifikátori. Zároveň je riešený problém, kedy rôzne ISVS môžu tú istú entitu reprezentovať odlišným identifikátorom a je potrebné tieto identifikátori zladiť a zjednotiť.

**Číselníky –** vytvorenie podmienok pre masívne zjednotenie využitia číselníkov a to nie len základných ale akýchkoľvek. Prechod na referencovateľné identifikátori zjednoduší implementačným tímom orientáciu sa v číselníkoch a zavedie potrebnú metodiku do ich použitia.

**Otvorené údaje –** zvýšenie využiteľnosti otvorených údajov prechodom na 4-5 hviezdičiek. Tým bude jednoduché robiť dátové mashupy, čo podporí tvorbu inovatívnych aplikácií.

**Elektronické služby štátu –** mapovanie služieb a dát, ktoré tečú v službách umožní analyzovať dopady zmien v dátových prvkoch (napr. legislatívne zmeny), ako aj zjednoznační dokumentáciu či umožní globálnu validáciu dát v službách.

## Potrebné kroky

Pre systematický nábeh Centrálneho modelu údajov do praxe, je potrebné aby boli vykonané aj nasledovné kroky:

* MetaIS vyhlásený ako referenčný register jednotných referencovateľných identifikátorov
* Dopracovanie potrebnej funkcionality do MetaIS
* Na základe metodiky „Metodika dátového modelovania“ [[2]](#footnote-2)na Ministerstve vnútra vypracovať metodické usmernenie pre jej aplikáciu v praxi
* Presadenie konceptu dereferenciácie v prostredí eGov ISVS
* Vyžadovať komunikáciu na rozhraniach pomocou štandardizovaných dátových prvkov definovaných v MetaIS
* Vyžadovať popisovanie rozhraní, databázových schém, zákonov prostredníctvom dátových prvkov CMÚ
* Každý referenčný register musí povinne poskytovať jednotné referencovateľné identifikátory pre entity ktoré spravuje ako aj dátový model definovaný v MetaIS.
* Vynucovanie referencovania základných číselníkov, ktoré majú povahu ako referenčné údaje

## Záver

Centrálny model údajov je jedným z hlavných stavebných kameňov eGov infraštruktúry. Vzhľadom na jeho presah na rôzne odvetvia eGovu ako aj jeho prepojenosť s nadchádzajúcimi projektmi, je potrebné tento koncept čo najviac rozšíriť smerom na všetkých správcov ISVS ako aj ich dodávateľov. CMU je potrebné chápať nie ako byrokratickú záťaž, ale ako spôsob, ktorý má po úspešnom nasadení uľahčiť manažment dát pre jednotlivé OVM. Tým bude do budúcna zredukovaná náročnosť na správu dát pričom sa pri jednotení dátového jazyku predpokladá zníženie nárokov ako na integráciu tak aj na prevádzku ISVS.

UPVII ako koordinačný úrad pre informatizáciu plánuje podporovať implementáciu CMU do jednotlivých nových ako aj existujúcich systémov. Vznik tohto dokumentu preto slúži ako prvotná kuchárka pre zainteresované osoby. I keď sa môže zdať implementácia na prvý pohľad komplikovaná, boli vykonané kroky aby boli vstupné náklady čo najnižšie. Jedným z dôležitých elementov je vznik Dátovej kancelárie, ktorá bude jednotlivým OVM pomáhať pri implementácii konceptu CMU do praxe. Na základe vstupov od jednotlivých správcov ISVS bude táto kuchárka neustále aktualizovaná o detailnejšie opisy resp. nové oblasti použitia.

V rámci predstavenia CMU boli vytvorené aj prehľadové tabuľky, kde pre jednotlivé OVM boli vypracované odporúčania a návrhy dátových prvkov, ktoré by mali byť štandardizované. Tieto zoznamy bude spravovať UPVII a plánuje ich neustále aktualizovať na základe vstupov zástupcov jednotlivých OVM.

# Súčasný stav

## Strategické dokumenty

* Národná koncepcia informatizácie verejnej správy (http://www.informatizacia.sk/narodna-koncepcia-informatizacie-verejnej-spravy/1306s)
* Lepšie dáta : Manažment údajov - Strategická priorita Manažment údajov
* Lepšie dáta : Otvorené údaje - Strategická priorita Otvorené údaje

## Prieskumy a požiadavky

Počas vypracovania tohto dokumentu boli uvažované relevantné dokumenty, ktoré sa problematikou dát zaoberajú.

**Prioritné životné situácie (Detailný akčný plán Informatizácie verejnej správy**

www.informatizacia.sk/ext\_dok-detailny\_akcny\_plan.../26030c

**Dátové požiadavky na službu Moje dáta**

https://metais.vicepremier.gov.sk/studia/detail/d2112fadb259-ae0c-157b-8587d43fa79a

**Hlavné dáta požadované verejnosťou**

https://docs.google.com/spreadsheets/d/13uVKtcS4vDHZISeLZvv-0E3Cefmutc8P-CVk0cEf1RM/edit#gid=0

**Rezortné analýzy potrebných údajov (požiadavky konzumentov)**

https://metais.vicepremier.gov.sk/studia/detail/6a3eda58-d61a-33e8-4034-33afda4eb51a

## Štandardizácia

Oblasť Centrálneho modelu údajov verejnej správy založeného na ontológiách prešla dlhodobým vývojom. Práce na tomto modely začali ešte v roku 2015 kedy sa vytvorili prvé dátové štruktúry. Tieto štruktúry ako aj celý obraz o sémantických dátových štandardoch bol postupne diskutovaný na rôznych pracovných skupinách v rámci UPVII. Samotný proces sa snažil byť v čo najväčšej možnej miere otvorený a preto boli jednotlivé návrhy zverejňované aj pre verejnosť. Táto metodika do značnej miery vychádza a používa definované časti v štandardizačnej skupine[[3]](#footnote-3) a novele výnosu[[4]](#footnote-4).

Výstup jednotlivých častí bol predkladaný do štandardizačnej skupiny PS1 – dátové štandardy na UPVII. Mnohé z výsledkov štandardizácie boli pretavené do novely výnosu o štandardoch 55/2014, kde boli jednotlivé časti aktualizované v dvoch novelách. Kľúčové časti, ktoré novela o štandardoch definuje v súvislosti s Centrálnym modelom údajov a sémantickými dátovými štandardmi sú:

* Použitie Centrálneho modelu údajov na popis dátových prvkov
* Nahradenie historického Katalógu dátových prvkov prvkami Centrálneho modelu údajov
* Definícia referencovateľných a jednotných referencovateľných identifikátorov
* Definovanie Metainformačného systému ako zodpovedného systému za dielčie časti v sémantických dátových štandardov (referenčný register jednotných referencovateľných identifikátorov, nástroj na popis dátových prvkov služieb prostredníctvom CMU, dereferenciácia, ...)
* Dereferenciácia
* Povolenie sémantických formátov na komunikáciu medzi systémami
* Podporenie vytvárania otvorených údajov s použitím sémantických dátových štandardov
* Štandardizácia a implementácia referencovateľných identifikátorov v prostredí Modulu Elektronických formulárov
* Návrhy generalizovanej podoby štruktúry číselníkov zakomponovaním sémantických dátových štandardov

## Centrálny metainformačný systém

V rámci OPIS bol vytvorený Metainformačný systém verejnej správy, ktorý plní prvotný technologický komponent, ktorý zadefinoval základné rámce pre prácu s referencovateľnými identifikátormi ako aj so sémantickými dátovými schémami. I keď je existujúca podpora zo strany MetaIS do značnej miery limitovaná oproti existujúcim požiadavkám, základnú prácu registrácie referencovateľných identifikátorov poskytuje už v súčasnosti (bez podpory modelovania, dereferenciácie a správy dát,...). Pre maximálne vyťaženie z Centrálneho modelu údajov a použitia referencovateľných identifikátorov je potrebná radikálna prestavba Metainformačného systému, ktorý by mal už na najnižšej úrovni podporovať sémantickú prácu s dátami.

## Metodika dátového modelovania

V priestore dátového modelovania dlhodobo chýbala centrálna metodika v oblasti modelovania dát a spracovania dát. Pri vypracovaní EVS projektu Optimalizácia procesov verejnej správy bola táto požiadavka adresovaná. Výsledkom je ucelená metodika dátového modelovania, ktorá slúži ako prvotný podklad pre naštartovanie procesov v rámci vznikajúcej Centrálnej dátovej kancelárie na UPPVII[[5]](#footnote-5).

Metodika nepokrýva úplne zavedenie Centrálneho modelu údajov a zameriava sa skôr na edukačnú a inicializačnú časť tvorby dátových modelov. Zároveň však kladie do kontextu komplexný prehľad dátového modelovania ako aj kompetentné inštitúcie za jednotlivé oblasti.

# Predstavenie centrálneho modelu údajov

## Koncept (FAQ)

### Sémantické dátové štandardy

Sémantické dátové štandardy reprezentujú celú množinu štandardov, ktorých výstupom sú :

* Tvorba referencovateľných a jednotných referencovateľných identifikátorov
* Dereferenciácia
* Ontológie - Centrálny model údajov
* Procesy spojené so správou centrálneho modelu údajov a prideľovania identifikátorov

### Triplet - Trojica

Základná informačná jednotka v sémantickom webe. Je reprezentovaná trojicou: subjekt, predikát, objekt. V zjednodušenej forme je to možné pochopiť prostredníctvom príkladu :

Mama má Emu

Nakoľko sa v priestore sémantického webu používajú URI identifikátory (referencovateľné identifikátory) technický zápis by mohol byť nasledovný :

https://vicepremier.sk/id/physical-person/MAMA https://vicepremier.sk/def/ontology/physical-person/isInRelationShipWith https://vicepremier.sk/id/physical-person/EMA

https://vicepremier.sk/id/physical-person/MAMA = Mama

https://vicepremier.sk/def/ontology/physical-person/isInRelationShipWith = má (je vo vzťahu)

https://vicepremier.sk/id/physical-person/EMA = Ema

### Ontológie

Ontológie reprezentujú dátovú schému linkovaných dát. Hlavným rozlíšením oproti bežným relačným tabuľkám je, že ontológia popisuje dáta v grafovej štruktúre a je preto univerzálnejšia ako tabuľkový prístup k dátami. Ďalším veľmi dôležitým prvkom ontológií je použitie URI identifikátorov na identifikáciu jednotlivých popisovaných entít a ich vzťahov. Vďaka tejto vlastnosti je možné robiť globálne dátové schémy v ktorých URI identifikátor zabezpečuje ich jednoznačnosť v celom priestore.

Základné dátové prvky ontológie sú Trieda, Objektová vlastnosť a Dátová vlastnosť.

**Trieda** - popisuje množinu objektov s rovnakými vlastnosťami. Príkladmi tried sú napríklad Fyzická osoba, Právnická osoba, Zákon, Liek. Triedu je možné chápať ako typ entity.

 <owl:Class rdf:about="&pper;PhysicalPerson">

        <rdfs:subClassOf rdf:resource="&dct;PeriodOfTime"/>

        <rdfs:label xml:lang="en">Physical Person</rdfs:label>

        <rdfs:label xml:lang="sk">Fyzická osoba</rdfs:label>

        <rdfs:label xml:lang="de">natürliche Person</rdfs:label>

        <adms:identifier rdf:resource="https://data.gov.sk/id/identifier/metais-standard/538" />

        <rdfs:subClassOf rdf:resource="http://www.w3.org/ns/person/Person" />

        <rdfs:isDefinedBy rdf:resource="https://data.gov.sk/def/ontology/physical-person/2018-03-29.owl"/>

    </owl:Class>

**Dátová vlastnosť** – vlastnosť viažuca sa na konkrétnu triedu alebo inštanciu, ktorá reprezentuje literálny typ tj. číslo, reťazec znakov, dátum a podobne. Príklady sú meno osoby, dátum narodenia, dátum vytvorenia spoločnosti, číslo bankového účtu a podobne.

<owl:DatatypeProperty rdf:about="&pper;familyName">

        <rdfs:label xml:lang="en">family name</rdfs:label>

        <rdfs:label xml:lang="sk">priezvisko</rdfs:label>

        <rdfs:label xml:lang="de">Nachname</rdfs:label>

        <rdfs:domain rdf:resource="&pper;PhysicalPerson"/>

        <rdfs:range rdf:resource="&xsd;string"/>

        <rdfs:subPropertyOf rdf:resource="&pper;name"/>

        <rdfs:description xml:lang="sk">Obsahuje priezvisko osoby.</rdfs:description>

        <adms:identifier rdf:resource="https://data.gov.sk/id/identifier/metais-standard/255" />

        <adms:identifier rdf:resource="https://data.gov.sk/id/identifier/kdp/D.1.1.1.6" />

        <rdfs:isDefinedBy rdf:resource="https://data.gov.sk/def/ontology/physical-person/2018-03-29.owl"/>

    </owl:DatatypeProperty>

**Objektová vlastnosť** – vlastnosť relácie medzi objektami. Napríklad vzťah medzi dvoma Osobami môže byť : súrodenec, manžel/ka, šef/zamestnanec, jeDieťa.

<owl:ObjectProperty rdf:about="&pper;father">

        <rdfs:subClassOf rdf:resource="&pper;relative" />

        <rdfs:label xml:lang="en">father</rdfs:label>

        <rdfs:label xml:lang="sk">otec</rdfs:label>

        <rdfs:label xml:lang="de">Vater</rdfs:label>

        <rdfs:domain rdf:resource="&pper;PhysicalPerson"/>

        <rdfs:range rdf:resource="&pper;PhysicalPerson"/>

        <rdfs:description xml:lang="sk">Relácia osoby na jej otca.</rdfs:description>

        <adms:identifier rdf:resource="https://data.gov.sk/id/identifier/metais-standard/422" />

        <rdfs:isDefinedBy rdf:resource="https://data.gov.sk/def/ontology/physical-person/2018-03-29.owl"/>

    </owl:ObjectProperty>

### Linkované údaje

Linkované údaje je spôsob, kedy sú štruktúrované dáta vyjadrené v grafovej štruktúre. Okrem dát samotných je aj ich schéma definované v grafovej štruktúre. Jednotlivé uzly v grafe ako aj hrany používajú URI identifikátor, ktorý zabezpečuje to, že uzly v grafe ako aj hrany môžu byť prostredníctvom URI identifikátora referencované.

### SPARQL

Dopytovací jazyk nad sémantickými úložiskami. Jazyk je oficiálnym W3C štandardom [[6]](#footnote-6)a jeho posledná finálna verzia je 1.1. S rozvojom sémantických štandardov aj v rámci Európskej únie vznikol tlak na ďalšie rozširovanie možností tohto dopytovacieho jazyka, čo viedlo k vytvoreniu novej skupiny pre verziu 1.2[[7]](#footnote-7).

### Čo to je Centrálny model údajov verejnej správy

Centrálny model údajov verejnej správy predstavuje sadu štandardizovaných dátových modelov (napr. dátový model fyzickej osoby, právnickej osoby, lokácie,...) a prvkov popísaných prostredníctvom ontológií, ktoré sú navzájom prepojené a ktoré využívajú na svoju identifikáciu jednotné referencovateľné identifikátory - URI. Tvorí ho množina ontológií popisujúcich kľúčové dáta (dáta s vysokým potenciálom na prepoužitie-integráciu) v štáte a ich mapovania na medzinárodné štandardy ako napr. ISA2 Core Public Vocabularies a ostatné.

Stratégiou tvorby centrálneho modelu je udržiavanie národných ontológií primárne pre základné číselníky, referenčné dáta, a na ostatné štandardné údaje ako popis datasetov, elektronických služieb, číselníkovej štruktúry sa používa odporučený štandard v danej oblasti.

Centrálny model údajov verejnej správy nemá ambíciu pokryť dátovo celú množinu dátových prvkov a objektov evidencie v inštitúciách verejnej správy. Jeho dôraz je na prvky, ktoré sa často používajú v integračných väzbách medzi informačnými systémami. Prílišný a nezmyselný rozmach dátového modelu by mal kontraproduktívne účinky najmä kvôli zhoršeniu nájditeľnosti jednotlivých dátových prvkov.

### Referencovateľný identifikátor

Reprezentuje globálny identifikátor v rámci priestoru ISVS. Je reprezentovaný URI identifikátorom, ktorý má v novele výnosu 55/2014 § 2 písmeno t) jednoznačne definovaný formát.

https://[domain]/[type]/[classification]+/[identity](/[version])?([pathToFile])?

[domain] - použitá doména. Entity centrálneho modelu údajov verejnej správy alebo jednotné referencovateľné identifikátory používajú doménu data.gov.sk

[type] =

* “**def**“ ak URI reprezentuje definičnú entitu v len sémantickom svete (ontológia, entita ontológie, pravidlá, abstraktný číselníkový typ)
* “**id**“ ak URI reprezentuje indivíduum
* “**set**“ ak URI reprezentuje dataset
* “**doc**“ ak URI reprezentuje jednoznačný identifikátor pre úložisko dokumentov.

[classification]+ = trieda alebo koncept, pričom klasifikácia môže mať viac úrovní. Účelom viacúrovňovej definície je umožnenie presnejšej definície a zabezpečenie jednoznačnosti referencovateľného identifikátora.

[identity] = číselný alebo textový identifikátor entity, ktorý priraďuje vlastník šablóny referencovateľného identifikátora.

[version] = Len indivíduá, ktoré môžu existovať súčasne v rôznych verziách môžu mať aj verziu. Vzhľad ako vyzerá verzia je nutné špecifikovať regulárnym výrazom, aby bolo možné riešiť dereferenciáciu pre verzionovateľné entity, kde chceme byť vždy naviazaní na poslednú verziu (štandardné správanie pri neuvedení verzie)

 [pathToFile] = tento parameter je používaný len v prípade dokumentov, ktoré majú parameter type = doc

**Príklad:** RPO (register právnických osôb) si v systéme MetaIS zaregistruje šablónu https://data.gov.sk/id/legal-subject/[identity] a tým sa stáva jej vlastníkom. V prípade RPO je preto parameter identity reprezentovaný IČO, čo vyjadruje jednoznačný referencovateľný identifikátor právnickej osoby.

**Jednotný referencovateľný identifikátor**

Jednotný referencovateľný identifikátor predstavuje koncept unikátneho globálneho identifikátora v rámci celého priestoru Slovenskej republiky. Tento identifikátor je zapísaný vo forme URI a vzťahujú sa naňho pravidlá definované vo výnose 55/2014 Z.z. Pre zabezpečenie unikátnosti týchto identifikátorov bola určená centrálna autorita (Metainformačný systém na UPVII), ktorá na požiadanie prideľuje menné priestory URI.

Jednotný referencovateľný identifikátor (JRI) má všetky vlastnosti referencovateľného identifikátora. Jeho hlavné odlišnosti oproti referencovateľným identifikátorom sú:

* nutné zabezpečenie unikátnosti v rámci celého priestoru SR,
* každá entita s priradením identifikátorom musí reprezentovať výhradne jednu entitu tj. identifikátor nemôže byť priradený ak nevieme zabezpečiť unikátnosť identifikátora vždy pre jedinečnú entitu,
* každý referenčný údaj musí mať priradený JRI, iné údaje a schémy len v prípade, že spĺňajú všetky náležitosti JRI definované štandardom,
* povinná registrácia v Centrálnom Metainformačnom systéme.

|  |  |  |
| --- | --- | --- |
|  | **Referencoveľný identifikátor** | **Jednotný referencovateľný identifikátor** |
| **Unikátnosť identifikátora** | Lokálna (vrámci OVM) | Globálna vrámci celej SR |
| **Uplatnenie pravidiel tvorby identifikátorov** | Áno | Áno |
| **Povinná registrácia šablóny identifikátora v MetaIS** | Nie | Áno |

Príklady jednotných referencovateľných identifikátorov:

* https://data.gov.sk/id/legal-subject/50349287 - globálny identifikátor právneho subjektu Úrad podpredsedu vlády Slovenskej republiky pre investície a informatizáciu, ktorý vydáva RPO
* https://data.gov.sk/id/egov/isvs/9676 - globálny identifikátor informačného systému verejnej správy, ktorý vydáva MetaIS
* https://data.gov.sk/id/nuts3/SK010 - globálny identifikátor Bratislavského kraja, ktorý vydáva Štatistický úrad

### Stotožňovanie entít

Jedným z problémov pri integrácii je používanie rôznych identifikátorov pre tú istú vec. Sémantické dátové štandardy prinášajú odpoveď aj na tento typ problému. Jedným z týchto problémov sú aj hodnoty číselníkov štatistického úradu, kde 3 rôzne číselníky odkazujú entitu Bratislavský kraj.

V uvedenom príklad teda platí, že ak sa definuje:

<https://data.statistics.sk/id/stat-region/1> owl:sameAs <https://data.gov.sk/id/nuts3/SK010>

a súčasne aj

<https://data.statistics.sk/id/stat-region/1> owl:sameAs <https://data.statistics.sk/id/region/2>

Na základe týchto dvoch vzťahov sú vytvorené vzájomné väzby medzi všetkými identifikátormi obojsmerne. Pre ilustrovanie sily sémantických vzťahov a prínosu sémantických dátových štandardov v oblasti integrácií sa na obrázku nachádzajú aj iné entity, ktoré odkazujú vždy niektorú z variant Bratislavského kraja. Za referenčný identifikátor považujeme ten podľa európskeho NUTS identifikátora (na obrázku znázornený zelenou farbou).



Obrázok 1: Príklad stotožnenia prostredníctvom sémantických väzieb

Na základe týchto vzťahov je dopytovanie sa na ktorýkoľvek identifikátor Bratislavského kraja unifikovaný tj. všetky tri identifikátori sú ako kebyže považované za jednu entitu a vzťahy sú naviazané na ňu.



Obrázok 2: Ukážka výsledku stotožnenia sémantickými vzťahmi

Použitie sémantických mapovacích mostíkov predstavuje riešenie problému postupného nábehu na referencovateľné identifikátori resp. transformáciu interných identifikátorov jednotlivých systémov na globálne referenčné identifikátori vyjadrené jednotným referencovateľným identifikátorom.

### SHACL

Sémantické dátové štandardy si najmä kvôli dátovej reprezentácii vo forme grafu (sieť) vyžadujú nadnárodný spôsob dátovej a sémantickej validácie. Pre tieto účely je v novele výnosu o štandardoch definovaný európsky štandard SHACL[[8]](#footnote-8).

### Formáty serializácie sémantických dátových štandardov

Pri práci so sémantickými dátami je potrebné používať štandardizovaný formát na výmenu dát. Pre tento účel existuje množstvo výmenných formátov, ktoré sú schopné pokryť problematiku otvorených údajov. My sme vybrali tri najpoužívanejšie a tie sme sa snažili porovnať. Serializáciu sémanticky popísaných dát (zápis dát a dátových štruktúr do štandardizovaného formátu) je možné vykonávať viacerými spôsobmi.

Medzi najčastejšie používané formáty patria:

* JSON-LD
* RDF/XML
* N-Triples
* Turtle
* N-Quads
* Trig

**JSON-LD**

Výhody

* Možnosť zachytiť aj podgraf/kontext grafu
* Jednoznačnosť pri poradí prvkov
* Úsporný formát na veľkosť
* JSON v praxi často používaný
* Najmodernejší z formátov

Nevýhody

**RDF/XML**

Výhody

* XML formát
* Ľahká čitateľnosť

Nevýhody

* „Ukecaný“ formát
* Nemožnosť zachytiť podgraf

**N3**

Výhody

* Natívne zapísané trojice

Nevýhody

* Neznámy formát
* „Ukecaný“ formát
* Nemožnosť zapísať podgraf

I keď je použitie závislé od možností integrujúcich sa strán, pri serializácii je preferovaný formát JSON-LD, ktorý je schopný pokryť linkované údaje ako z pohľadu obsahu tak aj použiteľnosti pri programovaní. Príklady jednotlivých serializácií sa nachádzajú v prílohe (kapitola 7.3).

## Otvorené údaje: 4 vs 5 hviezdičkové data

### Definícia

Stupeň interoperability medzi publikovanými údajmi je možné definovať na základe jasne stanovenej tabuľky.

|  |  |  |
| --- | --- | --- |
| **Úroveň** | **Popis** | **Formáty** |
| 1★ | Sprístupnenie údajov (v akomkoľvek formáte) pod otvorenou licenciou | pdf, doc |
| 2★ | Údaje sú sprístupnené v štruktúrovanej forme (napr. Excel namiesto obrázka s tabuľkou) | xls |
| 3★ | Sprístupnenie údajov v otvorenom formáte (napr. CSV, alebo ODS namiesto Excelu) | odt, csv, tsv, json, xml,  GeoJson, GML, WKT, KML |
| 4★ | Použitie URI na identifikáciu ľubovoľných entít (dátové prvky, údaje) tak, aby ich bolo možné referencovať. Údaje sú popísané ontológiami, avšak tie nepatria do Centrálneho modelu údajov verejnej správy. | rdf/xml, trig, json-ld, trix |
| 5★ | Zladenie/prepojenie údajov s ostatnými, aby bol jasný ich význam. V prípade otvorených verejných údajov SR ide o ich popísanie Centrálnym modelom údajov verejnej správy jednotne mapovaným na medzinárodné ontologické štandardy.  Pre dosiahnutie tejto úrovne je potrebné, aby dataset obsahoval minimálne 60% údajov, ktoré sú referenčné alebo sú súčasťou Centrálneho modelu údajov, zvyšné údaje musia spĺňať nároky na 4 hviezdičky. | rdf/xml, trig, json-ld, trix[[9]](#footnote-9) |

Dôležitým rozdielom medzi kvalitou interoperability je rozdiel medzi úrovňami 4 a 5, ktoré podliehajú sémantickým dátovým štandardom.

### Rozdiel medzi 4 a 5 hviezdičkovými dátami :

**4 hviezdičky**

* použitie referencovateľných identifikátorov na identifikáciu entít
* použitie vlastných ontológií na dátový model
* nutnosť mapovania ontológií v najväčšej možnej miere na 5 hviezdičiek na zachovanie čo najväčšej interoperability
* dereferenciácia referencovateľných identifikátorov je voliteľná
* zverejnenie vlastných ontológií je odporúčané
* zaregistrovanie šablón referencovateľných identifikátorov na MetaIS je voliteľné
* pri registrácii je nutné vždy požiadať Centrálnu dátovú kanceláriu na UPPVII o zaradenie dát medzi 5 hviezdičkové. Len v prípade odmietnutia sa stávajú 4 hviezdičkové.
* negarantujú dátovú interoperabilitu medzi systémami nakoľko nie sú súčasťou centrálneho modelu údajov
* referencovateľné identifikátory musia byť tvorené metodikou, ktorá je súčasťou Sémantické dátové štandardy ISVS

**5 hviezdičiek**

Všetko zo 4 hviezdičiek, ale :

* reprezentujú centrálny model verejnej správy prostredníctvom ontológií
* sú používané jednotné referencovateľné identifikátorom
* model údajov je pod prísnou správou pracovnej skupiny PS1 a zaručuje úplnú interoperabilitu údajov v prostredí verejnej správy
* doména všetkých URI je https://data.gov.sk
* dereferenciácia je zabezpečená integráciou MetaIS a portálu data.gov.sk

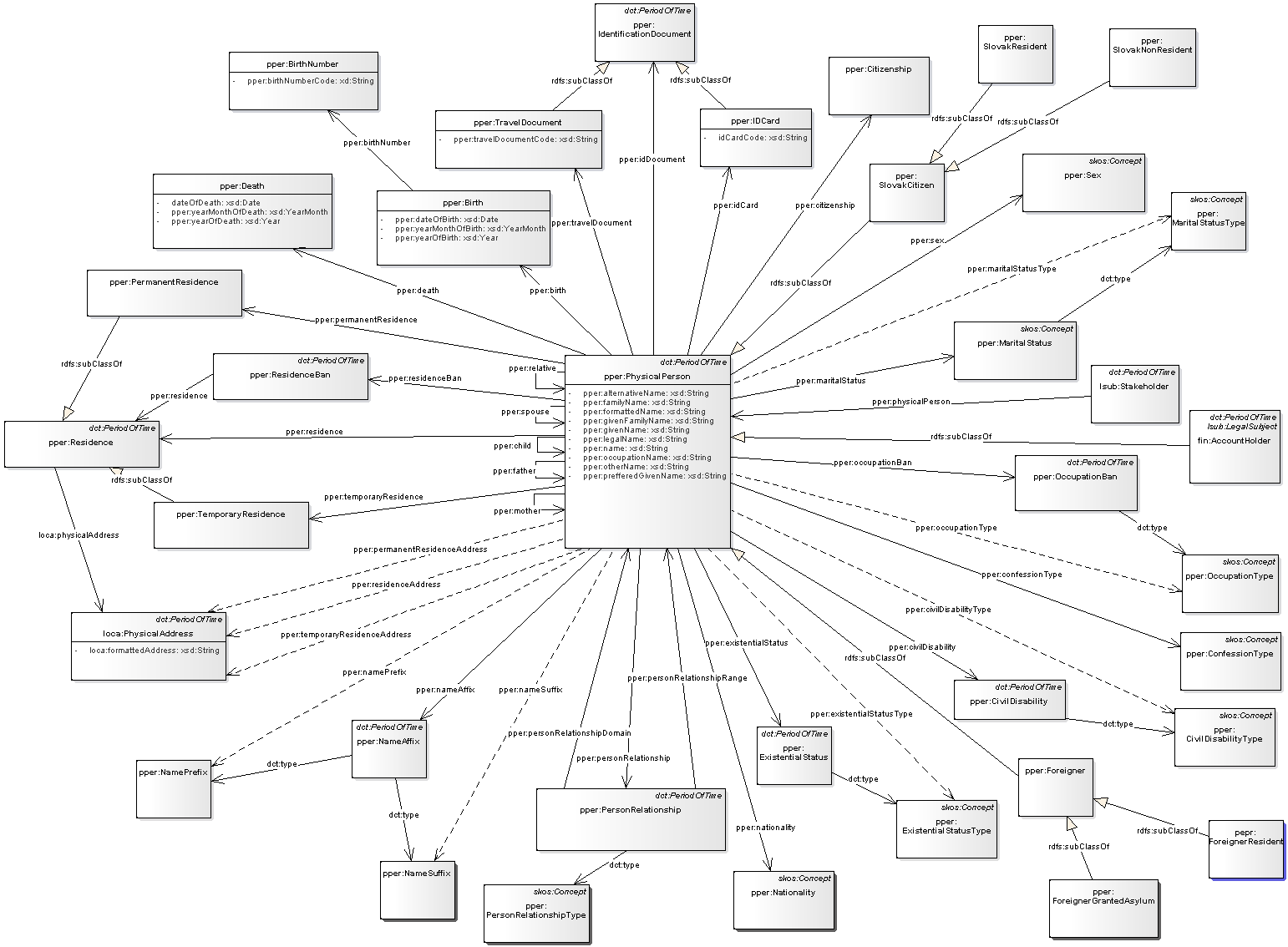
Dôležitým elementom je, že každá entita musí mať ambíciu stať sa 5 hviezdičkovou entitou. Pracovná skupina PS1 bude dozerať a schvaľovať žiadosti o priradenie entity do centrálneho modelu údajov.

## Štandardizované prvky

Centrálny model údajov prešiel počas doby svojej tvorby viacerými pracovnými ako aj finálnymi verziami. Aktuálna verzia Centrálneho modelu údajov je 1.1 a pracuje sa na ďalšom rozširovaní. Pre zvýšenie prehľadnosti o aktuálnom stave sa v dokumente nachádzajú aktuálne UML diagramy, ktoré reprezentujú existujúci stav Centrálneho modelu údajov a jeho rozsah. Diagramy v nasledovných kapitolách sú popísané prostredníctvom UML notácie. Neprerušované čiary reprezentujú priame spojenia medzi dátovými prvkami. Prerušované čiary reprezentujú odvodenú (implicitnú) reláciu medzi dátovými prvkami.

### Ontológia fyzickej osoby

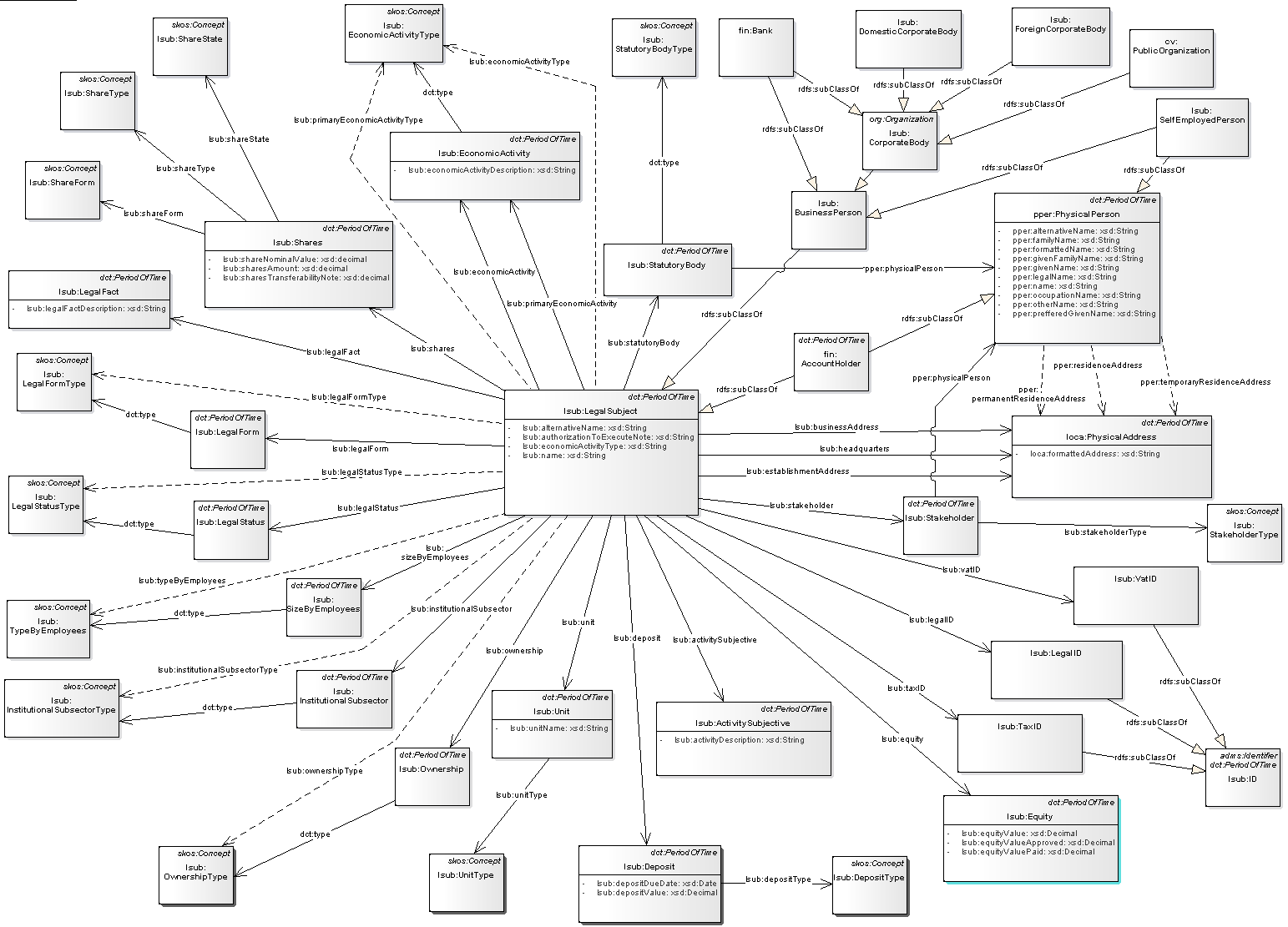
Register fyzických osôb je aktuálne vyhlásený za referenčný register. V rámci tejto aktivity boli vyhlásené mnohé referenčné údaje, ktoré k danému referenčného registru prislúchajú. Na základe vyhlásených referenčných údajov bol vytvorený model fyzickej osoby, ktorý bol zaradený do Centrálneho modelu údajov. Jeho RDF/XML reprezentácia je dostupná na stránkach UPVII[[10]](#footnote-10).



Obrázok 3: Dátová schéma fyzickej osoby

### Ontológia právneho subjektu

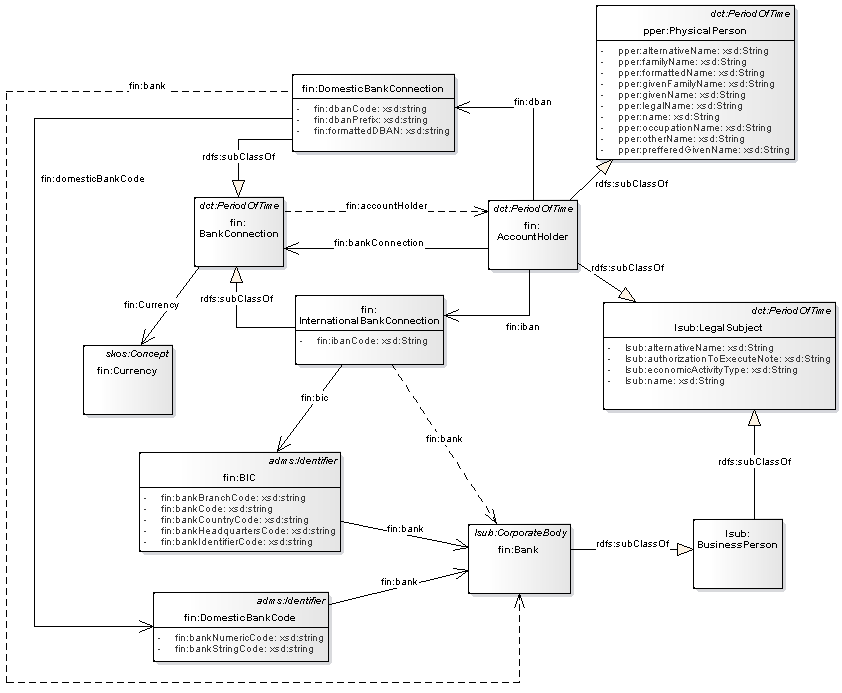
Register právnických osôb je aktuálne vyhlásený za referenčný register. V rámci tejto aktivity boli vyhlásené mnohé referenčné údaje, ktoré k danému referenčného registru prislúchajú. Na základe vyhlásených referenčných údajov bol vytvorený model právneho subjektu, ktorý bol zaradený do Centrálneho modelu údajov. Jeho RDF/XML reprezentácia je dostupná na stránkach UPVII[[11]](#footnote-11).



Obrázok 4: Dátová schéma právneho subjektu

### Ontológia finančných entít

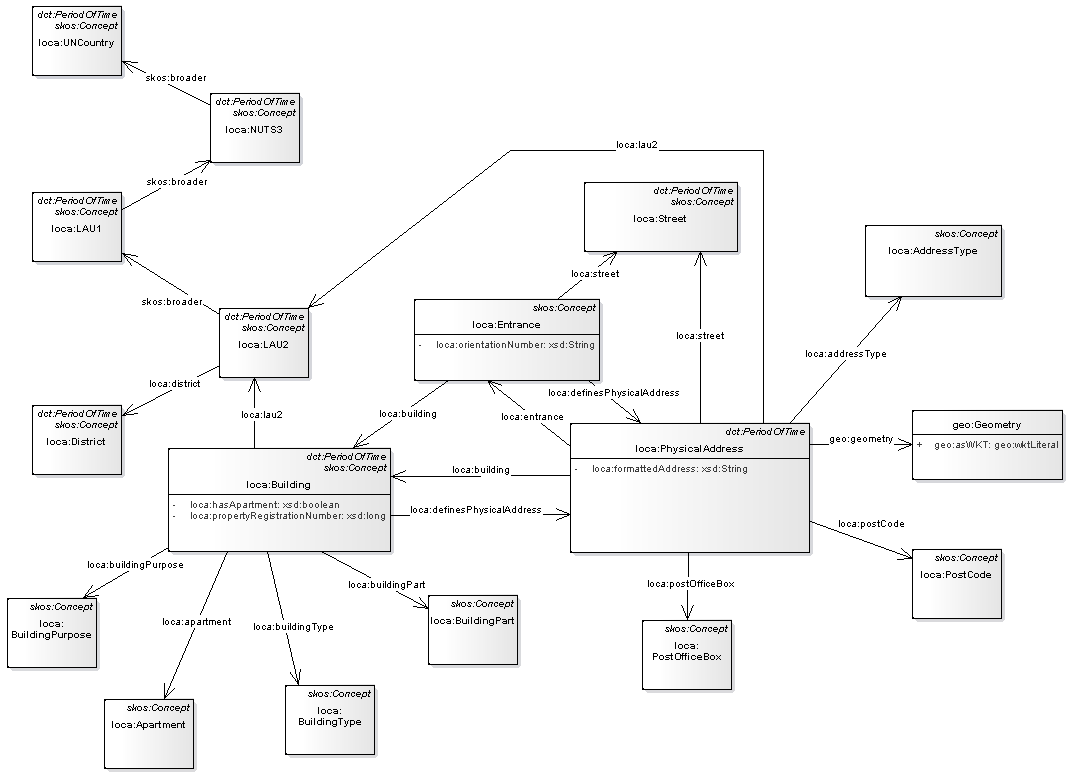
Na vyhlasovania referenčných údajov v oblasti Finančnej správy bol vyhlásený len jeden referenčný údaj – IČ DPH. V tomto ohľade šiel Centrálny model údajov nad rámec referenčných údajov a dátová schéma obsahuje aj iné, často používané, dátové entity. Jeho RDF/XML reprezentácia je dostupná na stránkach UPVII[[12]](#footnote-12).



Obrázok 5: Dátová schéma finančných entít

### Ontológia lokácie

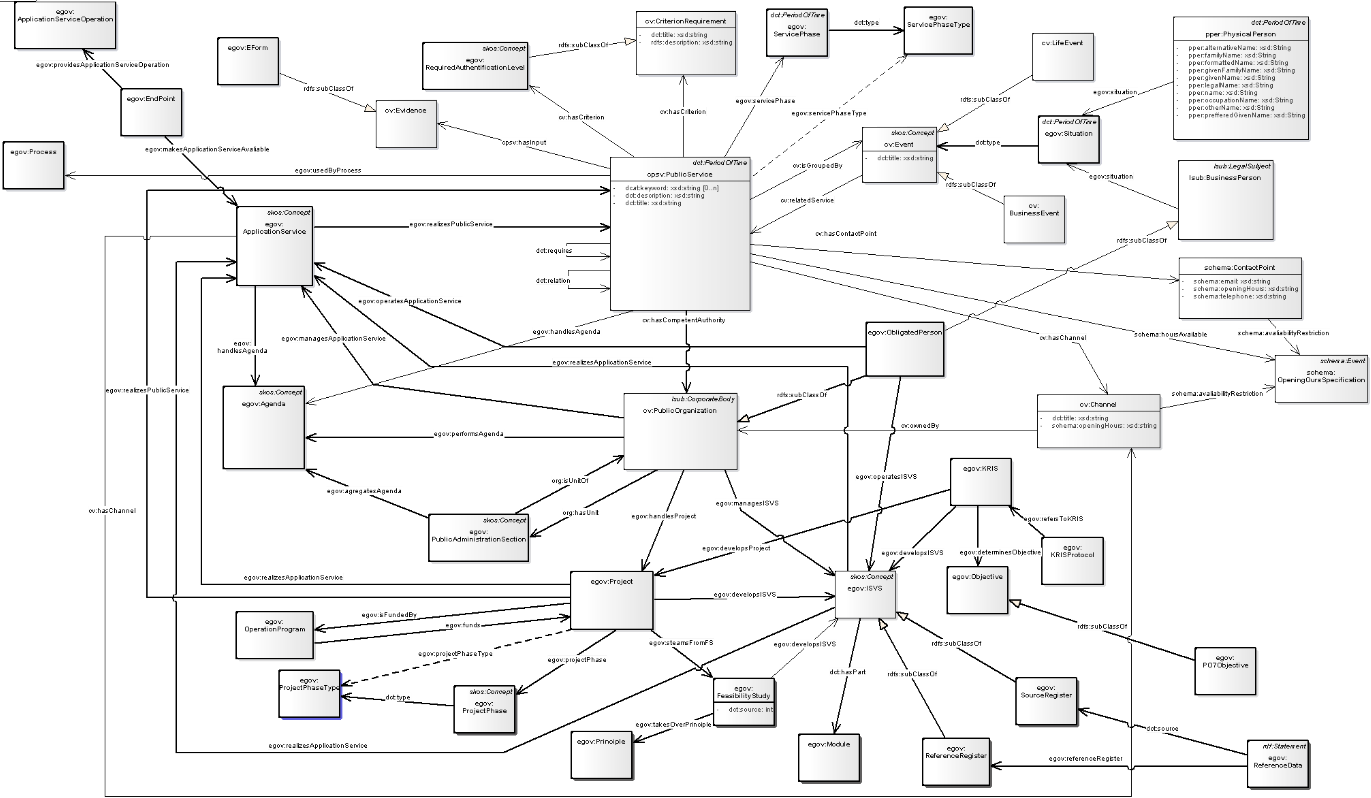
Referenčné údaje v oblasti lokality aktuálne absentujú. Jedným z dôvodov je nejednoznačnosť referečnosti pre jednotlivé dátové entity. V tejto oblasti je prienik v referečnosti medzi základnými číselníkmi (obce, okresy, kraje, štáty) a pripravovaného referenčného Registra adries (adresný bod, ulica, mestská časť). Centrálny model údajov v oblasti entít lokality predbehol proces vyhlasovania a definoval dátový model entít a ich prepojení. Jeho RDF/XML reprezentácia je dostupná na stránkach UPVII[[13]](#footnote-13).



Obrázok 6: Dátová schéma lokácie

### Ontológia egovernment entít

V oblasti eGovernment entít neprebehol proces vyhlasovania referenčných údajov. Aktuálne je táto problematika adresovaná v pripravovanej dopytovej výzve na Metainformačný systém, ktorý má v tejto oblasti priniesť určitý posun. Nakoľko je problematika eGovernment entít veľmi dôležitá v tejto oblasti bol vytvorený komplexný dátový model, ktorý je súčasťou Centrálneho modelu údajov. Jeho RDF/XML reprezentácia je dostupná na stránkach UPVII[[14]](#footnote-14).



Obrázok 7: Dátová schéma eGovernment entít

## Medzinárodné ontológie

Pri tvorbe Centrálneho dátového modelu údajov sa nevytvárajú vždy nové dátové prvky alebo vzťahy. Množstvo všeobecných modelov, ktoré majú medzinárodný presah:

* RDF
* RDFS
* OWL
* Dublin Core
* Dublin Core Terms
* SKOS
* Foaf
* Schema.org
* DCAT
* ADMS
* Core Vocabularies – ISA2

## Katalóg dátových prvkov

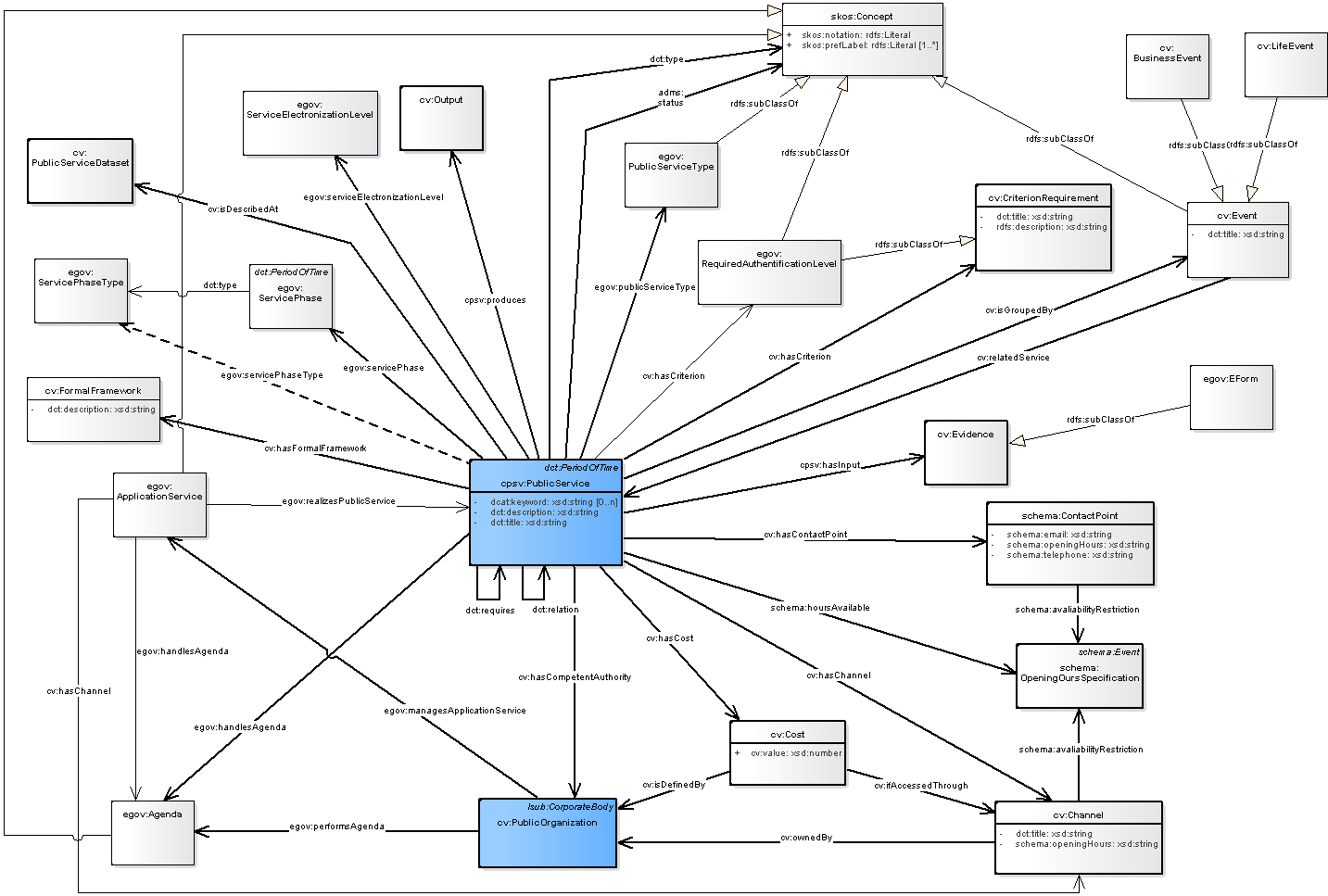
Vývoj dátových štandardov má dlhú históriu a už v minulosti nastali pokusy ako štandardizovať dátové prvky, ktoré sú vymieňané v štáte. Priamym predchodcom Centrálneho modelu údajov verejnej správy je práve Katalóg dátových prvkov.

Daný štandard obsahoval nevýhody, ktoré boli práve prostredníctvom CMU adresované :

* Len pre XML, bez uvažovania iných formátov
* Nereferencovateľnosť schémy – nepoužívanie globálnych identifikátorov dátových prvkov (Jednotný referencovateľný identifikátor)
  + nemožnosť na centrálnej úrovni validovať dátové prvky v službách, formulároch, (textová dĺžka jednotlivých dátových prvkov, dátové typy, ...)
* Nízka úroveň abstrakcie – Katalóg dátových prvkov vyžadoval presné používanie daných elementov. V praxi je potrebné si definovať štruktúry modulárne tj. posielanie len tých údajov, ktoré majú pre danú službu zmysel
* Chýba koordinačný orgán, ktorý tvorí podporu pri zavádzaní – existovala iba štandardizačná komisia, ktorá nemala kapacitu na každodenné potreby pri implementácii a osvete

## Metaúdaje katalógu, datasetu, distribúcie otvorených údajov

Štruktúra metaúdajov je definovaná vo výnose o štandardoch 55/2014. Je možné je vizualizovať prostredníctvom diagramu tried nasledovne:



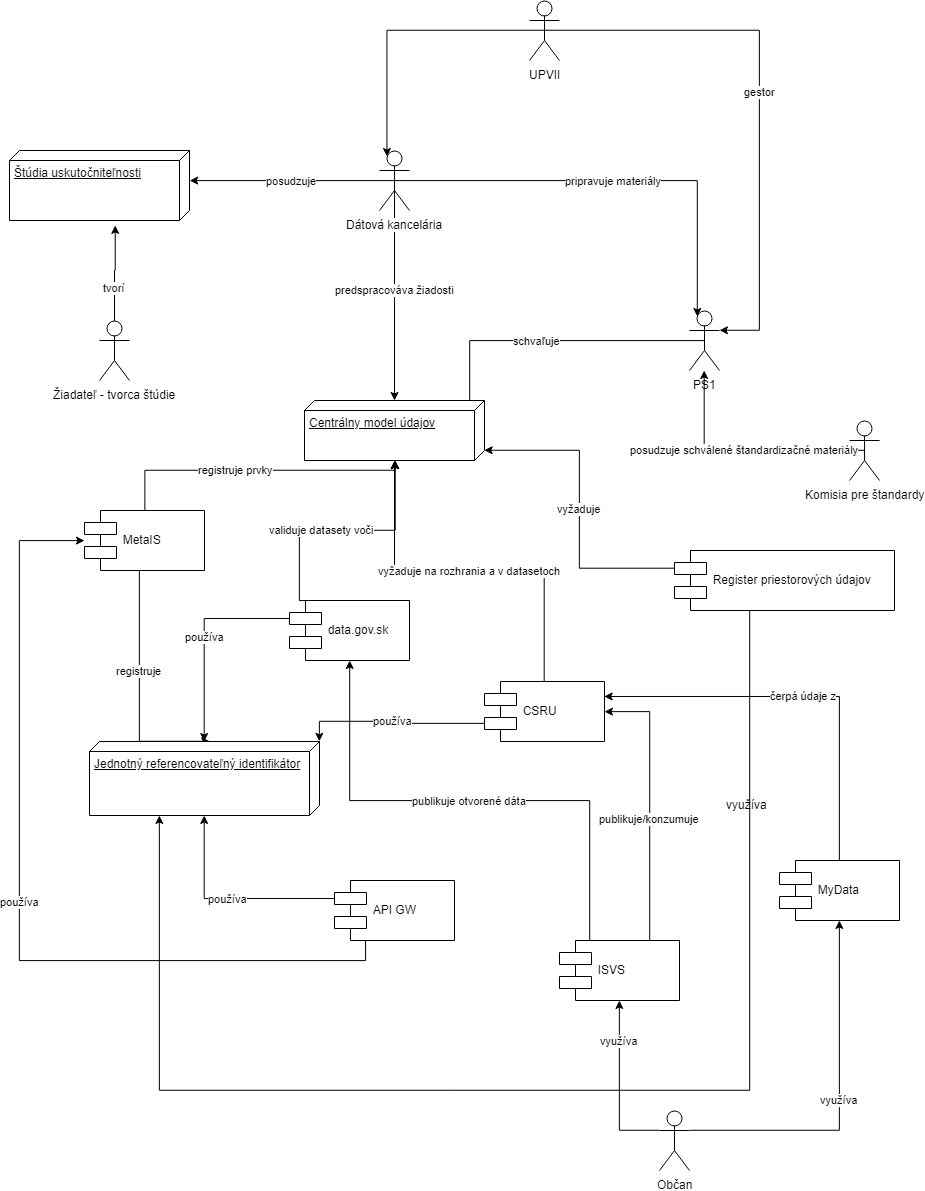
Obrázok 8: Dátová schéma metaúdajov otvorených údajov

Povinnosť ako aj popis atribútov je definovaný v prílohe 9 k výnosu o štandardoch 55/2014. Pre prehľadnosť je daná tabuľka súčasťou aj tohto dokumentu v prílohe (kapitola 7.4).

# Aplikácia Centrálneho modelu údajov v eGov

## Kompetencie

Uvedené kompetencie a vzťahy reprezentujú finálny stav aplikácie CMU do praxe. Aktuálne nie sú všetky vzťahy plne legislatívne alebo metodicky predpísané. Nakoľko má tento dokument s časti slúžiť aj ako podklad pre budúce metodické usmernenie v tejto oblasti, popísané kompetencie v ňom budú detailne rozpísané.



Obrázok 9: Koncepcia využitia Centrálneho modelu údajov

**UPPVII -**hlavný orgán zodpovedný za sémantické dátové štandardy, koordináciu a tvorbu Centrálneho modelu údajov verejnej správy založeného na ontológiách

**Dátová kancelária** je centrálny koordinačný a výkonný aparát, ktorý vykonáva nasledovné aktivity:

* Manažment kvality údajov,
* Riadenie vyhlasovania referenčných údajov a prepájanie údajov,
* Podpora realizácie služby Moje dáta a správa údajov o subjekte,
* Podpora zavádzania analytického spracovania údajov,
* Podpora zvýšenia rozsahu a kvality otvorených údajov,
* Podpora pri úprave legislatívy na zabezpečenie efektívnejšieho riade a správy dát,
* Organizovanie školení a workshopov ako pristupovať k údajom.

Pri jej práci je dôležitá hlavne spolupráca s všetkými aktérmi a systémami popísanými nižšie.

**Pracovná skupina PS1** je schvaľovateľ štandardov a garant kvality. Návrhy dátových prvkov a jednotných referencovateľných identifikátorov a  tvoria do pracovnej skupiny robia

**MetaIS** vystupuje v úlohe referenčného registra URI (jednotných referencovateľných identifikátorov) a štandardov viažucich sa na URI. Zároveň je zasadený do pozície technologického podporného komponentu pre dátovú kanceláriu, ktorá pomáha v správe Centrálneho modelu údajov a problematiky identifikátorov a sémantických dátových štandardov v štátnej správe. Dôležitým prvom je prepojenosť MetaIS s nástrojom BizzDesign, kde prostredníctvom metodiky pre dátové modelovanie je možné systematickým spôsobom zberať a manažovať dáta v národných ISVS.

**CSRÚ -**centrálna integračná platforma, ktorá aplikuje sémantické dátové štandardy pri integračných zámeroch. Prostredníctvom zbernice dátovej integrácie vyžaduje používanie jednotných referencovateľných identifikátorov na rozhraniach ako aj popísanie služieb použitím dátových prvkov Centrálneho modelu údajov.

**API GW** - API gateway komponent, ktorý zavádza sémantické dátové štandardy na úrovni služieb

**OpenData -**platforma otvorených údajov (data.gov.sk) ktorá svojou činnosťou a softvérovými prostriedkami pomáhajú zavedeniu Centrálneho modelu údajov a sémantických dátových štandardov pre oblasť otvorených údajov

**Moje Dáta -**platforma údajov o občanovi, ktorá zavádza sémantické dátové štandardy na úrovni verejnosti

**Register priestorových informácií -** zjednocuje tvorbu, poskytovanie a aktualizáciu metaúdajov o priestorových informáciách, ako aj vytvorenie a aplikovanie princípov harmonizácie a interoperability priestorových informácií. Jedným z dôležitých rysov RPI pre manažment údajov je zosúladenie sa s Centrálnym metainformačným systémom verejnej správy, od ktorého bude RPI preberať pridelené Jednoznačné referencovateľné identifikátory pre priestorové údaje, čím sa dosiahne celková vzájomná interoperabilita údajov z RPI a ostatných referenčných registrov. Identifikátory tvoria základ pri validácii metaúdajov spravovaných v registri. RPI predstavuje jeden zo základných nástrojov na manažment priestorových údajov v Slovenskej republike a jeho aktívnym využívaním je možné do značnej miery eliminovať nedostatky v súčasnom stave priestorových údajov v SR.[[15]](#footnote-15)

## Číselníky

Materiál je návrhom členov štandardizačnej skupiny PS1. Aktuálne návrh čaká štandardizačný proces a jeho zapracovanie do výnosu o štandardoch.

Jedným zo základných dátových elementov vo formulároch je Codelist element. Element je prostredníctvom XSD schémy definovaný cez typ [CodelistCType](https://wiki.finance.gov.sk/pages/viewpage.action?pageId=14713058) . Na základe harmonizácie použitia sémantických dátových štandardov v prostredí verejnej správy je nutné rozšíriť správanie daného elementu. Definícia schémy číselníkov je definovaná v prílohe 2 vo výnose o štandardoch 55/2014.

**Všeobecné podmienky použitia**

* Je vyžadované použitie číselníkových hodnôt v Codelist namiesto NonCodelistData Elementom vždy keď je to len možné aby sa minimalizovali nečíselníkové integrácie. Za výnimku je možné považovať zoznam zahraničných miest. Tento číselník má neustále rastúcu tendenciu nakoľko je povinnosť zapísať mestá tak ako ich oznámi občan. Nakoľko je zoznam zahraničných miest obrovský a nie je možné ho považovať v dohľadnej dobe za úplný nie je možné použiť Codelist Element a je nutné použíť NonCodelist Element.
* Codelist Element sa nesmie používať na enumeration, nakoľko enumeration je základnou funkcionalitou XSD. Rozdiel medzi číselníkom a enumeration je, že číselník musí mať URI identifikátor pre každú položku číselníka. Enumeration nemá identifikátor.
* Codelist Element obsahuje ľubovoľný číselník, ktorý je zaregistrovaný v MetaIS (nemusí byť základný) a jeho tvorca garantuje jeho stálosť a trvalú unikátnosť - nebude možné daný URI identifikátor prideliť inému číselníku a zároveň nebude možné zmeniť ani jeho štruktúru. Každý číselník, ktorý má byť použitý v Codelist Elemente musí byť v MetaIS zaregistrovaný, aby mohol byť používaný rôznymi systémami a mohol byť dereferencovaný.
* Nové služby alebo systémy, ktoré vzniknú a používajú Codelist Element majú povinnosť používať URI identifikátor pre číselník bez verzie. Zároveň aj položky číselníka sa musia reprezentovať ako jednoznačný URI identifikátor aj pre jednotlivé položky číselníka v prípade, že dané položky majú referenčné URI.

Pre maximalizáciu integračného potenciálu je potrebné zachovávať entity z predchádzajúcich verzií neustále v ďalších verziách. Pokiaľ je daná entita už v danej verzii neplatná je možné využiť vlastnosť z ontológie Resource.

|  |
| --- |
| URI\_ZNEPLATNENEJ\_POLOZKY\_CISELNIKA https://data.gov.sk/def/ontology/resource/isValid "false"^^xsd:boolean |

Ak je potrebné z integračných dôvodov vedieť aj časovú platnosť položky číselníka, je možné použiť vzťahy z resource ontológie :

https://data.gov.sk/def/ontology/resource/validFrom

https://data.gov.sk/def/ontology/resource/validTo

**Element CodelistCode**

* obsahuje identifikáciu číselníka prostredníctvom URI a to bez verzie. Hlavným dôvodom je spôsob dereferenciácie, kde identifikátor číselníka bez verzie je automaticky považovaný za poslednú verziu. Systémy pri používaní deklarujú, že pracujú s poslednou verziou číselníka, pričom rozhodujúci je čas a dátum dopytu na systém, ktorý vracia Codelist element.
* V prípade nereferenčných číselníkov, ktoré rozširujú existujúce referenčné/základné číselníky je nutné zabezpečiť aby tie položky číselníka, ktoré majú pridelené referenčné URI, tak práve toto URI bolo súčasťou číselníka.

ItemCode - ak existuje URI identifikátor položky, tak sa používa ten.

Príklad Codelist elementu pre základný číselník:

<Codelist>

<CodelistCode>https://data.gov.sk/set/codelist/CL004001</CodelistCode>

<CodelistItem>

<ItemCode>https://data.gov.sk/def/identifier-type/7</ItemCode>

<ItemName>IČO (Identifikačné číslo organizácie)</ItemName>

</CodelistItem>

</Codelist>

**Príklad Codelist Elementu pre číselník RFO obcí**

* tento číselník rozširuje základný číselník obcí LAU2 : https://data.gov.sk/set/codelist/CL000025
* nakoľko číselník rozširuje existujúci referenčný číselník, vracia referenčné URI identifikátory v časti ItemCode v prípade, že ide o položku, ktorá sa nachádza aj v základnom číselníku a vlastný URI identifikátor pre prvky, ktoré sa v danom číselníku nenachádzajú

**Príklad s URI identifikátorom prvku v číselníku, ktorý nie je súčasťou základného číselníka**

Ak nereferenčný číselník nie je rozšírením referenčného číselníka (ako v príklade RFO číselníka obcí), používajú sa vždy URI identifikátory nereferenčného URI charakteru (nemajú prefix https://data.gov.sk/...)

<Codelist>

<CodelistCode>https://rfo.sk/set/codelist/ciselnik-obci</CodelistCode>

<CodelistItem>

<ItemCode>https://rfo.sk/id/lau2/123456</ItemCode>

<ItemName>Názov\_nejakej\_obce\_ktora\_nieje\_v\_zakladnom\_ciselniku\_CL000025</ItemName>

</CodelistItem>

</Codelist>

**Príklad s URI identifikátorom položky číselníka, ktorá je súčasťou základného číselníka:**

<Codelist>

<CodelistCode>https://rfo.sk/set/codelist/ciselnik-obci</CodelistCode>

<CodelistItem>

<ItemCode>https://data.gov.sk/id/lau2/SK0233500011</ItemCode>

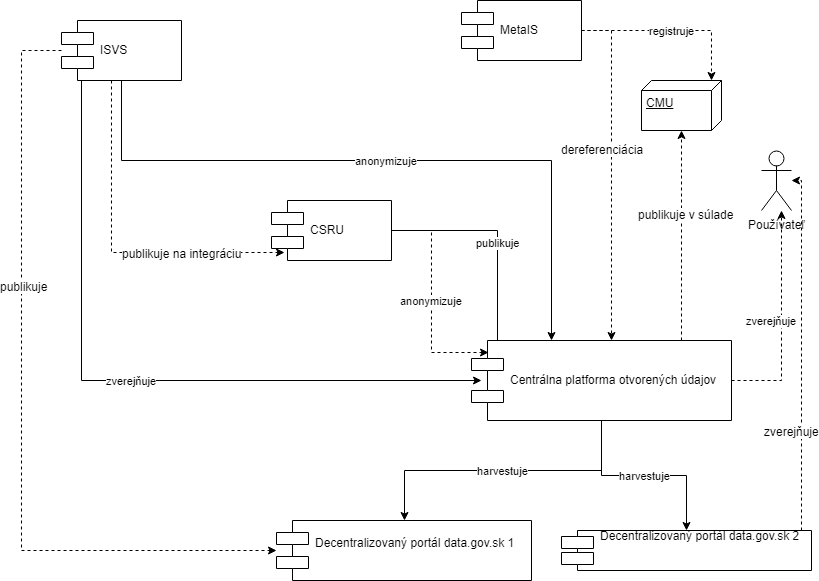
<ItemName>Nitra</ItemName>

</CodelistItem>

</Codelist>

## CSRÚ a otvorené údaje

Otvorené údaje sú integrálnou súčasťou stratégie OPII. Na základe tohto predpokladu bola celková stratégia v oblasti otvorených údajov zakomponovaná do modelu platformy dátovej integrácie. Komunikácia pri publikácii otvorených údajov na základe existujúcich štúdií uskutočniteľnosti a výstupov štandardizačných skupín a skupín orientujúcich sa na strategické priority je nasledovná.



Obrázok 10: Komunikačná schéma kooperácie platformy otvorených údajov a CSRÚ

## Dereferenciácia

### Predstavenie

Dereferenciácia je proces, kedy je referencovateľný identifikátor (URI) entity zároveň možné použiť aj ako URL lokátor entity. Tým je možné zabezpečiť zobrazenie entity vo webovom prehliadači ako aj získať informácii o entite v štruktúrovanom formáte (RDF/XML, JSON, XML...).

Samotné využitie dereferenciácie je možné ilustrovať na príklade referencovateľného identifikátora právneho subjektu (firma, živnosť, nezisková organizácia... ). Organizácia s IČO 12345678 má jednotný referencovateľný identifikátor :

https://data.gov.sk/id/legal-subject/12345678

Bez procesu dereferenciácie by tento identifikátor bol len strojovo čitateľný napríklad na úrovni databázového identifikátora. Vďaka dereferenciácii je možné tento identifikátor vložiť do webového prehliadača. Na základe technických prostriedkov sa zobrazí po vložení do prehliadača detail konkrétnej firmy tj. všetky údaje, ktoré eviduje RPO o danom subjekte. Tým že referencovateľný identifikátor vyzerá ako bežná URL linka, keď občan uvidí takýto identifikátor nebude zmätený a bude ho preto možné používať aj ako identifikátor pri komunikácii s občanom.

Dereferenciácia je proces s veľmi širokým využitím. Druhou skupinou používateľov sú technický pracovníci. Tí môžu na základe špecifikovania Content-Type alebo query parametrov využívať rozšírené možnosti. Ako príklad môže slúžiť znova jednotný referencovateľný identifikátor subjektu s IČO 12345678.

**Príkaz:**

GET https://data.gov.sk/id/legal-subject/12345678

vo webovom prehliadači vráti defaultne HTML vizualizáciu entity tj. detail subjektu ako bežná webová stránka.

**Príkaz:**

Header : Content-Type: application/json

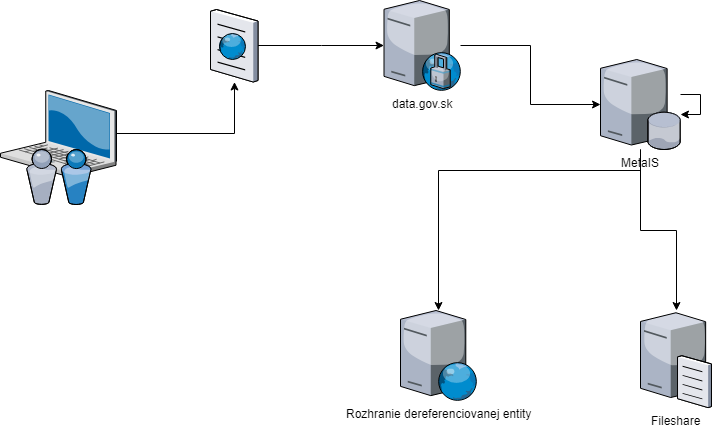
GET https://data.gov.sk/id/legal-subject/12345678

Vráti štruktúrovaný objekt vo formáte JSON, ktorý môže využívať akákoľvek aplikácia či systém a používať ho pre svoje potreby. Tým dochádza k unifikácii identifikátora pre použitie v databázach, pri komunikácii s občanom ako aj pri technických integráciách. Koncepčne je možné dereferenciáciu chápať ako unifikovaný spôsob využiteľný aj pre OpenAPI, pričom jeho prínosy sú pre všetky skupiny.

### Koncept realizácie

Pre zabezpečenie dereferenciácia sú potrebné minimálne 2 nosné informačné systémy :

* Portál otvorených údajov data.gov.sk
* Metainformačný systém.



Obrázok 11: Základná komunikačná schéma dereferenciácie

Úlohou portálu data.gov.sk je presmerovanie volaní spĺňajúce nasledovný vzor (regulárny výraz)

https://data.gov.sk/(id|def|set|doc)/.\*

do Metainformačného systému na adresu:

Nakoľko je MetaIS referenčným registrom referencovateľných identifikátorov a tým pádom sú všetky menné priestory registrované v ňom, MetaIS sa stáva primárnym kameňom pre dereferenciáciu pridelených menných priestorov referencovateľných identifikátorov.

https://metais.finance.gov.sk/refid/(id|def|set|doc)/.\*

Žiadateľ môže v systéme MetaIS požiadať o pridelenie celého menného priestoru. Napr. RPO požiada o zaregistrovanie menného priestoru :

https://data.gov.sk/id/legal-subject/{id}

pre potreby jednotnej referencovateľnej identifikácie RPO entít. Samotný menný priestor je

Po zadaní tohto menného priestoru do webového prehliadača, portál data.gov.sk presmeruje dopyt na :

https://metais.finance.gov.sk/refid/id/legal-subject/

https://data.gov.sk/id/legal-subject/

Nakoľko nie je zadaná presná entita ale menný priestor, MetaIS zobrazí stránku o popise menného priestoru s údajmi ako kto je vlastníkom daného menného priestoru, dátum registrácie a pod.

### **Rozšírené parametre dereferenciácie**

Žiadateľ môže v systéme MetaIS požiadať o pridelenie celého menného priestoru tj. identický proces ako v bode 5.4.2.

**Neverzionovateľné referencovateľné identifikátory**

Tento typ identifikátor sa používa na časovo stále objekty tj. napríklad identifikátor právneho subjektu alebo identifikátor osoby je časovo stály. Okrem týchto identifikátorov sú aj také, ktoré verzionujeme. Typickým príkladom je identifikátor formulárov, v ktorých je verzia dôležitou súčasťou.

RPO požiada o zaregistrovanie menného priestoru :

https://data.gov.sk/id/legal-subject/{id}

pre potreby tvorby jednotného referencovateľného identifikátora RPO entít. Samotný vlastník menného priestoru RPO, môže definovať šablónu na presmerovania pri dereferenciácii v dvoch možnostiach

a) Použitím celého jednotného referencovateľného identifikátora tj. využitie nahradzujúceho vzoru {uri} napr.

https://rpo.statistics.sk/rpo/view?uri\_parameter={uri}

// Konkrétny príklad pre UPVII

https://rpo.statistics.sk/rpo/view?uri\_parameter=https://data.gov.sk/id/legal-subject/50349287

b) Použitím ID časti tj. využitie nahradzujúceho vzoru {id} .  Pozn. názvy parametrov

https://rpo.statistics.sk/rpo/view?id\_parameter={id}

alebo

https://rpo.statistics.sk/rpo/view/{id}

// Konkrétny príklad pre UPVII

https://rpo.statistics.sk/rpo/view?id=50349287

https://rpo.statistics.sk/rpo/view/50349287

Entita Úradu podpredsedu vlády SR pre investície a informatizáciu pre daný menný priestor by vyzerala nasledovne :

https://data.gov.sk/id/legal-subject/50349287

Po zadaní tohto jednotného referencovateľného identifikátora do webového prehliadača, portál data.gov.sk presmeruje dopyt na :

V prípade, že RPO vyplní šablónu na presmerovanie, MetaIS spraví presmerovanie na stránku registrátora podľa definovaného zdroja.

https://rpo.statistics.sk/rpo/view?id\_parameter=50349287

alebo

https://rpo.statistics.sk/rpo/view/50349287

https://metais.finance.gov.sk/refid/id/legal-subject/50349287

V prípade, že RPO nevyplní šablónu na presmerovanie, MetaIS zobrazí stránku o popise menného priestoru s údajmi ako napríklad kto je vlastníkom daného menného priestoru, dátum registrácie a pod.

Názvy parametrov id\_parameter, uri\_parameter a verzia\_parameter v príklade sú ilustratívne a vlastník šablóny jednotného referencovateľného identifikátora si ich môže pomenovať ľubovoľne. Placeholdery {id},{uri},{version} sú dané a na ich základe je možné vytvoriť šablónu na presmerovanie.

### **Dereferenciácia súborov**

Primárne bola dereferenciácia súborov požadovaná pre pokrytie požiadaviek elektronických formulárov a ich súčastí. Samotný princíp bol pri štandardizácii abstrahovaný aby dokázal pokryť oblasť dereferenciácie súborov aj pre typy použitia. Preto akýkoľvek zdroj súborových údajov, ktoré je potrebné referencovať medzi systémami má vytvorené potrebné štandardizačné podhubie. Pri aplikovaní tohto štandardu sa zjednoduší a stransparentňuje spôsob referencovania súborov v celom eGov.

Verzionovateľné entity typu doc predstavujú špeciálnu kategóriu pre potreby dereferenciácie. Hlavným rozdielom je referencia statických súborov v namapovanom úložisku.

Žiadateľ môže v systéme MetaIS požiadať o pridelenie celého menného priestoru. Napr. MEF (modul elektronických formulárov) požiada o zaregistrovanie menného priestoru pre úložisko na dereferenciáciu súborov z definičného balíčka elektronických formulárov:

https://data.gov.sk/doc/egov/eform/{id}/{version}

Samotný unikátny menný priestor je

https://data.gov.sk/doc/egov/eform/

Entita v konkrétnej verzii, ktorý je zároveň jednotným referencovateľným identifikátorom pre úložisko je

https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.1

kde za verziou nasleduje cesta v úložisku na súbor schema.xsd

https://data.gov.sk/doc/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.1/schema.xsd

Nakoľko pre dokumenty typu /doc nie je časť referencovateľného identifikátora {id} ani {version} požadovaná, je možné si zaregistrovať aj úložisko ako (ide čisto o príklad, nie reálny referencovateľný identifikátor pre UPVII dokumenty)

A následne je možné pre toto úložisko referovať súbor ako :

https://data.gov.sk/doc/egov/filesystem/upvii/nazov\_dokument.pdf

## Elektronické formuláre

V oblasti elektronických formulárov prebehol rad štandardizačných stretnutí na ktorom bolo komplexne pokryté začlenenie sémantických dátových štandardov. V nasledujúcich kapitolách sa nachádzajú definície formulárov, XSD schém formulárov ako aj XML reprezentácie formulárov. Menný priestor elektronických formulárov je definovaný štandardom a ich predpis súčasťou zoznamu JRI v MetaIS. MEF (Modul elektronických formulárov) vytvára identifikátor na základe ktorého je možné vytvoriť JRI. Tento identifikátor sa následne používa aj pri definícii jednotlivých súčastí balíčka elektronického formulára a jeho menného priestoru.

### [Referencovateľný identifikátor XML menného priestoru a elektronického formulára](https://wiki.vicepremier.gov.sk/pages/viewpage.action?pageId=19661413)

**Vzor referencovateľného identifikátora URI XML menného priestoru a referencovateľného identifikátora elektronického formulára:**

https://[domain]/id/egov/eform/[id]/[version]

[domain] - použitá doména. Entity centrálneho modelu údajov verejnej správy alebo jednotné referencovateľné identifikátory používajú doménu data.gov.sk tj. referencovateľný identifikátor spravovaný modulom elektronických formulárov zaregistrované cez ÚPVS majú vždy doménu data.gov.sk

[id] *=* identifikátor elektronického formulára (vo výnose označovaný ako referencia ) v tvare : ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia

* vytvára sa v slovenskom jazyku bez diakritiky,
* „ICO\_poskytovatela“ obsahuje IČO poskytovateľa elektronického formulára, ak je potrebné doplniť aj nižšiu hierarchickú úroveň, oddeľuje sa bodkou,
* „nazov\_elektronickeho\_formulara“ je vecný názov elektronického formulára podľa bodu 2.2.1 písm. a) prílohy č. 3 výnosu MF SR č. 55/2014 Z. z., určený jeho poskytovateľom, pričom jednotlivé slová sa oddeľujú podtržníkmi,
* „jazyk“ elektronického formulára v zmysle bodu 2.2.1 písm. f) prílohy č. 3 výnosu MF SR č. 55/2014 Z. z. sa uvádza podľa príslušného kódu položky základného číselníka CL010076 Jazyky; neskoršie pridanie novej jazykovej verzie v rámci jedného elektronického formulára nemení pôvodný identifikátor; ak jeden elektronický formulár obsahuje viac ako jednu jazykovú verziu, neuvádza sa žiadna hodnota,
* „doplnujuca\_informacia“ môže byť viacnásobná, oddelená bodkami alebo nemusí existovať.

[version] *=*verzia formulára, ktorá je previazaná a identická s verziou prislúchajúcej XSD schémy

**Príklad referencovateľného identifikátora elektronického formulára v konkrétnej verzii :**

https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.0

Referencovatelné identifikátor XML menného priestoru pre daný formulár je identické s referencovateľným identifikátorom formulára. Referencovateľný identifikátor XML menného priestoru sa používa pre deklaráciu menného priestoru v atribútoch "xmlns" a "targetNamespace".

**Príklad referencovateľného identifikátora XML menného priestoru je:**

https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.0

**Príklad XML hlavičky:**

<ROOT\_ELEMENT\_TAG xmlns="https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

    xsi:schemaLocation="https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.0 https://data.gov.sk/doc/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.0/schema.xsd">

...

</ROOT\_ELEMENT\_TAG>

### [Referencovateľný identifikátor pre súčasti elektronického formulára](https://wiki.vicepremier.gov.sk/pages/viewpage.action?pageId=19661404)

**Vzor jednotného referencovateľného identifikátora pre XSD schému elektronického formulára je:**

https://[domain]/doc/egov/eform/[id\_formulara]/[version\_formulara]/schema.xsd

|  |
| --- |
|  |

[domain] *=* použitá doména. Entity centrálneho modelu údajov verejnej správy alebo jednotné referencovateľné identifikátory používajú doménu data.gov.sk

[id\_formulara] = identifikátor, ktorý je prebraný z parametra id v šablóne elektronického formulára

[version\_formulara] = verzia formuláru, ktorá je prebratá z verzie prislúchajúceho formulára (parameter version v šablóne elektronického formulára) a zároveň je v súlade s metodikou pre tvorbu XSD. Verzia musí byť vždy vyjadrená prostredníctvom regulárneho výrazu a zapísaná v systéme MetaIS.

**Referencovateľné identifikátor elektronického formuláru bez verzie.**

V Kontajneri XML údajov (XMLDataContainer) podľa prílohy č. 11 Výnosu o štandardoch pre IS VS sa používa v atribúte "Identifier" v dátovom prvku XMLData**.**

Nakoľko obsah elektronických formulárov sa podpisuje, Jednotný referencovateľný identifikátor elektronického formuláru bez verzie sa nedereferencuje do poslednej verzie nakoľko tá sa môže meniť v čase čo je pre podpisovanie neprijateľné.

https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia

**Jednotný referencovateľný identifikátor elektronického formuláru v konkrétnej verzii (referencovateľný identifikátor je zároveň aj XML namespace elektronického formuláru)**

https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.1

**URI súboru XSD schémy.**V Kontajneri XML údajov (XMLDataContainer) podľa prílohy č. 11 Výnosu o štandardoch pre IS VS sa používa v dátovom prvku UsedXSDReference v UsedSchemasReferenced**.**

https://data.gov.sk/doc/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.1/schema.xsd

**URI súboru manifest.xml**

**URI transformacie.**V Kontajneri XML údajov (XMLDataContainer) podľa prílohy č. 11 Výnosu o štandardoch pre IS VS sa používa v dátovom prvku UsedPresentationSchemaReference v UsedSchemasReferenced.

https://data.gov.sk/doc/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.1/Content/nazov-suboru.xslt

**URI zip balíčka elektronického formuláru**

https://data.gov.sk/doc/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.1/eform.zip

Ak referencovateľný identifikátor elektronického formulára resp. prislúchajúcej XSD schémy elektronického formulára používa doménu data.gov.sk, je považovaný za jednotný referencovateľný identifikátor. Tým pádom sa na nich vzťahuje dereferenciácia tj. napr. referencovateľný identifikátor XSD schémy je zároveň aj URL danej schémy.

**Príklad hlavičky XSD schémy:**

<?xml version="1.0" encoding="UTF-8"?>

<xs:schema xmlns:xs="[http://www.w3.org/2001/XMLSchema"](http://www.w3.org/2001/XMLSchema) xmlns="https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.0" targetNamespace="https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.0" elementFormDefault="qualified">

...

...

</xs:schema>

### Metaúdaje elektronického formulára

V rámci sémantických dátových štandardov bola navrhnutá sémantická reprezentácia meta.xml súboru, ktorý bude súčasťou zmien v Module elektronických formulárov.

<egov:EForm rdf:about="[https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.1"](https://data.gov.sk/id/egov/eform/ICO_poskytovatela.nazov_elektronickeho_formulara.jazyk.doplnujuca_informacia/1.1)>

<dct:type rdf:resource="[http://purl.org/adms/assettype/ServiceDescription"](http://purl.org/adms/assettype/ServiceDescription)/>

<dct:title xml:lang="sk">Slovenský názov formulára</dct:title>

<dct:title xml:lang="en">Anglický názov formulára</dct:title>

<res:gestor rdf:resource="[https://data.gov.sk/id/legal-subject/12345678"](https://data.gov.sk/id/legal-subject/12345678)/>

<dct:description xml:lang="sk">Popis elektronického formulára</dct:description>

<dct:created rdf:datatype="&xsd;dateTime">2017-01-01T00:00:00Z</dct:created>

<dct:creator rdf:resource="[https://data.gov.sk/id/legal-subject/12345678"](https://data.gov.sk/id/legal-subject/12345678) />

<dct:isVersionOf rdf:resource="[https://data.gov.sk/id/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia"](https://data.gov.sk/id/egov/eform/ICO_poskytovatela.nazov_elektronickeho_formulara.jazyk.doplnujuca_informacia)/>

<dcat:downloadURL rdf:resource="[https://data.gov.sk/doc/egov/eform/ICO\_poskytovatela.nazov\_elektronickeho\_formulara.jazyk.doplnujuca\_informacia/1.1"](https://data.gov.sk/doc/egov/eform/ICO_poskytovatela.nazov_elektronickeho_formulara.jazyk.doplnujuca_informacia/1.1)/>

<owl:versionInfo rdf:datatype="&xsd;string">1.1</owl:versionInfo>

<adms:versionNotes rdf:datatype="&xsd;string">Poznámky k zmenám vo verzii elektronického formulára</adms:versionNotes>

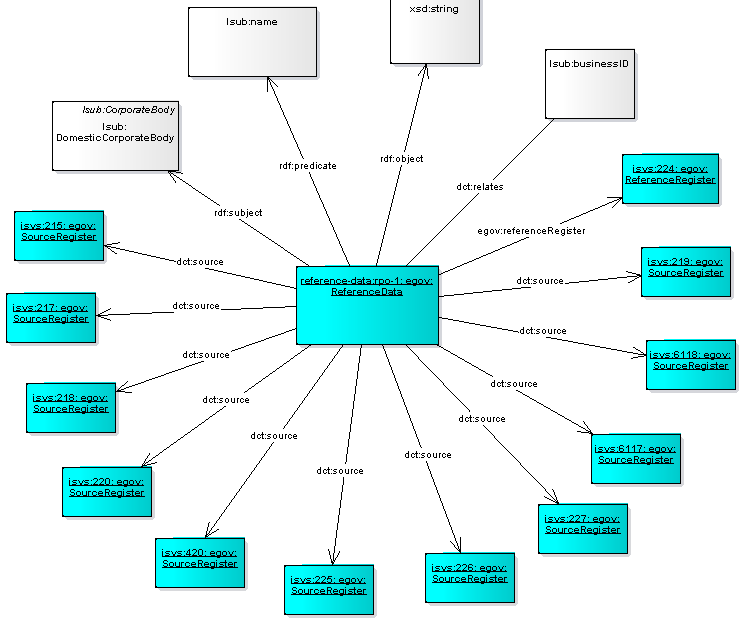
<adms:status rdf:resource="[http://purl.org/adms/status/Completed"](http://purl.org/adms/status/Completed)/>

<res:language rdf:resource="[https://data.gov.sk/def/language/sk"](https://data.gov.sk/def/language/sk)/>

</egov:EForm>

## Referenčné údaje

Pri vyhlasovaní referenčných údajov je potrebné aby ku každému údaju boli definovaná sada metaúdajov. Tieto údaje sú prepísané do sémantickej podoby tak aby aj na tejto úrovni bolo možné používať jednotné referencovateľné identifikátory. Pre ilustráciu dátového modelu je na obrázku 12 zobrazený referenčný údaj názov firmy.



Obrázok 12: Dátová schéma záznamu o referenčnom údaji

Technický zápis vo formáte RDF/XML je definovaný nižšie.

<egov:ReferenceData rdf:about="https://data.gov.sk/def/reference-data/rpo-1">

<rdf:subject rdf:resource="&lsub;DomesticCorporateBody" />

<rdf:predicate rdf:resource="&lsub;name" />

<rdf:object rdf:resource="&xsd;string" />

<dct:relates rdf:resource="&lsub;businessID" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/219" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/6118" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/225" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/226" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/227" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/220" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/217" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/6117" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/215" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/218" />

<dct:source rdf:resource="https://data.gov.sk/id/egov/isvs/224" />

<egov:referenceRegister rdf:resource="https://data.gov.sk/id/egov/isvs/420" />

<rdfs:isDefinedBy rdf:resource="https://data.gov.sk/def/reference-data/rpo/2018-03-29"/>

</egov:ReferenceData>

# Procesy

Pre správu sémantických dátových štandardov (pod ktoré spadá aj Centrálny model údajov) sú navrhnuté procesy, ktoré pokrývajú životný cyklus rozširovania, kompetencií a údržby jednotlivých prvkov.

## Priradenie jednotného referencovateľného identifikátora

### Popis

**Pravidlá**

1. Prideľovanie a schvaľovanie dátových prvkov a jednotných referencovateľných identifikátorov prebieha výhradne prostredníctvom systému MetaIS, kde sa rozhodne, či sa prvok zaradí do Centrálneho modelu údajov verejnej správy, alebo či bude zaradený medzi 4 hviezdičkové dáta (vtedy bude iba referencovateľným identifikátorom a nie jednotným referencovateľným identifikátorom).
2. Každý zaregistrovaný dátový prvok alebo ontológia, ktoré sú schválené ako súčasť Centrálneho modelu údajov sú automaticky referencovateľné na základe integrácie MetaIS a portálu data.gov.sk podľa príslušných pravidiel dereferenciácie
3. Menný priestor je pre indivíduá je možné označiť za referenčné, len v prípade že je zdrojom referenčný register.
4. Každý referenčný identifikátor musí byť zaregistrovaný a schválený v MetaIS.
5. Štandardizácia ontológií centrálneho modelu údajov je schvaľovaná na pracovnej skupine PS1.
6. Koordinačný mechanizmus správy Centrálneho modelu údajov má na starosti Centrálna dátová kancelária na ÚPPVII
7. Čiastkové úpravy alebo doplnenia štandardizovaných dátových prvkov môžu prejsť cez zrýchlené schvaľovacie konanie s upovedomením všetkých členov PS1 do lehoty 15 dní od požiadania.
8. Každá žiadosť o zriadenie štandardizovaného dátového prvku do Centrálneho modelu údajov prechádza schvaľovacím konaním. Žiadosť môže byť zamietnutá v prípade

* dátový prvok s rovnakou sémantikou už existuje a je vyžadované jeho prepoužitie
* dátový prvok nemá predpoklad používania vo viac ako 1 aplikačných službách pre viac ako 1 integračného partnera resp. viacerými ISVS a tým pádom je súčasťou 4 hviezdičkovej verzie (a je navrhnutý referencovateľný identifikátor) tj. lokálneho dátového modelu žiadateľa

**Proces**

1) Žiadateľ podá žiadosť o jednotný referencovateľný identifikátor prostredníctvom vytvorenia žiadosti o zaregistrovanie v Centrálnom metainformačnom systéme.

https://metais.vicepremier.gov.sk/uri/newurirequest/

2) Správca jednotných referencovateľných identifikátorov z centrálnej dátovej kancelárie žiadosť zachytí a začne komunikovať so žiadateľom o identifikátor v prípade rozporov so štandardami alebo v prípade duplicity identifikátora resp. iných nezrovnalostí.

3) Ak spĺňa identifikátor na žiadosti všetky náležitosti, posunie žiadosť na schválenie štandardizačnej skupine PS1, ktorá dostane na zasadnutie predspracované žiadosti o ktorých bude rozhodovať. Následne bude hlasovať o pridelení jednotného referencovateľného identifikátora a schváli/zamietne žiadosť (systém MetaIS). Zoznam schválených žiadostí je dostupný na adrese

https://metais.vicepremier.gov.sk/uri/list/accepted

Typ identifikátorov o ktoré je možné žiadať :

* Ontológia - Registrácia URI na ontológiu
* Ontológia - Registrácia URI pre verziu ontológie
* Dátový prvok - Registrácia URI pre dátový prvok
* Šablóna URI - Registrácia URI pre individuum

### Príklad

**Definície**

// URI šablóna pre typy právnych foriem

https://data.gov.sk/def/legal-form/[STAT-ID]

// URI šablóna pre typy vlastníctva

https://data.gov.sk/def/ownership-type/[STAT-ID]

// Akciová spoločnosť

https://data.gov.sk/def/legal-form/[STAT-ID]

**Indivídua**

// URI šablóna pre RPO subjekty

https://data.gov.sk/id/legal-subject/[ICO]

// URI šablóna pre obce SR

https://data.gov.sk/id/municipality/[REG-ADDRESS-ID]

// URI šablóna pre kategorizované lieky

https://data.gov.sk/id/registered-drug/[SUKL-ID]

// URI šablóna pre ISVS

https://data.gov.sk/id/egov/isvs/[METAIS-ID]

# Prílohy

## Príloha 1 – Výstupy aktivity 1.2 a 1.3 – zoznam dátových entít a číselníkov

Formát výstupu je XLS súbor. Pri tvorbe tohto výstupu boli uvažované požadované dostupné zdroje. Pre sprehľadnenie výstupu sú jednotlivé zdroje agregované do troch celkov, ktoré pokrývajú požadovanú agendu.

Rozdelenie je nasledovné :

1. **Objekty evidencie, identifikované ako kľúčové údaje, ktoré sa nachádzajú v CSRÚ alebo je plánované ich pripojenie pre účely integrácie a zdieľania**

* Spracovaný je aktuálny stav ako aj nadchádzajúce plánované integrácie poskytovateľov dát
* Pre jednotlivé objekty evidencie sú identifikované prislúchajúce dátové entity
* Definuje, či sa entita nachádza v Centrálnom modely údajov
* Na základe predpokladaného počtu konzumentov ako aj významu samotnej entity je definovaná miera prepoužiteľnosti dátovej entity čím sa zvyšuje jej potenciál na zaradenie do Centrálneho modelu údajov
* Zoznam zahrnuje analýzu jednotlivých objektov evidencie a dátových entít z pohľadu kandidátov na Moje dáta
* Samotné prepojenie je odstupňovanie podľa dôležitosti

1. **Dáta, z pohľadu verejnosti (dáta vyžadované verejnosťou a dáta s vysokým potenciálom na prepoužitie)**

* Všetky datasety/objekty evidencie sú analyzované z pohľadu potenciálu na zaradenie do CMÚ
* Pri jednotlivých záznamoch je definovaná ich súčasná prístupnosť verejnosti
* Pre všetky objekty je definovaná ich dôležitosť z pohľadu zaradenie dátovej entity do CMÚ
* Jednotlivé záznamy majú prepojenie na plán CSRÚ (bod 1- Objekty evidencie, identifikované ako kľúčové údaje)

1. **Kľúčové číselníky**

* Kompletná analýza číselníkov, ktoré sú prostredníctvom integrácií v CSRÚ používané
* Analýza číselníkov z ďalších, dnes využívaných zdrojov očistená o prípadné duplicity
* Pre vyhlásené základné číselníky sa v zozname nachádzajú aj schválené/navrhované jednotné referencovateľné identifikátori
* Pri číselníkoch je uvádzaný spôsob poskytovania číselníkov tretím stranám
* Zoznam obsahuje a kategorizuje odporúčania pre všetky číselníky, ktoré nie sú základné s definovaním ich potenciálu stať sa základným číselníkom

## Príloha 2 – Modely dátových entít

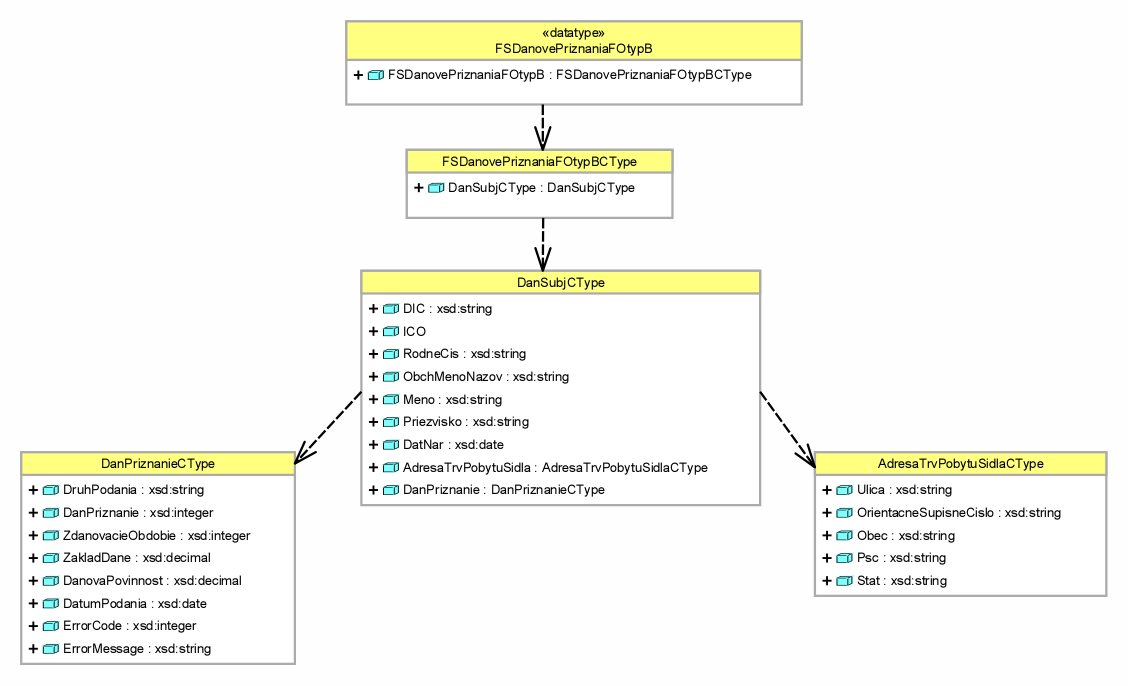
**Formát výstupu**

Výstup je prezentovaný prostredníctvom zdrojového súboru vo formáte BizzDesign Enterprise XMA, ktorý zabezpečuje prepoužiteľnosť výstupu a priebežné rozširovanie dátových modelov vzhľadom na ďalšie plánované integrácie.

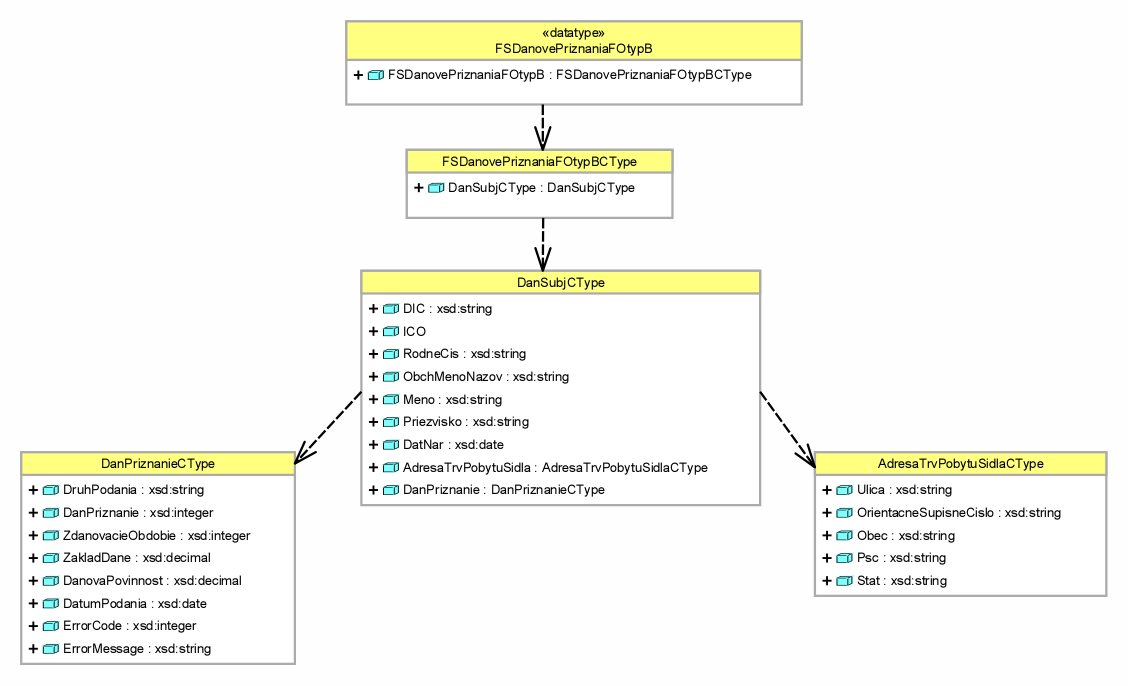
**Obsah**

Obsahuje modely objektov evidencie v CSRU, ktoré sú poskytované ako služba od rôznych poskytovateľov. Dátové modely nie sú len pre referenčné údaje ale aj pre iné objekty evidencie. I keď sú dátové prvky reprezentované prostredníctvom UML notácie, sú zapísané v BizzDesign Enterprise, čím je zabezpečená prepoužiteľnosť aj do budúcnosti(BizzDesign je nástroj používaný na modelovanie vrámci UPVII). V dokumente uvádzame aj samotné dátové modely, ktoré slúžia len pre účely prehľadu. Detailné informácie aj s popismi modelov sa nachádzajú v zdrojovom súbore BizzDesign.

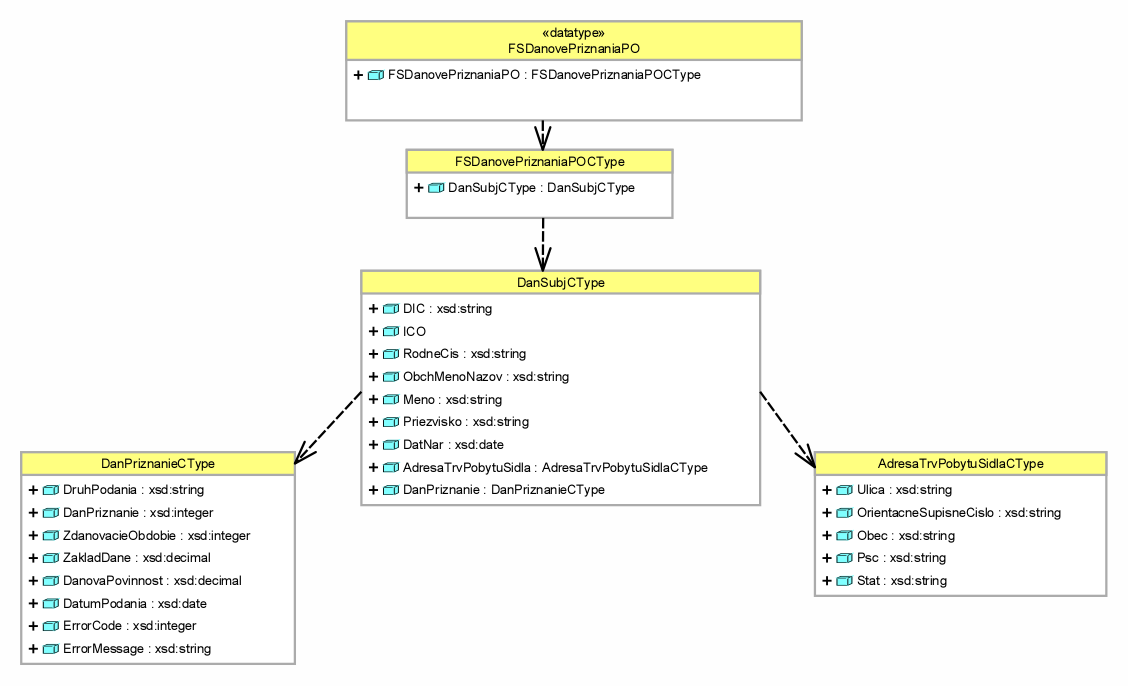
### Finančná správa



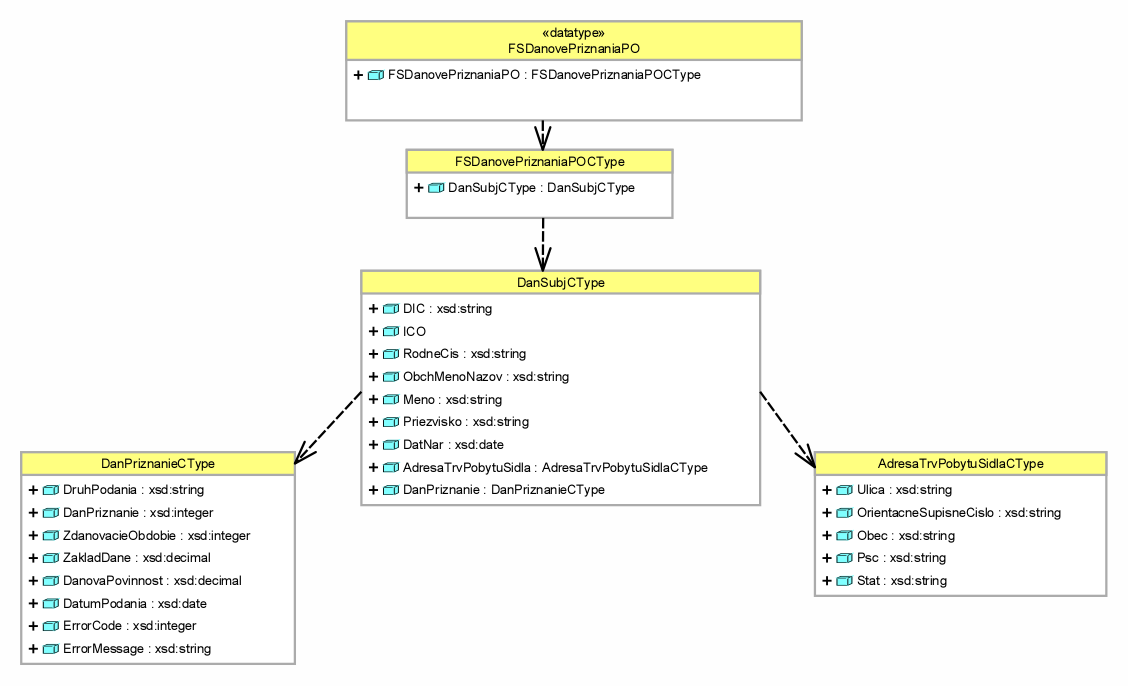
Obrázok 13: OE\_FSSR\_DanovePriznaniaFOB\_Asynch\_v003



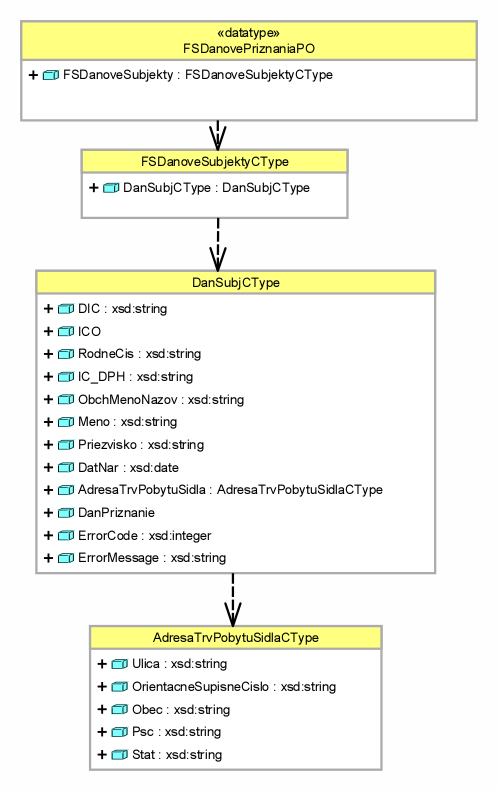
Obrázok 14: OE\_FSSR\_DanovePriznaniaFOB\_Synch\_v003



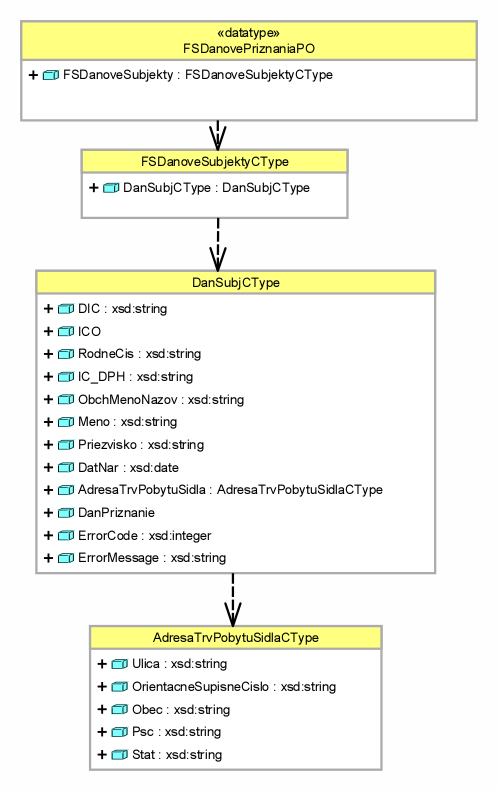
Obrázok 15: OE\_FSSR\_DanovePriznaniaPO\_Asynch\_v003



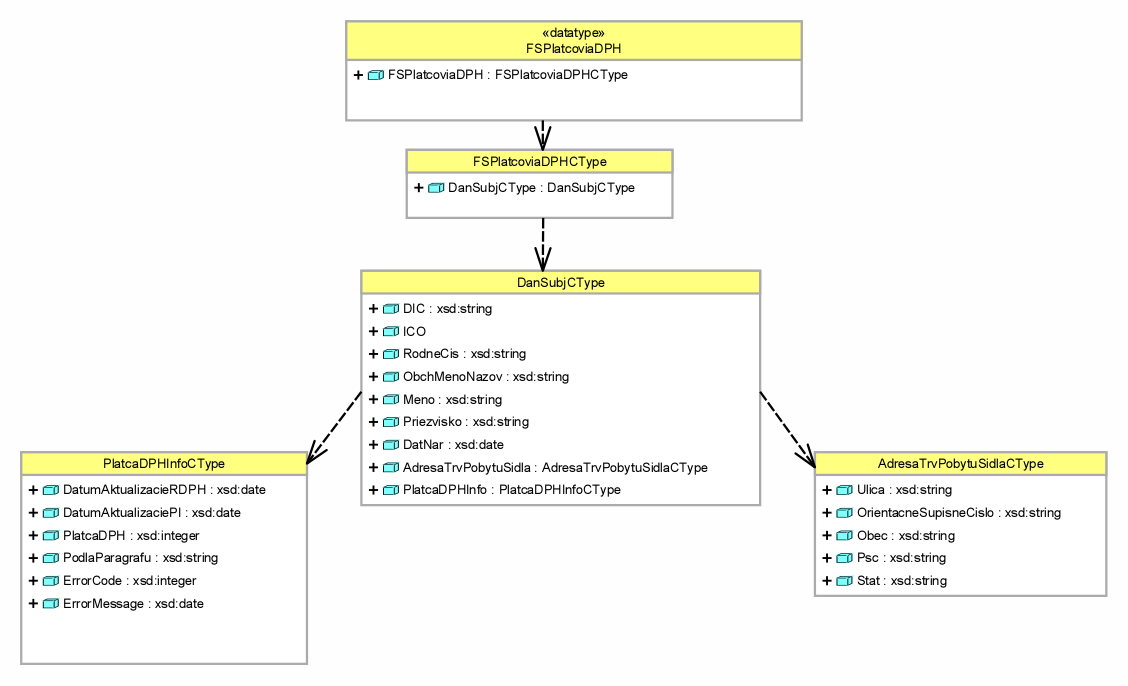
Obrázok 16: OE\_FSSR\_DanovePriznaniaPO\_Synch\_v003



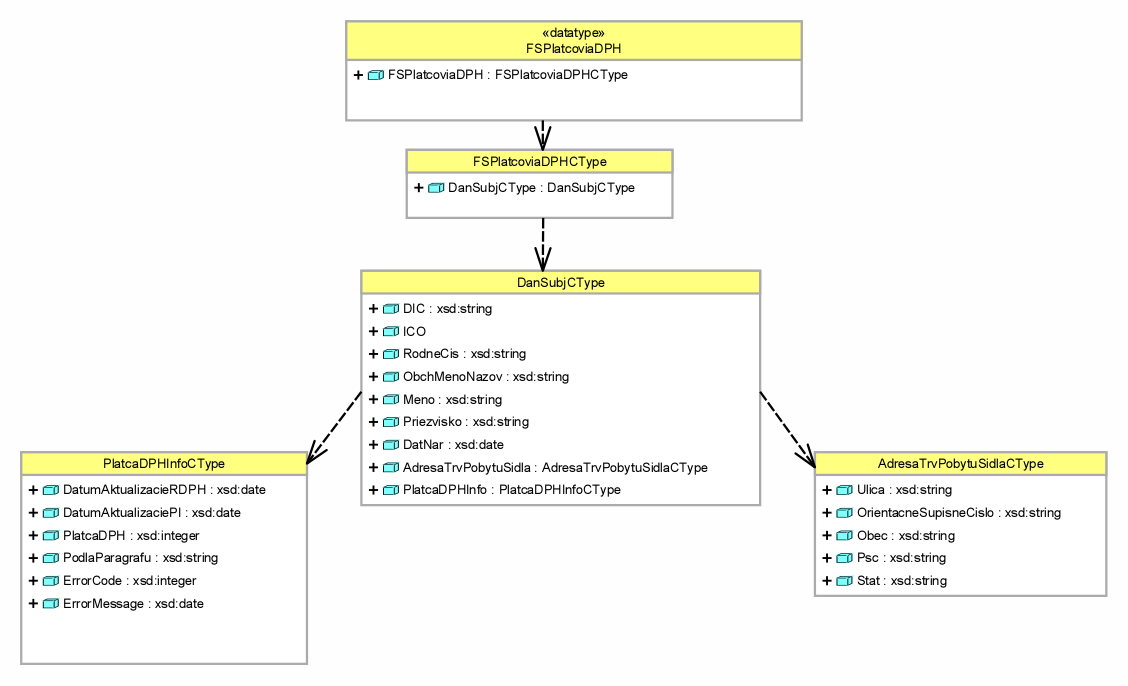
Obrázok 17: OE\_FSSR\_DanoveSubjekty\_Asynch\_v002



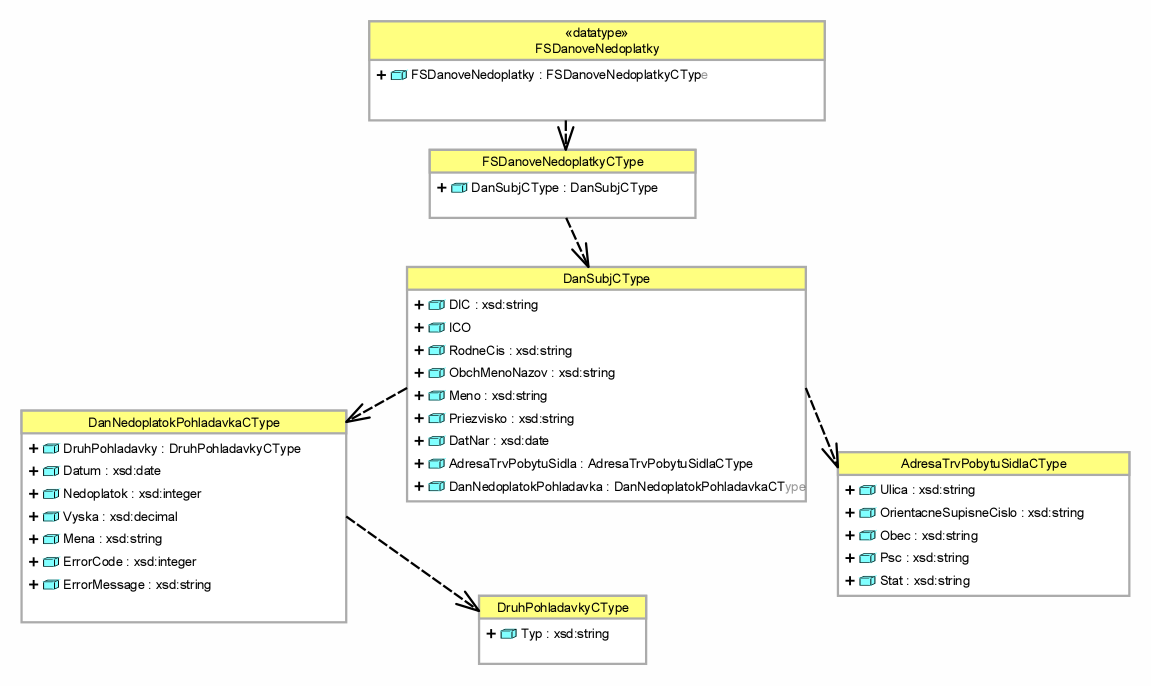
Obrázok 18: OE\_FSSR\_DanoveSubjekty\_Synch\_v002



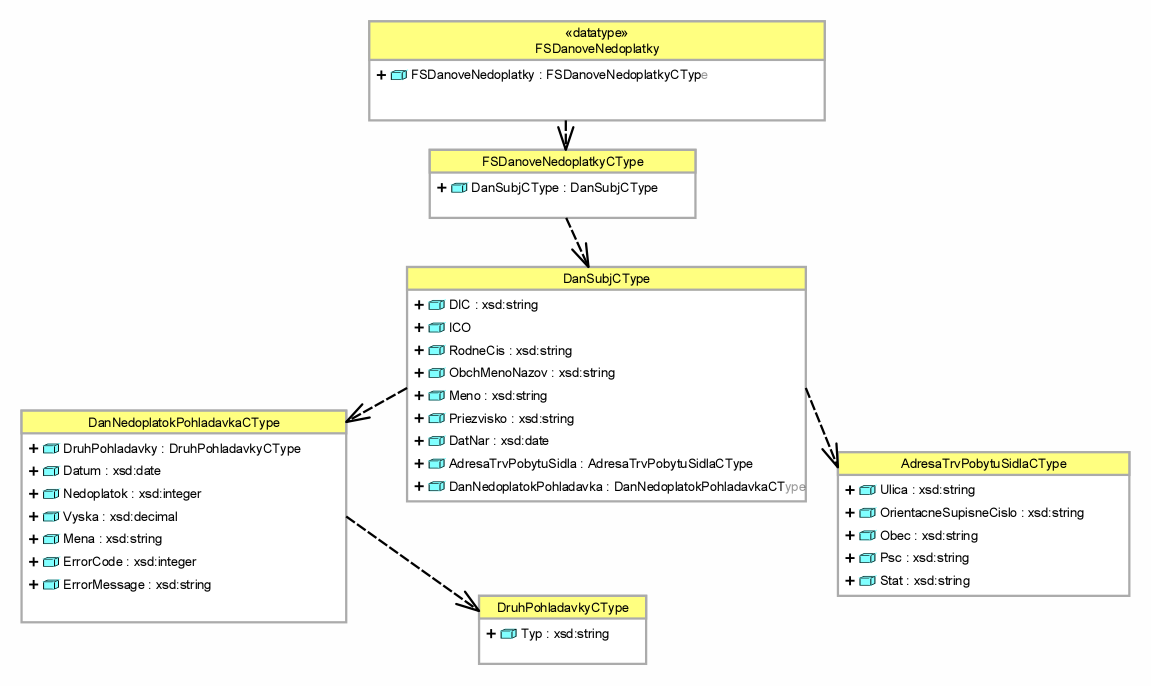
Obrázok 19: OE\_FSSR\_DanoveSubjektyRegisterDPH\_Asynch\_v003



Obrázok 20: OE\_FSSR\_DanoveSubjektyRegisterDPH\_Synch\_v003

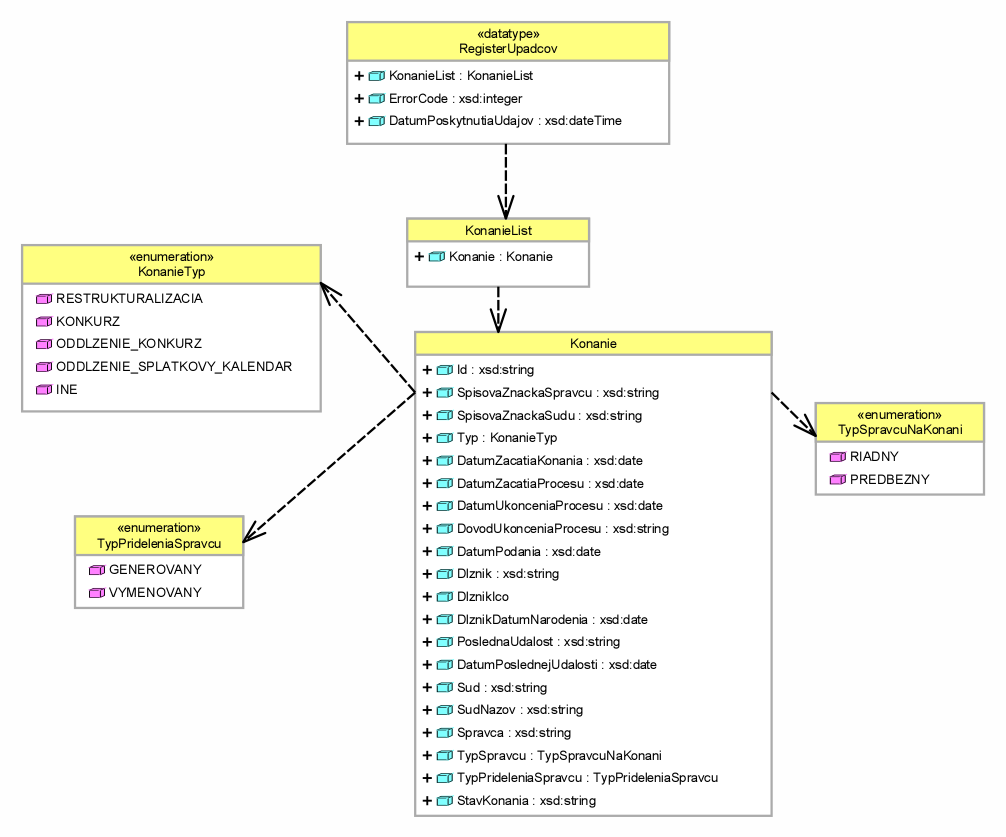


Obrázok 21: OE\_FSSR\_Nedoplatky\_Asynch\_v007



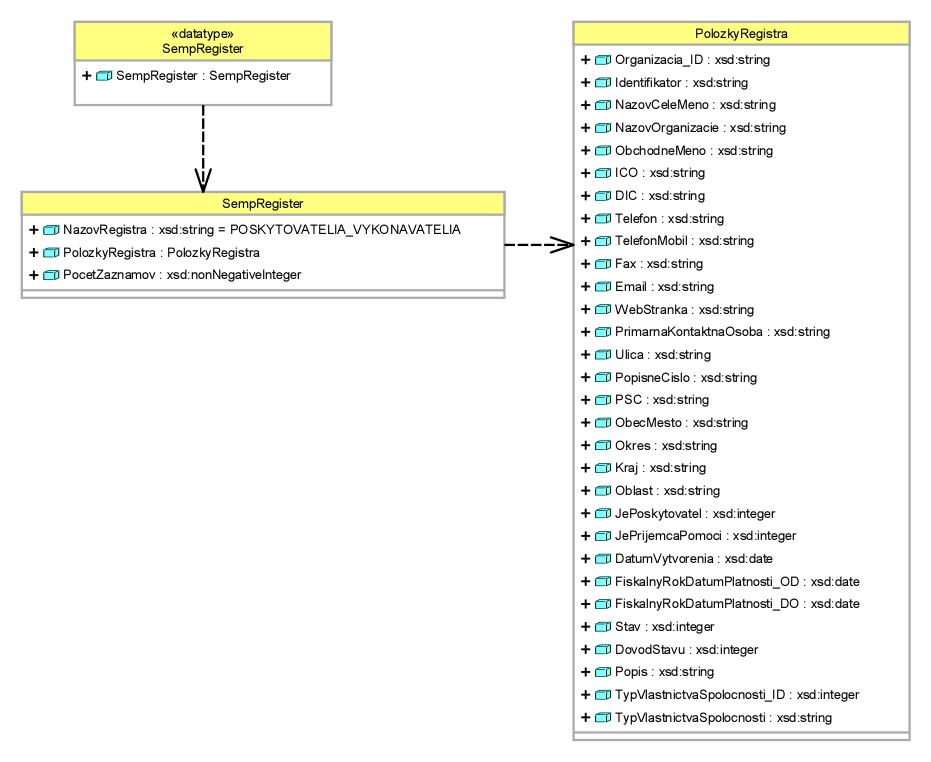
Obrázok 22: OE\_FSSR\_Nedoplatky\_Synch\_v007

### Register úpadcov



Obrázok 23: OE\_MinSprav\_RegisterUpadcov\_Sync\_v001

### Protimonopolný úrad



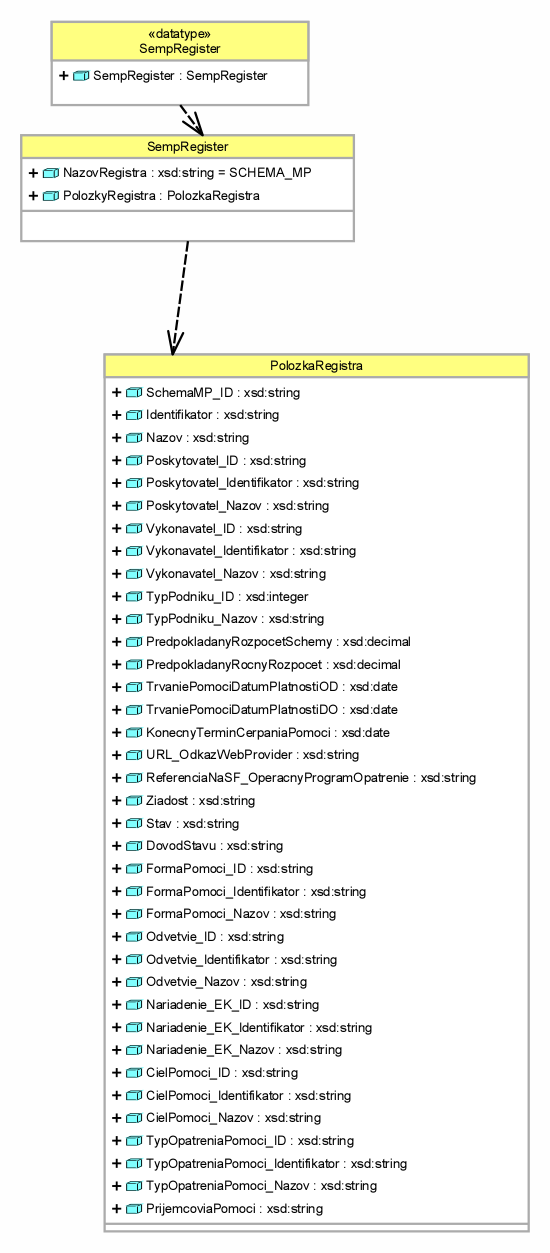
Obrázok 24: OE\_SEMP\_REGISTER\_POSKYTOVATELIA\_VYKONAVATELIA\_v001



Obrázok 25: OE\_SEMP\_REGISTER\_PRIJEMCOV\_POMOCI\_ASYNC\_v001



Obrázok 26: OE\_SEMP\_REGISTER\_PRIJEMCOV\_POMOCI\_SYNC\_v001



Obrázok 27: OE\_SEMP\_REGISTER\_SCHEMY\_MP\_v001



Obrázok 28: OE\_SEMP\_REGISTER\_SCHEMY\_SP\_v001

### Register adries



Obrázok 29: RA\_AddressWithAddressPointByBuildingIndex\_v20

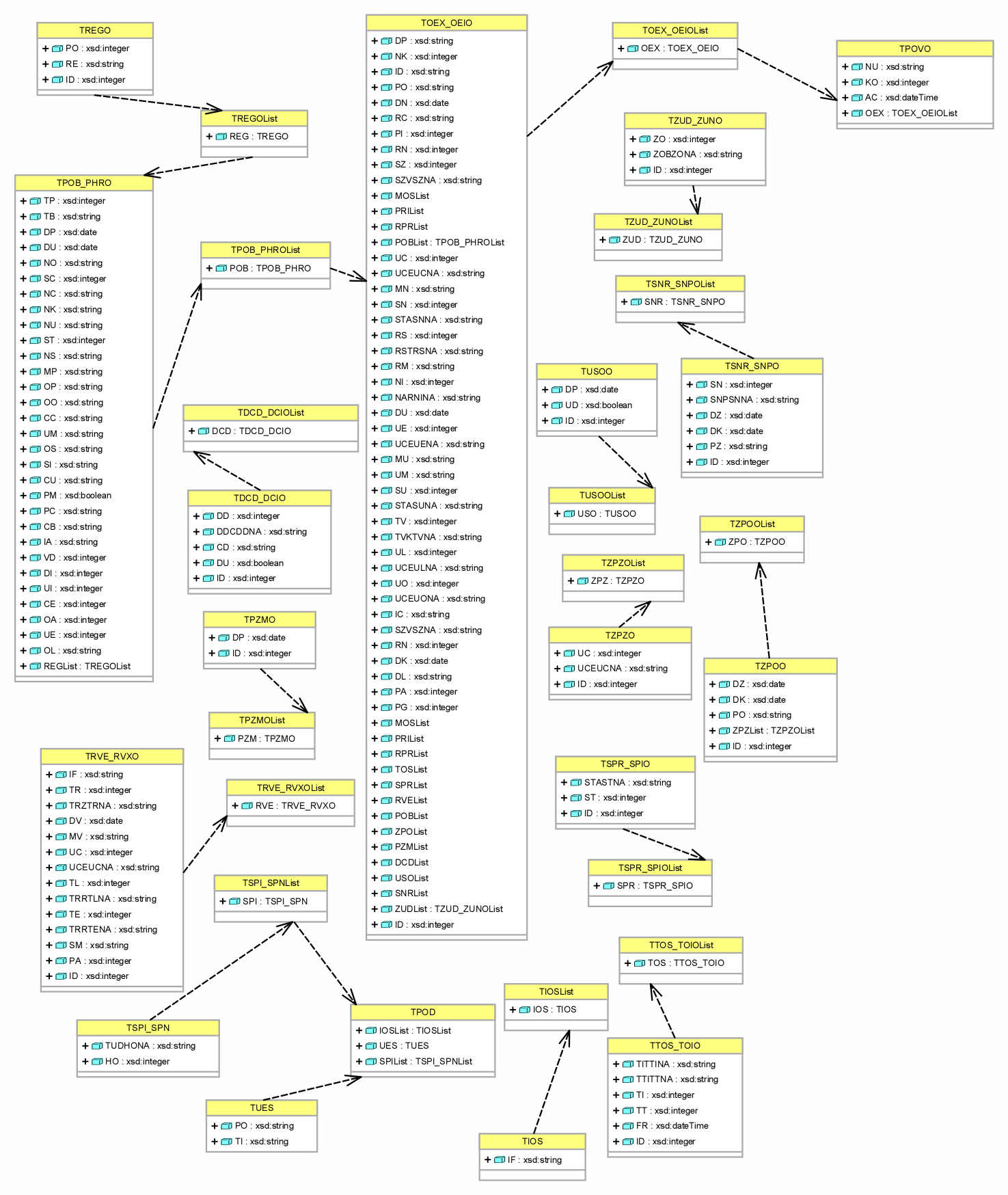


Obrázok 30: RA\_AddressesWithAddressPointsByAttributes\_v20

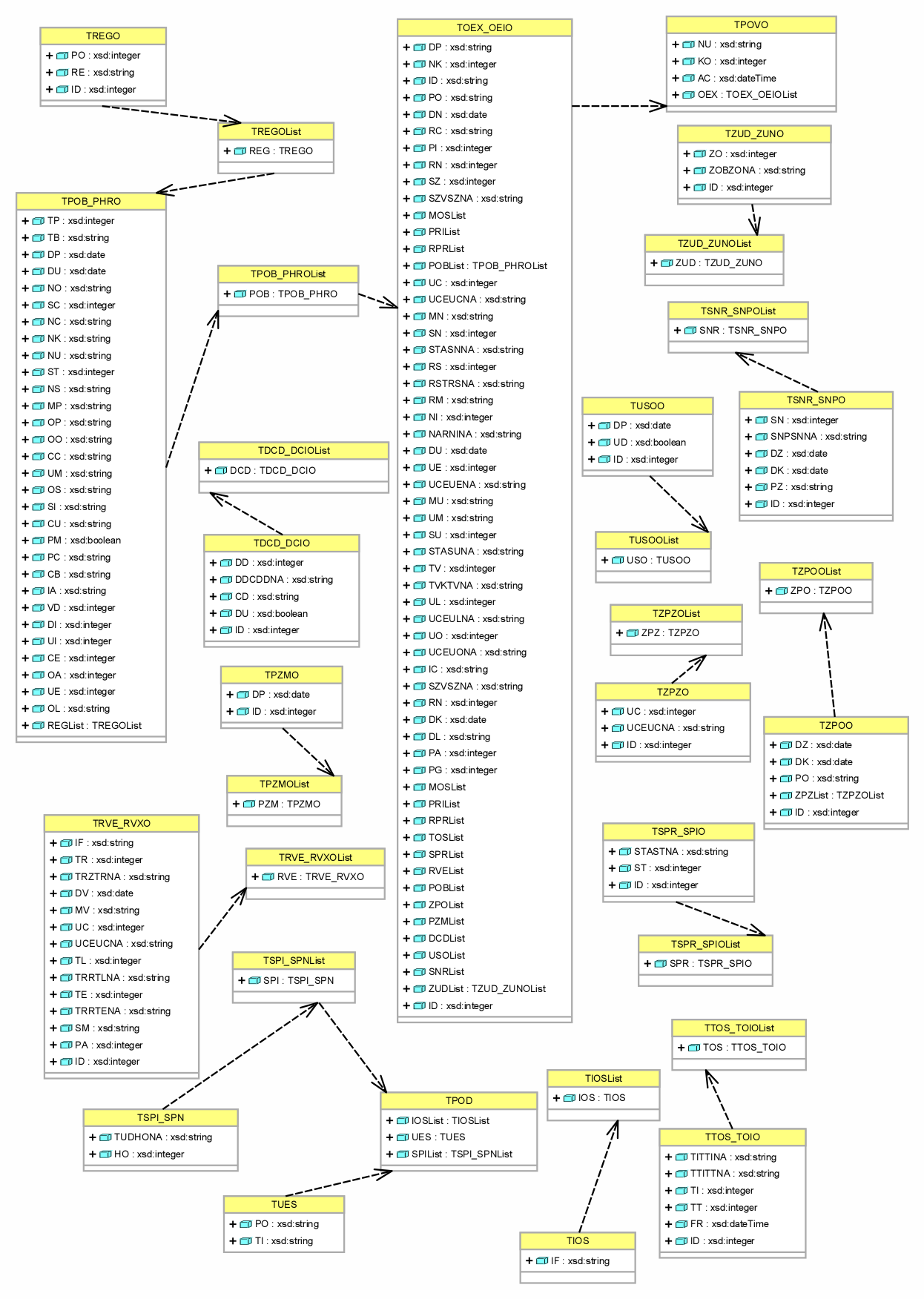


Obrázok 31: RA\_AddressesWithAddressPointsByBuildingIndexes\_v20

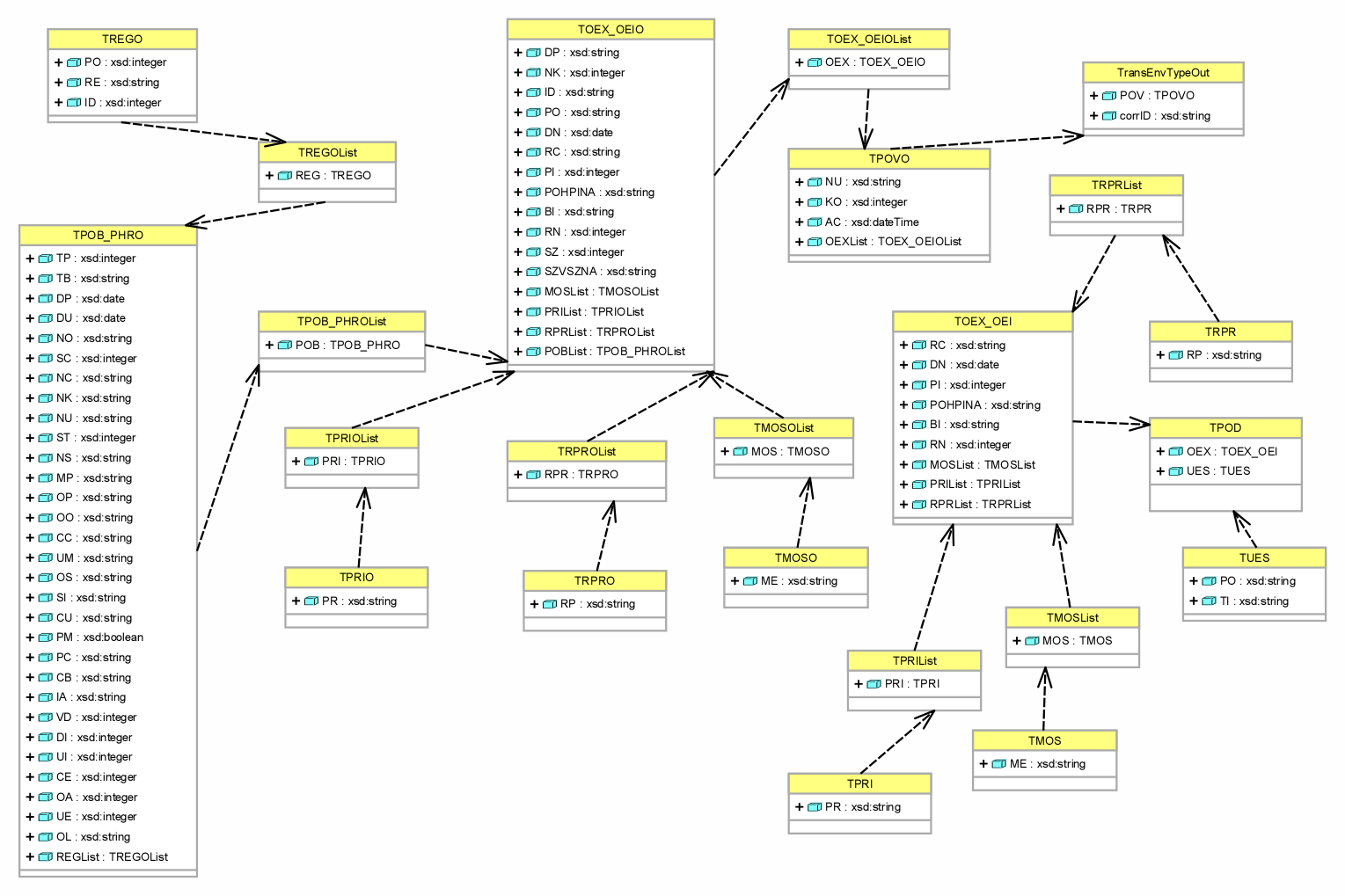
### Register fyzických osôb



Obrázok 32: PoskytnutieUdajovIFOOnlineWS-v1.0

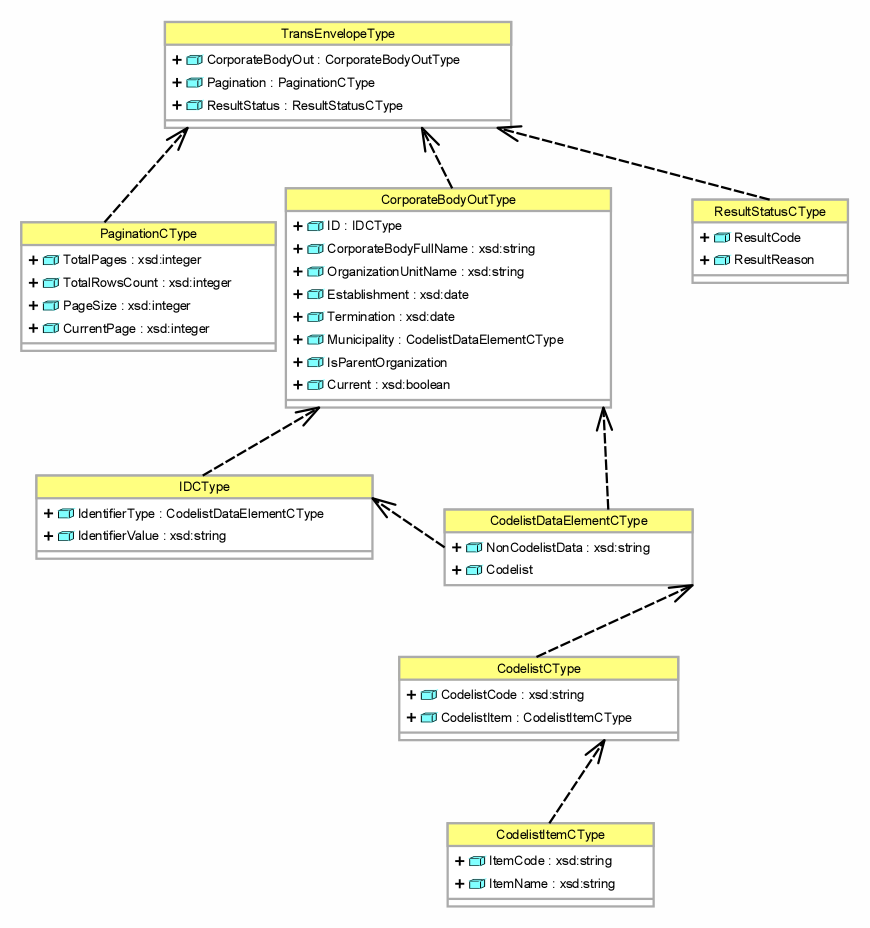


Obrázok 33: PoskytnutieUdajovIFOPodlaSpecifikacieWS-v1.0

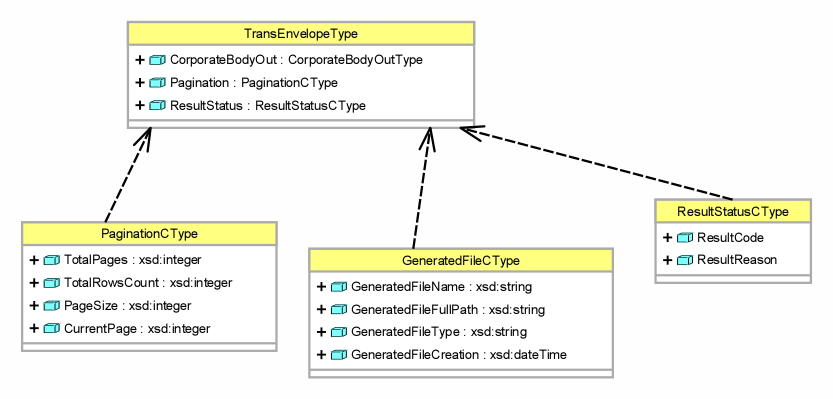


Obrázok 34: PoskytnutieZoznamuIFOPodlaVyhladavacichKriteriiWS-v1.0

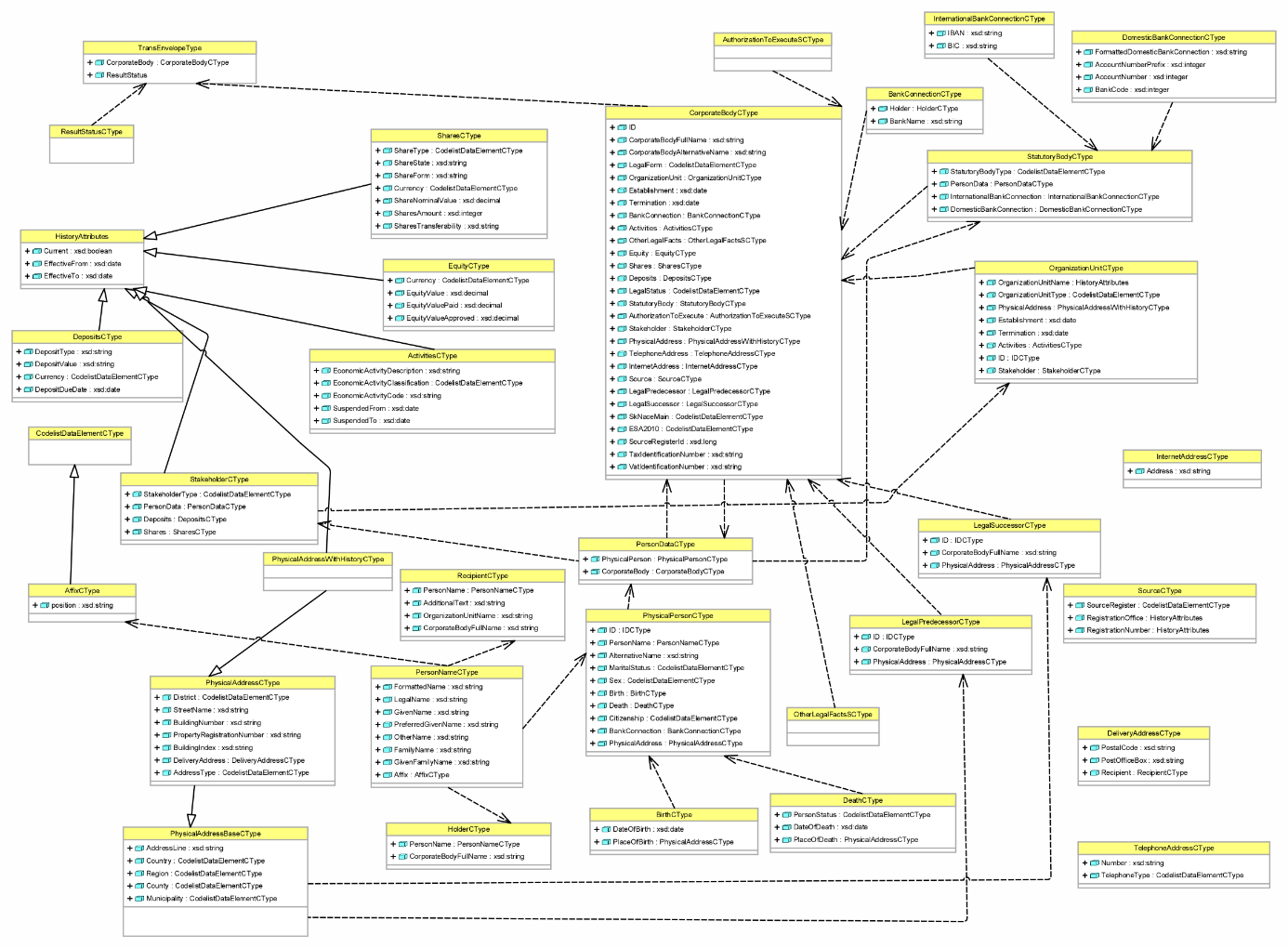
### Register právnických osôb



Obrázok 35: Poskytnutie\_IPO\_out-v2.4

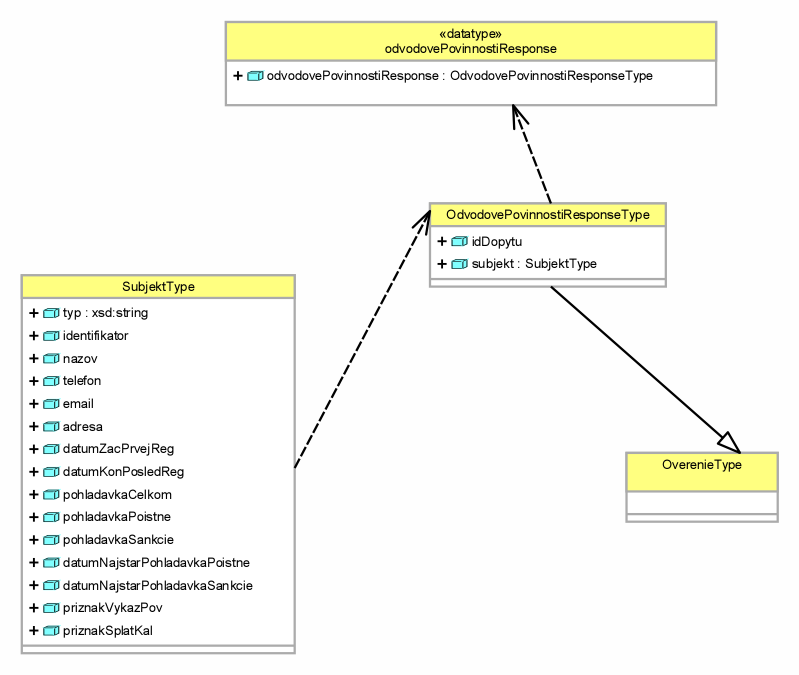


Obrázok 36: Poskytnutie\_suborov\_zmien\_PO\_out-v2.4



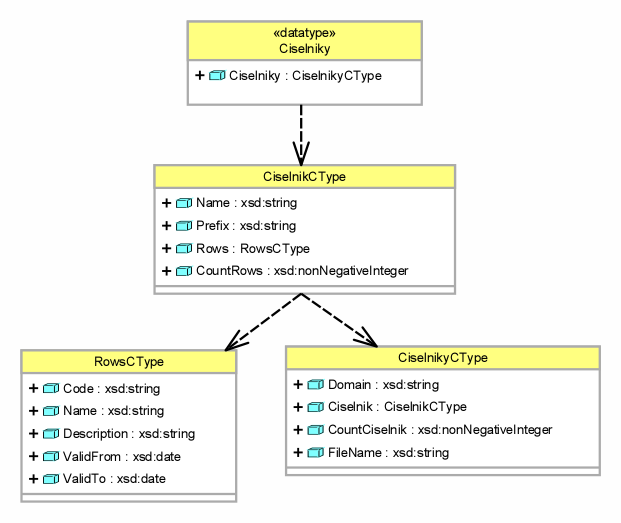
Obrázok 37: Poskytnutie\_udajov\_PO\_out-v2.4

### Sociálna poisťovňa

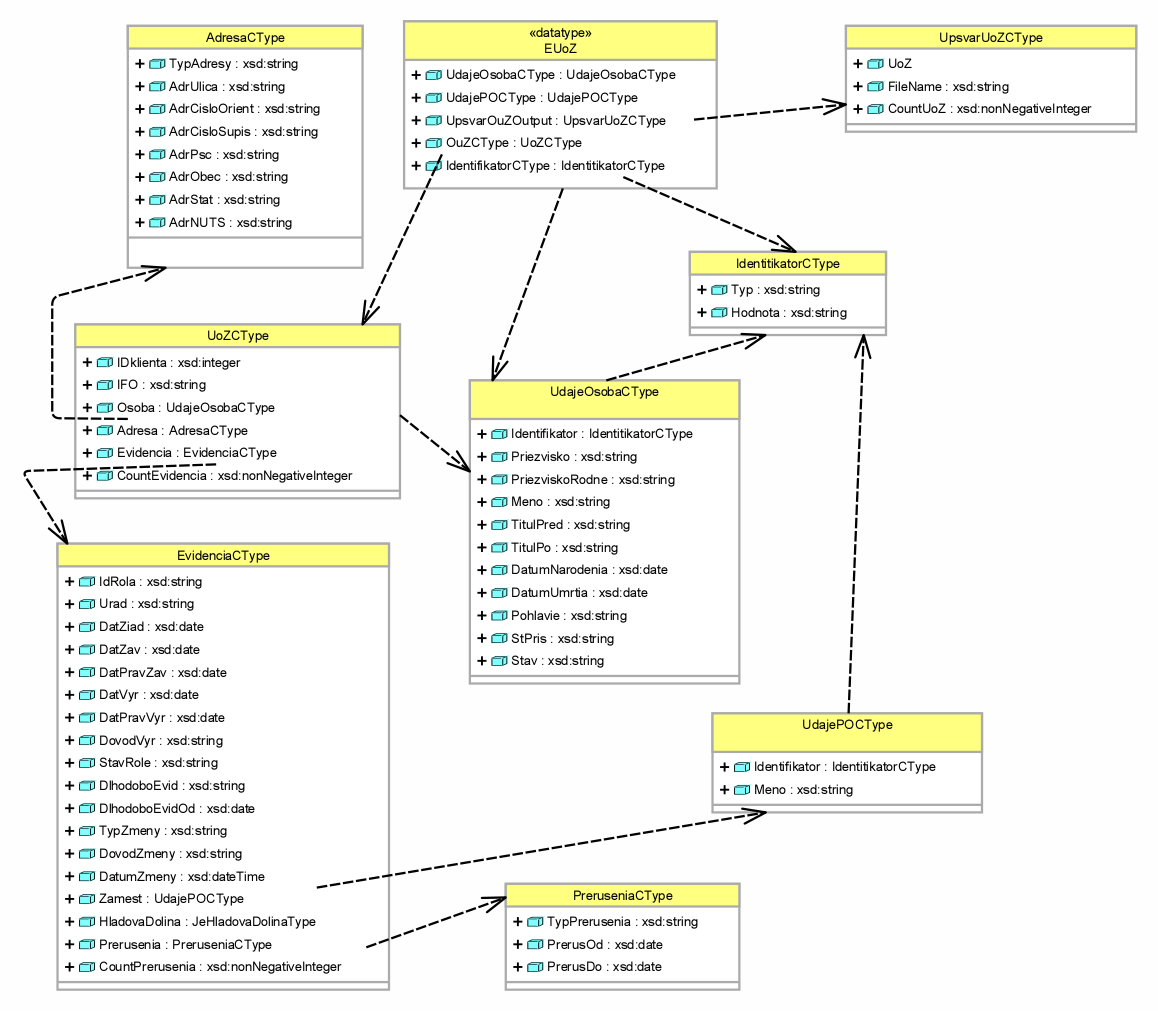


Obrázok 38: OE\_SocPoist\_OdvodovePovinnosti\_v002

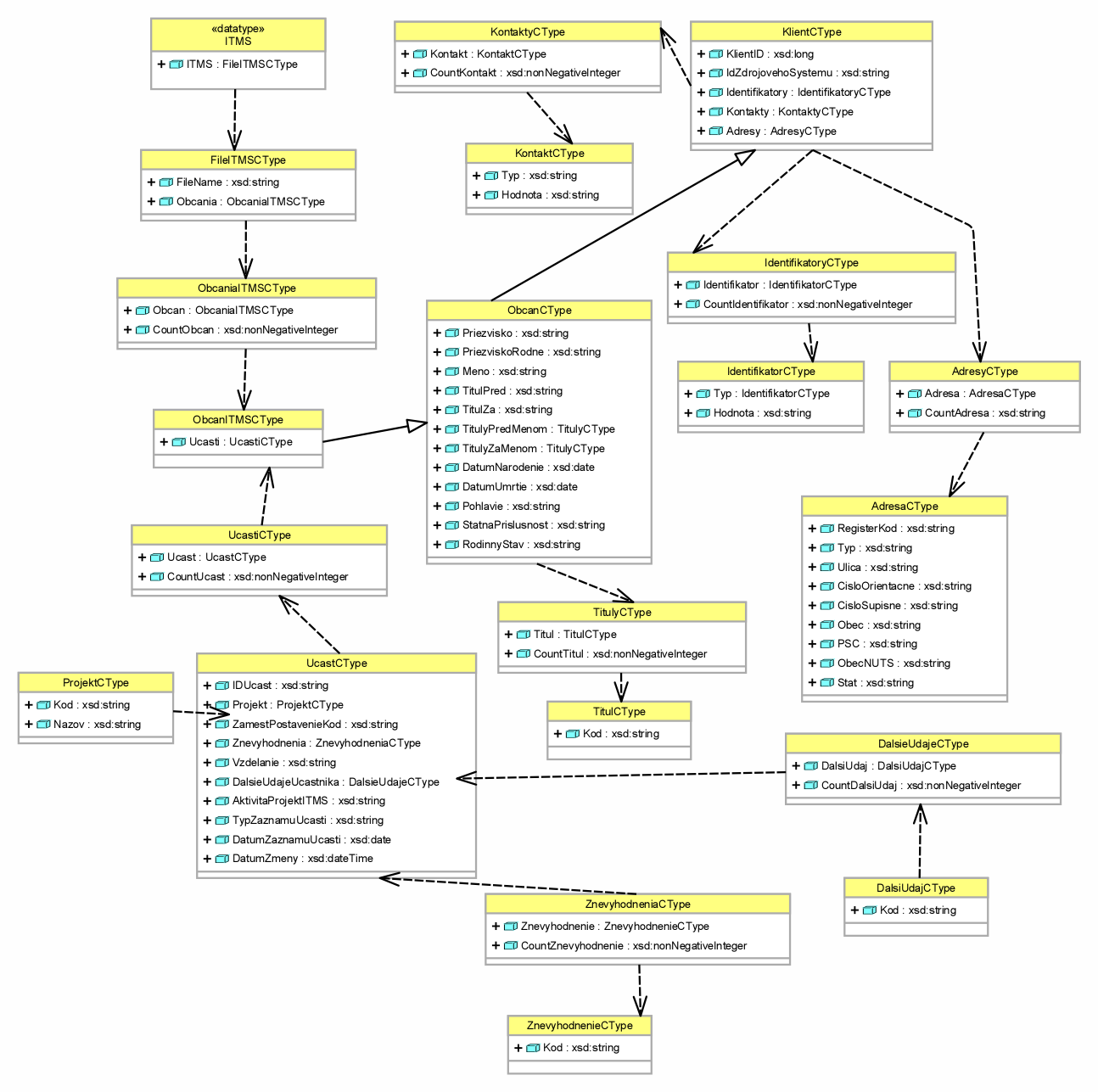
### UPSVaR



Obrázok 39: OE\_UPSVaR\_Ciselniky\_v001

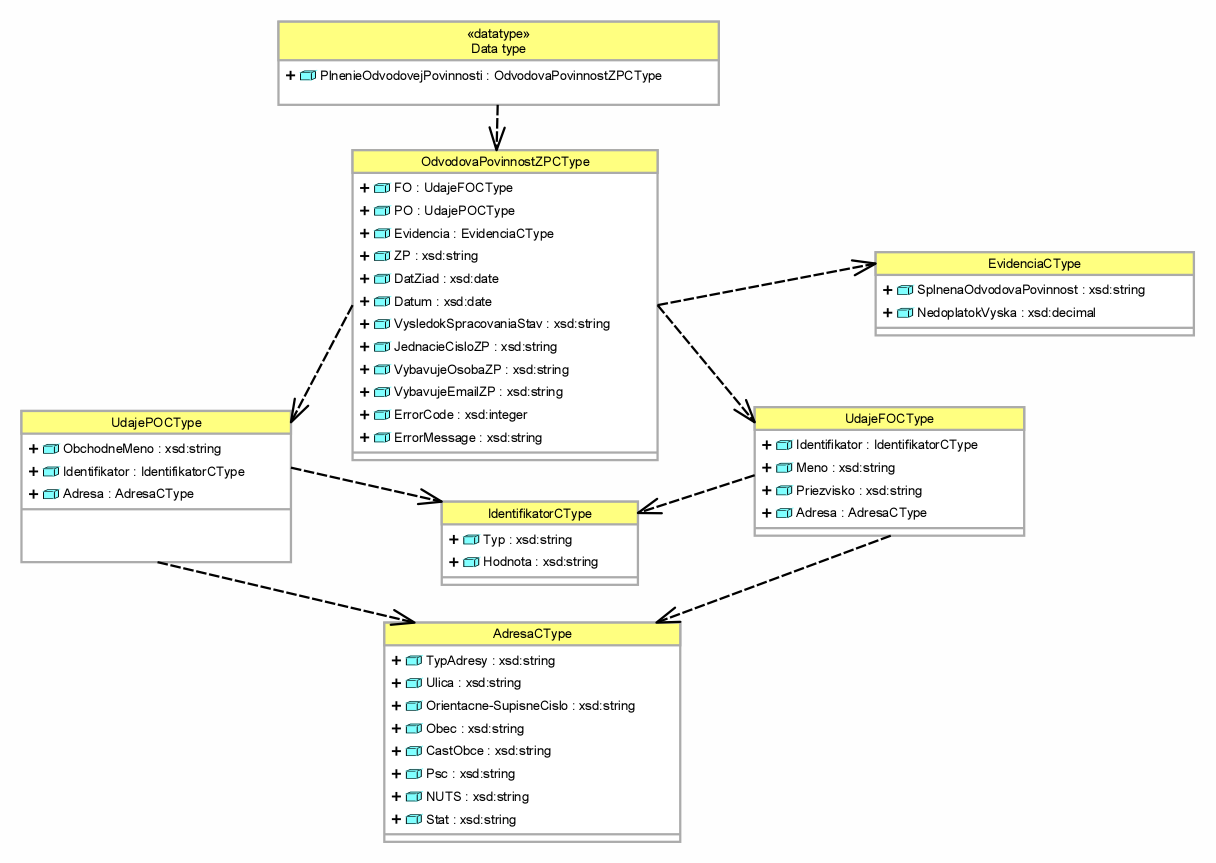


Obrázok 40: OE\_UPSVaR\_EUoZ\_v013

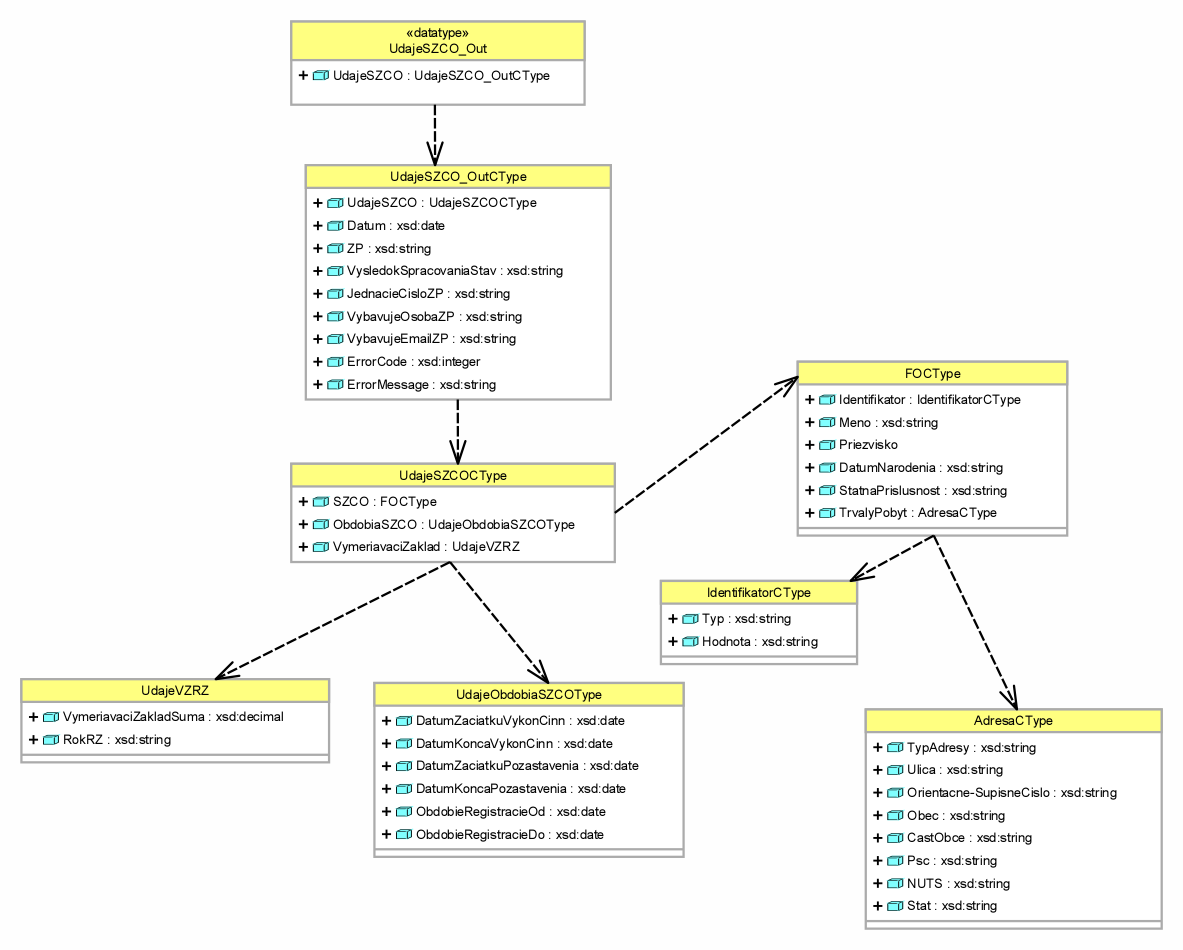


Obrázok 41: OE\_UPSVaR\_KartaUcastnika\_v012

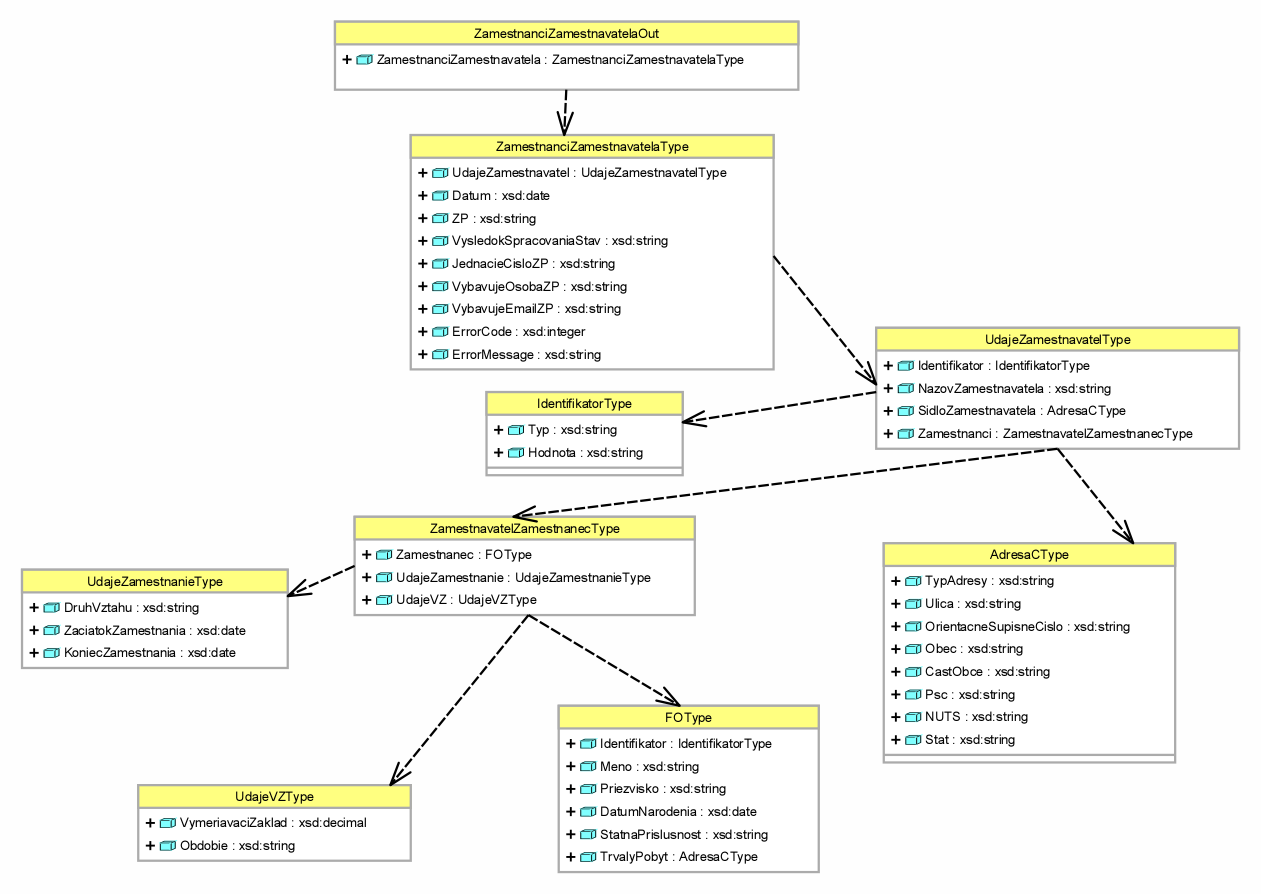
### Zdravotná poisťovňa



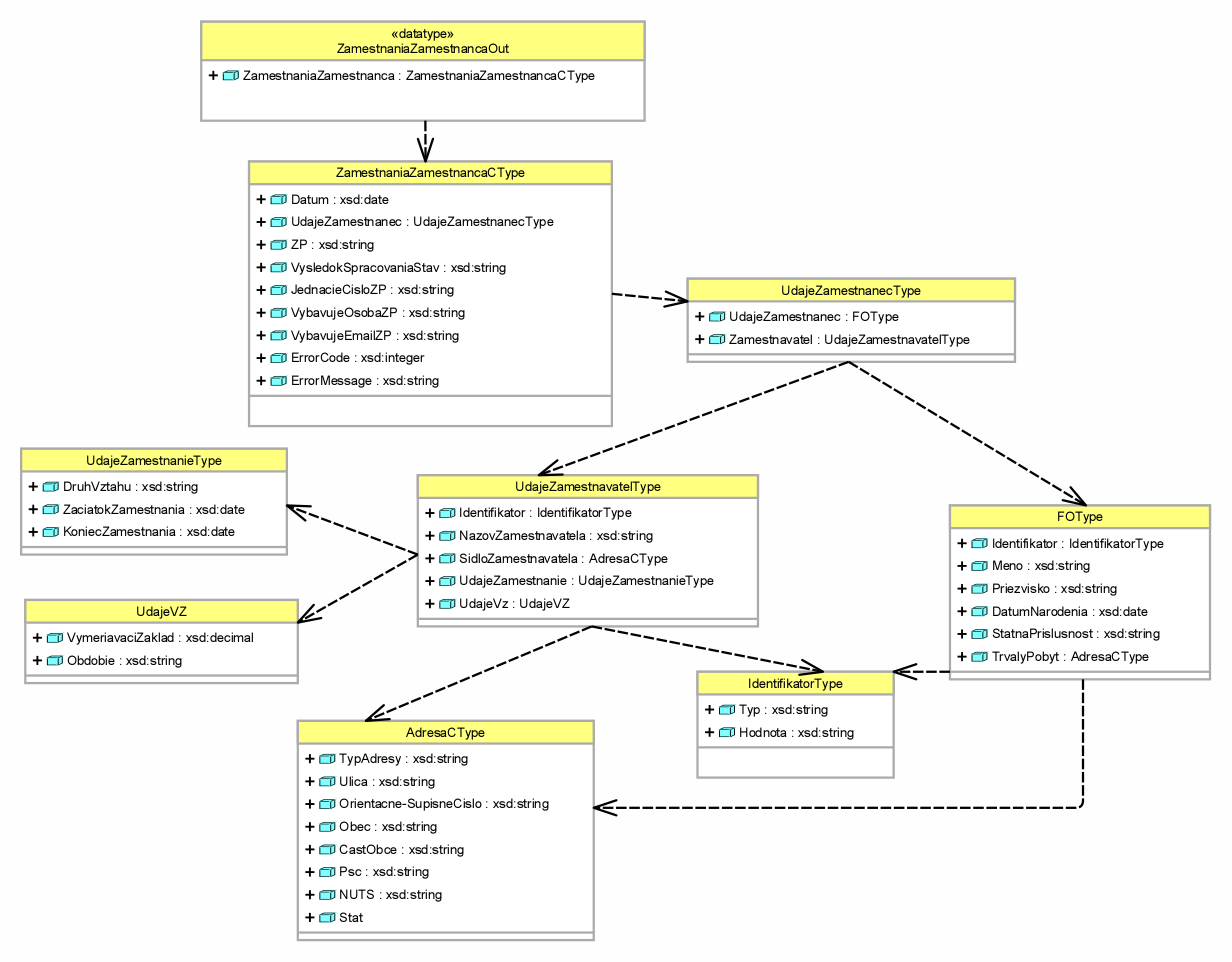
Obrázok 42: OE\_ZP\_Odvodova\_Povinnost\_v005



Obrázok 43: OE\_ZP\_UdajeSZCO\_v002



Obrázok 44: OE\_ZP\_ZamestnanciZamestnavatela\_v001



Obrázok 45: OE\_ZP\_ZamestnaniaZamestnanca\_v001

## Príloha 3 - Príklad serializácie dátového prvku

Serializácia dátového prvku Rodné priezvisko osoby:

### JSON-LD

Dátová serializácia do formátu JSON-LD podľa špecifikácie : <https://www.w3.org/TR/json-ld11/>

{

**"@context"**: {

**"adms"**: "http://www.w3.org/ns/adms#",

**"dcat"**: "http://www.w3.org/ns/dcat#",

**"dcterms"**: "http://purl.org/dc/terms/",

**"loca"**: "https://data.gov.sk/def/ontology/location/",

**"lsub"**: "https://data.gov.sk/def/ontology/legal-subject/",

**"owl"**: "http://www.w3.org/2002/07/owl#",

**"rdf"**: "http://www.w3.org/1999/02/22-rdf-syntax-ns#",

**"rdfs"**: "http://www.w3.org/2000/01/rdf-schema#",

**"skos"**: "http://www.w3.org/2004/02/skos/core#",

**"xsd"**: "http://www.w3.org/2001/XMLSchema#"

},

**"@id"**: "https://data.gov.sk/def/ontology/physical-person/familyName",

**"@type"**: "owl:DatatypeProperty",

**"rdfs:description"**: {

**"@language"**: "sk",

**"@value"**: "Obsahuje priezvisko osoby."

},

**"rdfs:domain"**: {

**"@id"**: "https://data.gov.sk/def/ontology/physical-person/PhysicalPerson"

},

**"rdfs:label"**: [

{

**"@language"**: "sk",

**"@value"**: "priezvisko"

}

],

**"rdfs:range"**: {

**"@id"**: "xsd:string"

},

**"rdfs:subPropertyOf"**: {

**"@id"**: "https://data.gov.sk/def/ontology/physical-person/name"

}

}

### RDF/XML

<?xml version="1.0" encoding="utf-8"?>

**<rdf:RDF**

xmlns:owl="http://www.w3.org/2002/07/owl#"

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

xmlns:adms="http://www.w3.org/ns/adms#"

xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

**>**

**<owl:DatatypeProperty** rdf:about="https://data.gov.sk/def/ontology/physical-person/familyName"**>**

**<rdfs:domain** rdf:resource="https://data.gov.sk/def/ontology/physical-person/PhysicalPerson"**/>**

**<rdfs:range** rdf:resource="http://www.w3.org/2001/XMLSchema#string"**/>**

**<rdfs:description** xml:lang="sk"**>**Obsahuje priezvisko osoby.**</rdfs:description>**

**<rdfs:label** xml:lang="sk"**>**priezvisko**</rdfs:label>**

**<rdfs:isDefinedBy** rdf:resource="https://data.gov.sk/def/ontology/physical-person/2018-03-29.owl"**/>**

**<rdfs:subPropertyOf** rdf:resource="https://data.gov.sk/def/ontology/physical-person/name"**/>**

**</owl:DatatypeProperty>**

**</rdf:RDF>**

### N-Triples

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/2000/01/rdf-schema#label> "priezvisko"@sk .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/2000/01/rdf-schema#range> <http://www.w3.org/2001/XMLSchema#string> .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/ns/adms#identifier> <https://data.gov.sk/id/identifier/kdp/D.1.1.1.6> .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/2000/01/rdf-schema#label> "family name"@en .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/2000/01/rdf-schema#isDefinedBy> <https://data.gov.sk/def/ontology/physical-person/2018-03-29.owl> .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#DatatypeProperty> .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/2000/01/rdf-schema#subPropertyOf> <https://data.gov.sk/def/ontology/physical-person/name> .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/2000/01/rdf-schema#label> "Nachname"@de .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/ns/adms#identifier> <https://data.gov.sk/id/identifier/metais-standard/255> .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/2000/01/rdf-schema#domain> <https://data.gov.sk/def/ontology/physical-person/PhysicalPerson> .

**<https://data.gov.sk/def/ontology/physical-person/familyName>** <http://www.w3.org/2000/01/rdf-schema#description> "Obsahuje priezvisko osoby."@sk .

## Príloha 4 – Metaúdaje otvorených údajov

### Katalóg

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Povinnosť použitia | Názov dátového prvku | Opis | Identifikátor dátového prvku | Referencia |
| Povinný | Katalóg | Katalóg (databáza) obsahujúca Datasety. | dcat:Catalog | <https://www.w3.org/TR/vocab-dcat/#class-catalog> |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Povinnosť použitia | Názov | Identifikátor dátového prvku | Dátový typ | Popis | Prípustný počet výskytov |
| Povinný | názov | dct:title | rdfs:Literal | názov katalógu s jazykovou príslušnosťou | 1..n |
| Povinný | opis | dct:description | rdfs:Literal | opis katalógu | 1..n |
| Povinný | vydavateľ | dct:publisher | foaf:Agent | entita zodpovedná za publikovanie katalógu | 1 |
| Povinný | dataset | dcat:dataset | dcat:Dataset | dataset patriaci do katalógu | 1..n |
| Nepovinný | domovská stránka | foaf:homepage | foaf:Document | domovská stránka pre katalóg | 0..1 |
| Nepovinný | jazyk | dct:language | dct:LinguisticSystem | predvolený jazyk používaný v katalógu | 0..n |
| Nepovinný | licencia | dct:license | dct:LicenseDocument | licencia pre katalóg | 0..1 |
| Nepovinný | dátum vydania | dct:issued | rdfs:Literal | dátum zverejnenia | 0..1 |
| Nepovinný | témy katalógu | dcat:themeTaxonomy | skos:ConceptScheme | použitie SKOS na klasifikáciu datasetov v katalógu | 0..n |
| Nepovinný | dátum modifikácie | dct:modified | rdfs:Literal | dátum poslednej aktualizácie katalógu | 0..1 |
| Nepovinný | má časť | dct:hasPart | dcat:Catalog | katalóg obsahuje vnorený katalóg | 0..n |
| Nepovinný | je súčasťou | dct:isPartOf | dcat:Catalog | katalóg je obsiahnutý v nadradenom katalógu | 0..1 |
| Nepovinný | práva | dct:rights | dct:RightsStatement | špecifikácia práv súvisiaca s daným katalógom | 0..1 |
| Nepovinný | geografická príslušnosť katalógu | dct:spatial | dct:Location | vlastnosť vyjadruje geografickú príslušnosť pre katalóg | 0..n |

### Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Povinnosť použitia | Názov dátového prvku | Popis | Identifikátor dátového prvku | Referencia |
| Povinný | Dataset | konceptuálna entita reprezentujúca publikovanú informáciu | dcat:Dataset | <https://www.w3.org/TR/vocab-dcat/#Class:_Dataset> |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Povinnosť použitia | Názov | Identifikátor dátového prvku | Dátový typ | Popis | Prípustný počet výskytov |
| Povinný | Názov | dct:title | rdfs:Literal | názov datasetu | 1..n |
| Povinný | Popis | dct:description | rdfs:Literal | popis datasetu | 1..n |
| Povinný | Vydavateľ | dct:publisher | foaf:Agent | entita zodpovedná za publikovanie datasetu | 1 |
| Povinný | Dátum vydania | dct:issued | rdfs:Literal | dátum zverejnenia | 1 |
| Povinný | Správca | dcatsk:maintaner | foaf:Agent | zodpovedný za kvalitu | 1 |
| Povinný | URL | dcat:landingPage | foaf:Document | URL na ktorej je Dataset sprístupnený so svojimi distribúciami | 1..n |
| Povinný | Verzia | owl:versionInfo | rdfs:Literal | informácia o verzii (číslo, alebo iná reprezentácia) | 1 |
| Povinný | Autor | dct:creator | foaf:Agent | tvorca daného datasetu | 1..n |
| Nepovinný | Kľúčové slová | dcat:keyword | rdfs:Literal | klúčové slová popisujúce dataset | 0..n |
| Nepovinný | Distribúcia datasetu | dcat:distribution | dcat:Distribution | relácia na distribúciu datasetu | 0..n |
| Nepovinný | Kontakt | dcat:contactPoint | vcard:Kind | kontaktné informácie pre posielanie komentárov k datasetu | 0..n |
| Povinný | Téma datasetu | dcat:theme | skos:Concept | súvisiace entity tvoriace tému datasetu | 1...n |
| Nepovinný | Prispievateľ | dct:contributor | foaf:Agent | prispievateľ k vytvoreniu datasetu | 0..n |
| Povinný | URI | dct:identifier | rdfs:Resource | URI datasetu | 1...n |
| Nepovinný | Periodicita aktualizácie | dct:accrualPeriodicity | dct:Frequency | periodicita aktualizácie datasetu, pričom sa musí uvádzať ak je známa | 0..1 |
| Nepovinný | Geografická príslušnosť datasetu | dct:spatial | dct:Location | vlastnosť vyjadruje geografickú príslušnosť regiónu pre dataset | 0..n |
| Nepovinný | Časová platnosť datasetu | dct:temporal | dct:PeriodOfTime | vlastnosť vyjadruje časovú platnosť datasetu, pričom sa musí uvádzať ak je známa | 0..n |
| Nepovinný | Práva | dct:accessRights | dct:RightsStatement | definovanie, či dataset je: opendata, obmedzený prístup alebo neprístupný | 0..1 |
| Nepovinný | Správnosť dát | dcatsk:validityType | dcatsk:ValidityType | údaj reprezentuje informáciu o kvalite obsahu údajov | 0..1 |
| Nepovinný | Zdrojový dataset | dct:source | dcat:Dataset | zdrojový dataset, ktorý tvorí podklad na dáta aktuálneho | 0..n |
| Nepovinný | Predchádzajúca verzia | dct:isVersionOf | dcat:Dataset | predošlá verzia datasetu | 0..n |
| Nepovinný | Poznámky k verzii | adms:versionNotes | rdfs:Literal | popis rozdielnosti verzie oproti predchádcajúcim verziám | 0..n |
| Nepovinný | Dokumentácia | foaf:page | foaf:Document | dokumentácia datasetu | 0..n |
| Nepovinný | Nasledujúca verzia | dct:hasVersion | dcat:Dataset | nasledovná verzia datasetu | 0..n |
| Nepovinný | Jazyk | dct:language | dct:LinguisticSystem | jazyk datasetu | 0..n |
| Nepovinný | Iný identifikátor | adms:identifier | adms:Identifier | Ďalší (externý) identifikátor datasetu | 0..n |
| Nepovinný | Informácia o pôvode datasetu | dct:provenance | dct:ProvenanceStatement | informácia o pôvode datasetu | 0..n |
| Nepovinný | Súvisiaci zdroj | dct:relation | rdfs:Resource | URI súvisiaceho zdroja | 0..n |
| Nepovinný | Vzorka údajov | adms:sample | dcat:Distribution | Vzorka dát z datasetu | 0..n |
| Nepovinný | Dátum modifikácie | dct:modified | rdfs:Literal | dátum poslednej aktualizácie datasetu | 0..1 |
| Povinný | Typ datasetu | dct:type | skos:Concept | Dataset je buď štandardný, alebo referenčný | 1 |

### Distribúcia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Povinnosť použitia | Názov | Popis | Identifikátor dátového prvku | Referencia |
| Nepovinný | Distribúcia | fyzická podoba datasetu v nejakom formáte | dcat:Distribution | <https://www.w3.org/TR/vocab-dcat/#class-distribution> |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Povinnosť použitia | Názov | Identifikátor dátového prvku | Dátový typ | Popis | Prípustný počet výskytov |
| Povinný | Adresa distribúcie | dcat:accessURL | foaf:Document | URL na ktorej je možné získať distribúciu | 1..n |
| Povinný | Licencia | dct:license | dct:LicenseDocument | licencia pre distribúciu | 1 |
| Povinný | Stav | adms:status | skos:Concept | vlastnosť vyjadruje možné stavy: kompletné, tvoriaci | 1 |
| Povinný | Formát | dct:format | dct:MediaTypeOrExtent | táto vlastnosť definuje súborový formát distribúcie | 0..1 |
| Povinný | Opis | dct:description | rdfs:Literal | popis distribúcie | 1..n |
| Nepovinný | URI | dct:identifier | rdfs:Resource | URI distribúcie | 0..1 |
| Nepovinný | Typ média (mimetype) | dcat:mediaType | dct:MediaTypeOrExtent | oficiálny IANA číselník | 0..1 |
| Nepovinný | Schéma | dct:conformsTo | dct:Standard | schéma pre distribúciu, odporúča sa uvádzať vždy, ak sa používajú údaje vyžadujúce štruktúrovaný popis | 0..n |
| Nepovinný | Veľkosť súboru (byte) | dcat:byteSize | rdfs:Literal | Veľkosť súboru v byte | 0..1 |
| Nepovinný | Digitálny odtlačok | spdx:checksum | spdx:Checksum | kontrolný súčet pre zabezpečenie nezmeniteľnosti údajov | 0..1 |
| Nepovinný | Dokumentácia | foaf:page | foaf:Document | dokumentácia distribúcie | 0..n |
| Nepovinný | Priamy odkaz na stiahnutie | dcat:downloadURL | rdfs:Resource | priamy odkaz na stiahnutie distribúcie | 0..n |
| Nepovinný | Dátum formálnej platnosti | dct:issued | rdfs:Literal | dátum formálneho vydania distribúcie | 0..1 |
| Nepovinný | Práva | dct:rights | dct:RightsStatement | špecifikácia práv súvisiaca s danou distribúciou | 0..1 |
| Nepovinný | Dátum modifikácie | dct:modified | rdfs:Literal | dátum poslednej aktualizácie distribúcie | 0..1 |

1. https://wiki.finance.gov.sk/pages/viewpage.action?pageId=21169133 [↑](#footnote-ref-1)
2. <https://sp.vicepremier.gov.sk/lepsie-data/_layouts/15/WopiFrame.aspx?sourcedoc=/lepsie-data/Zdielane%20dokumenty/2018/Metodiky/Modelovanie%20udajov/MVSR-Metodika-Modelovanie%20udajov.docx&action=default&Source=%2Flepsie%2Ddata%2F%5Flayouts%2F15%2Fmobile%2Fviewa%2Easpx%3FList%3Dfbbb64ab%2D2cea%2D4936%2Daa4d%2D9748b8d14cde%26View%3D8da26f6f%2D36a4%2D41ef%2D93a4%2D038cda92a926%26RootFolder%3D%252Flepsie%2Ddata%252FZdielane%2Bdokumenty%252F2018%252FMetodiky%252FModelovanie%2Budajov%26ViewMode%3DDetail%26wdFCCState%3D1> [↑](#footnote-ref-2)
3. https://wiki.finance.gov.sk/pages/viewpage.action?pageId=14712976 [↑](#footnote-ref-3)
4. https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2014/55/ [↑](#footnote-ref-4)
5. <https://sp.vicepremier.gov.sk/lepsie-data/_layouts/15/WopiFrame.aspx?sourcedoc=/lepsie-data/Zdielane%20dokumenty/2018/Metodiky/Modelovanie%20udajov/MVSR-Metodika-Modelovanie%20udajov.docx&action=default&Source=%2Flepsie%2Ddata%2F%5Flayouts%2F15%2Fmobile%2Fviewa%2Easpx%3FList%3Dfbbb64ab%2D2cea%2D4936%2Daa4d%2D9748b8d14cde%26View%3D8da26f6f%2D36a4%2D41ef%2D93a4%2D038cda92a926%26RootFolder%3D%252Flepsie%2Ddata%252FZdielane%2Bdokumenty%252F2018%252FMetodiky%252FModelovanie%2Budajov%26ViewMode%3DDetail%26wdFCCState%3D1> [↑](#footnote-ref-5)
6. https://www.w3.org/TR/sparql11-query/ [↑](#footnote-ref-6)
7. https://www.w3.org/community/sparql-12/ [↑](#footnote-ref-7)
8. https://www.w3.org/TR/shacl/ [↑](#footnote-ref-8)
9. http://www.informatizacia.sk/ext\_dok-sp\_otvorene\_udaje\_schvalena/26035c [↑](#footnote-ref-9)
10. https://wiki.vicepremier.gov.sk/pages/viewpage.action?pageId=24945729 [↑](#footnote-ref-10)
11. https://wiki.vicepremier.gov.sk/pages/viewpage.action?pageId=24945847 [↑](#footnote-ref-11)
12. https://wiki.vicepremier.gov.sk/pages/viewpage.action?pageId=24945835 [↑](#footnote-ref-12)
13. https://wiki.vicepremier.gov.sk/pages/viewpage.action?pageId=24945841 [↑](#footnote-ref-13)
14. https://wiki.vicepremier.gov.sk/pages/viewpage.action?pageId=24945827 [↑](#footnote-ref-14)
15. http://informatizacia.sk/ext\_dok-uppvii\_sp\_manazment\_udajov\_vfinal/24196c [↑](#footnote-ref-15)