Referenčné registre

Obsah

[2 Úvod 2](#_Toc505327374)

[3 Základné pojmy, Aktuálny stav 2](#_Toc505327375)

[3.1 Čo je referenčný údaj? 2](#_Toc505327376)

[3.2 Čo je referencovateľný identifikátor? 3](#_Toc505327377)

[3.3 Bezpečnosť, ochrana osobných údajov, GDPR 4](#_Toc505327378)

[3.4 Aplikačný pohľad na referenčné údaje a ich integráciu 4](#_Toc505327379)

[3.4.1 Má ISVS obsahovať kópiu referenčných údajov? 5](#_Toc505327380)

[3.4.2 Aké sú výhody centrálneho modulu integrácie údajov? 6](#_Toc505327381)

[4 Kvalita údajov 6](#_Toc505327382)

[4.1 Atribúty a kritériá kvality 6](#_Toc505327383)

[4.1.1 Referenčná integrita 6](#_Toc505327384)

[4.1.2 Strojová spracovateľnosť 7](#_Toc505327385)

[4.1.3 Unikátnosť 7](#_Toc505327386)

[4.1.4 Aktuálnosť 7](#_Toc505327387)

[4.1.5 Kompletnosť, Konzistentnosť, Správnosť 7](#_Toc505327388)

[4.1.6 Presnosť 8](#_Toc505327389)

[4.2 Musia mať údaje 100% kvalitu? 8](#_Toc505327390)

[4.2.1 Rozdeľuj a panuj 8](#_Toc505327391)

[4.2.2 Aká má byť kvalita údajov? 9](#_Toc505327392)

[5 Dátová politika 9](#_Toc505327393)

[5.1 Princípy dátovej kvality 9](#_Toc505327394)

[5.2 Príklad kritérií kvality 9](#_Toc505327395)

[5.3 Ako zvyšovať kvalitu údajov? 10](#_Toc505327396)

[6 Povinnosti 11](#_Toc505327397)

[6.1 Čo musí dodržať každý ISVS? 11](#_Toc505327398)

[6.2 Čo musí dodržať referenčný register? 11](#_Toc505327399)

[6.3 Ako postupovať pri vyhlásení referenčného registra? 11](#_Toc505327400)

[7 Governance z pohľadu údajov, integrácie a kvality údajov 12](#_Toc505327401)

[8 Kto to bude platiť? 12](#_Toc505327402)

[9 Definície pojmov a skratky 12](#_Toc505327403)

[10 Relevantné dokumenty a zdroje 12](#_Toc505327404)

# Úvod

Problem statement: Zdieľanie referenčných údajov je dôležité pre zvyšovanie kvality služieb (napr. pre princíp 1x a dosť), zvyšovanie efektivity procesov VS. Vyhlasovanie referenčných údajov ide pomaly, asi kvôli kvalite dát a malému tlaku

Nehovoríme o otvorených a analytických údajoch, aj keď je to blízko a veľa sa dá použiť

Príklady použité v tomto dokumente sú ilustračné, nie sú presné, záväzné.

# Základné pojmy, Aktuálny stav

## Čo je referenčný údaj?

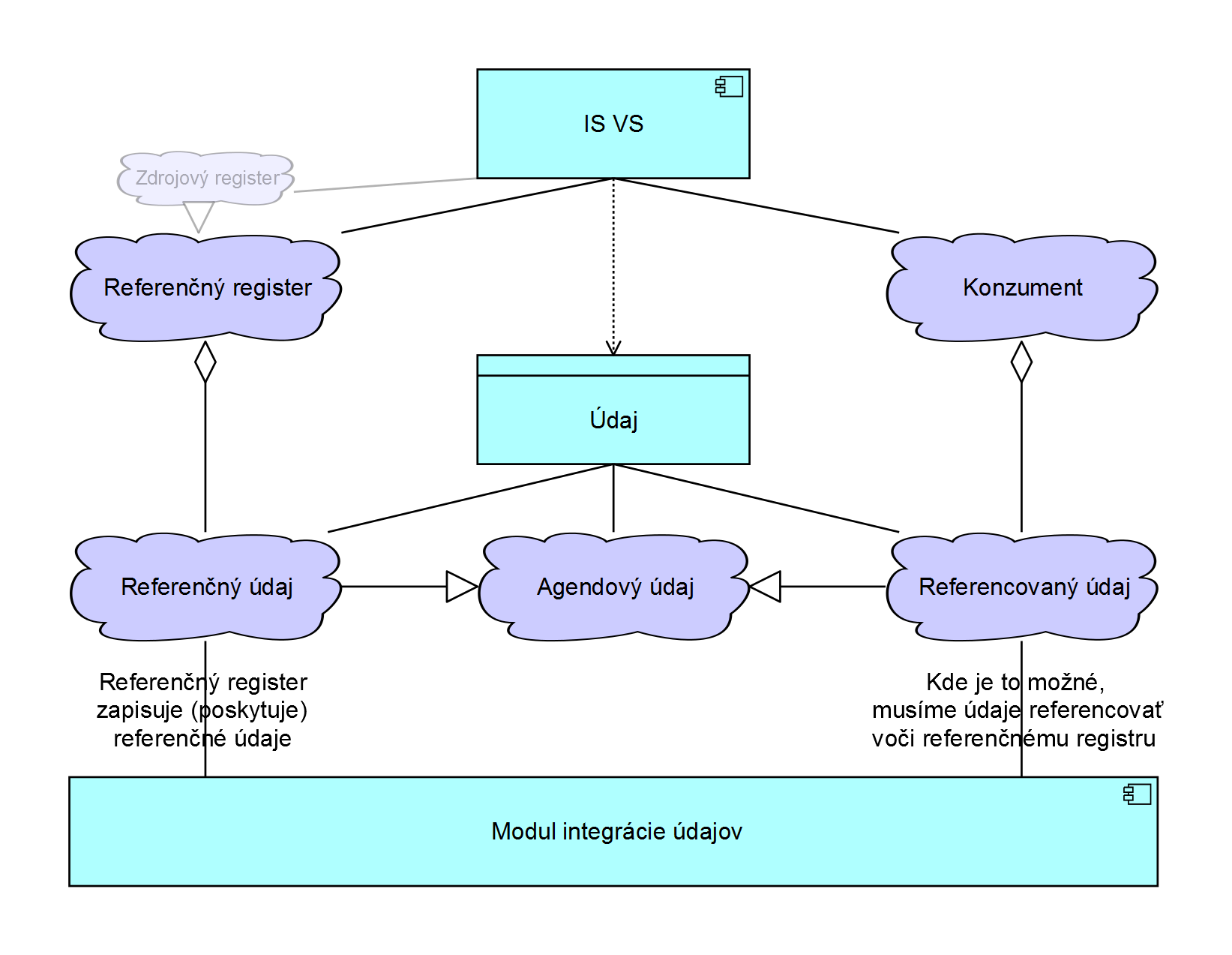
Informačný systém verejnej správy eviduje **agendové údaje**. Napríklad register vozidiel obsahuje informácie o vozidle ako typ vozidla, výrobca, rok výroby, vlastník vozidla, farba a podobne. Niektoré z týchto údajov sú dôležité pre iné ISVS alebo verejnosť; napr. evidenčné číslo vozidla (EČV) alebo majiteľ vozidla. Dá sa povedať, že z pohľadu iných systémov/verejnosti je zmyslom registra evidovať a poskytovať práve tieto údaje. Cieľom je aby, tieto najdôležitejšie údaje registra boli vyhlásené za **referenčné údaje**, čo ostatným zaručuje, že sa na ne môžu spoľahnúť.

Register eviduje aj iné údaje, napr. dátum evidencie vozidla, maximálnu záťaž vozidla a podobne. Niežeby tieto údaje neboli dôležité, ale ich význam pre okolie je menší, tí čo sa nevenujú priamo tej agende ich nutne nepotrebujú. Aj pri týchto údajoch sa musí ISVS snažiť o ich maximálnu možnú kvalitu, ale nie sú v špeciálnom režime ako referenčné údaje.

Niektoré údaje, napr. informácie o vlastníkovi vozidla, sa odkazujú na iné subjekty, v tomto príklade právnickú alebo fyzickú osobu. Nie je úlohou registra vozidiel, aby evidoval informácie o osobách. A už vôbec nie je žiadúce, aby pri zmene priezviska alebo názvu spoločnosti, musel občan oznamovať zmenu v príslušnom registri a ešte aj v registri vozidiel (princíp 1 krát a dosť). Preto, ak register potrebuje mať evidované údaje, ktoré sú doménou iného registra, tzv. **referencované údaje**, tak by sa mal iba odkazovať (referencovať) na údaje v registri, ktorý ich má na starosti. Presnejšie povedané, referenčné údaje iného registra musí ISVS referencovať. Tým sa zabezpečí, že jeho údaje sú správne a je možné naplniť princíp 1 krát a dosť.

Dá sa povedať, že IS VS vystupuje pri svojich dátach v niekoľkých roliach:

1. Referenčný register – pre referenčné údaje, za ktoré zodpovedá
2. Konzument – pre referencované údaje, ktoré musí preberať z iných registrov
3. Agendový systém – pre údaje, ktoré nie sú ani referenčné ani referencované. Aj tieto údaje môže register poskytovať navonok, ale nie sú v špeciálnom režime referenčných údajov.



Špeciálnym prípadom referenčného registra je **zdrojový register**. Príkladom je obchodný register, v ktorom sú evidované informácie o obchodných spoločnostiach (a iné). Aj keď je obchodný register primárnym miestom, kde sú tieto informácie evidované a manažované, posiela ich ďalej do registra právnických osôb. Až RPO je referenčný register, zhromažďuje informácie z viacerých zdrojových registrov, je zodpovedný za ich konsolidáciu a ich poskytovanie ako referenčných údajov. Z pohľadu konzumenta zdrojový register nie je zaujímavý, údaje sa získavajú z referenčného registra, za ktorým je zdrojový register „skrytý“. Čo sa týka kvality zdrojových údajov, je za ne zodpovedný zdrojový register, rovnako ako keby bol referenčným registrom.

## Čo je referencovateľný identifikátor?

Dobrým príkladom jednoznačného referencovateľného identifikátora je ičo pre právnické osoby. Tento identifikátor je

* jednoznačný – žiadne dve právnické osoby nemajú rovnaké ičo, ičo sa nepoužije druhý krát ani keď právnická osoba zanikne,
* jednoducho použiteľný – ide o jednoduchý údaj, ktorý nie je potrebné chrániť, ľahko sa zdieľa.

Ak sa informačný systém odkazuje na právnickú osobu, musí použiť práve tento jednoznačný identifikátor.

Ako identifikátor fyzických osôb sa používa jej rodné číslo v spojení s menom a priezviskom. Tento identifikátor nesie so sebou niekoľko problémov:

* rodné číslo v spojení s menom a priezviskom sa považuje za osobný údaj a teda nie je možné ho zverejňovať,
* rodné číslo nie je bezvýznamové, vieme sa z neho dozvedieť ďalšie informácie (dátum narodenia, pohlavie),
* priezvisko (a aj meno) osoby sa môže meniť, napriek tomu, že samotná fyzická osoba je tá istá, čo komplikuje referencovanie.

Väčšina ISVS eviduje informácie o fyzických osobách, preto definovanie vhodného jednoznačného referencovateľného identifikátora, ktorý by bol jednoducho použiteľný, je kľúčové pre naplnenie princípu 1 x a dosť.

Strategickým zámerom je použitie URI (Jednotný referencovateľný identifikátor) v takom formáte, aby bolo možné vytvárať Linked Data, čo smeruje k vytvorenie sémantického webu (web dát, nie web dokumentov), označovaný ako Web 3.0.

Množina prvkov ktorá opisuje nejakú doménu (napríklad organizácia má sídlo, názov a podobne) sa nazýva ontológia. Používaním štandardizovaných ontológií dochádza k takzvanej integrácii v dátach, pretože pokiaľ ktorýkoľvek subjekt zverejní svoje dáta popísané štandardizovanou ontológiou, tak tieto dáta je možné okamžite spracovať. Podrobnejšie informácie o ontológiách, linked data je možné nájsť XXXXX.

## Bezpečnosť, ochrana osobných údajov, GDPR

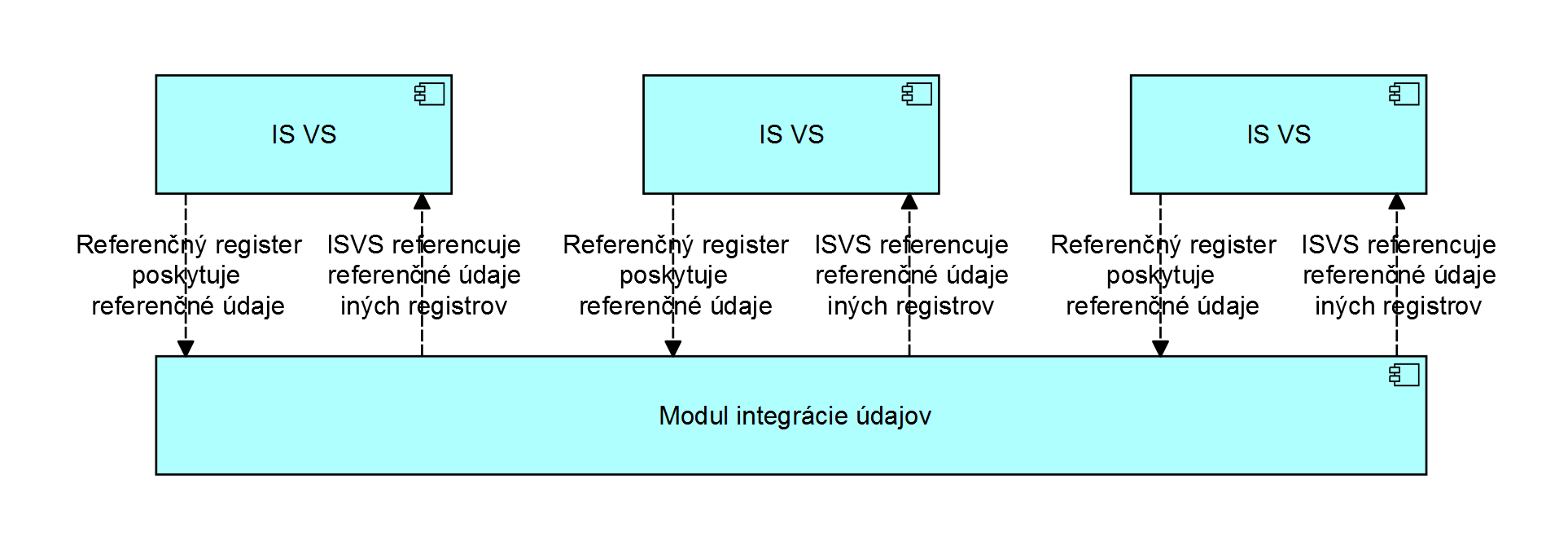
Aby bolo možné efektívne zlepšovať služby poskytované verejnou správou (napr. naplniť princíp 1 x a dosť) je nutné zabezpečiť zdieľanie údajov a ich poskytovanie v čo najširšej možnej miere. Na druhú stranu, ochrana súkromia a bezpečnosť údajov sú rovnako dôležité. Tieto dve sily – zverejňovanie údajov a ich ochrana – idú často proti sebe a je potrebné nájsť vhodný kompromis. Uveďme dva ilustračné príklady.

* Pokiaľ občan chce vybaviť nejakú vec, je pre neho užitočné, ak príslušný úradník má k dispozícii všetky potrebné údaje. Na druhú stranu, treba zabrániť tomu, aby úradník nemohol vyhľadávať údaje, v systémoch do ktorých má prístup, bez toho aby na to mal dôvod.
* V niektorých prípadoch je dôležité uviesť aj osobné údaje. Napríklad, ak sa fyzická osoba dostane do konkurzu, tak sa o nej podľa zákona č. 7/2005 zverejní meno, priezvisko, dátum narodenia a bydlisko. Predstavme si, že o tejto osobe sa v inom ISVS zverejní iba meno a priezvisko. Prepojením jednoznačných identifikátorov je možné doplniť aj ostatné údaje a tak je z pohľadu druhého IS o osobe známych viac údajov ako bolo zamýšľané.

Pomôcť riešiť tieto problémy by mal centrálny modul „Moje dáta“, kde občan môže manažovať informácie, ktoré o ňom štát eviduje.

## Aplikačný pohľad na referenčné údaje a ich integráciu

V zákone č. 305/2013 o e-Governmente je definovaný spoločný modul integrácie údajov (zodpovedá IS CSRU), pričom OVM sú povinné využívať tento spoločný komponent. To znamená, že ISVS musí referencovať referenčné údaje (napr. meno a ičo právnickej osoby) s využitím tohto komponentu, nie priamym napojením na referenčný register. Ak ISVS plní aj úlohu referenčného registra, tak musí poskytovať referenčné údaje tomuto centrálnemu modulu, ktorý ich bude poskytovať ďalej. Tento centrálny modul poskytuje aj centrálne číselníky.



Aby konzument mohol referencovať referenčné údaje, musí ich vedieť vyhľadať a následne získať. Preto modul integrácie údajov musí poskytovať

* metódu search na vyhľadanie objektu evidencie na základe určených kritérií a
* metódu getObj na získanie konkrétneho objektu.

Aby mohol konzument efektívne udržiavať referencované údaje aktuálne, musí modul integrácie údajov poskytovať

* metódu getChanged, ktorá vráti sadu objektov zmenených za obdobie.

Pre referenčný register to znamená, že musí implementovať tieto metódy aby ich centrálny modul prevolával alebo ich centrálny modul využil na udržiavanie svojej vlastnej kópie referenčných údajov.

### Má ISVS obsahovať kópiu referenčných údajov?

Teoreticky by konzument mal mať uložený iba identifikátor objektu evidencie (napr. ičo právnickej osoby) a pri zobrazení získať potrebné referenčné údaje (napr. obchodný názov, adresa) z referenčného registra (cez CSRU). Tým je zaručené, že zobrazované údaje sú vždy aktuálne a správne. Naráža to však na niekoľko praktických problémov:

1. Rýchlosť zobrazenia údajov. Ak by mal konzument uložený obchodný názov, adresu, vedel by záznam zobraziť rýchlejšie ako keď si má tieto údaje vypýtať z referenčného registra a až potom zobraziť.
2. Typicky budú používatelia vyhľadávať nie podľa identifikátora (aj keď v prípade ičo je to nie tak nezvyčajné), ale podľa iných údajov, napr. práve podľa obchodného mena, na čo je nutné mať tieto atribúty uložené. Typickým príkladom je vyhľadávanie právnických osôb podľa obchodného názvu. Riešiť túto úlohu dotazovaním cez MUK na RPO predstavuje diskomfort, pretože práve rýchle vyhľadávanie (ideálne predpĺňanie výsledkov) je presne to čo používateľ očakáva.
3. Ak si dá používateľ zobraziť zoznam objektov, napr. zoznam firiem, ktoré spolupracovali s konkrétnou štátnou firmou, tak očakáva, že tento zoznam bude abecedne zoradený. Dá sa to realizovať aj dotiahnutím jednotlivých údajov a ich následným zoradením, ale je to veľmi neefektívne, v prípade väčších zoznamov až nerealizovateľné.

Preto konzument vo väčšine prípadov potrebuje mať lokálne uložené aspoň niektoré referenčné údaje z iných registrov a tieto údaje udržiavať aktuálne.

Je na zvážení gestora systému, aby rozhodol

* či sa mu oplatí držať vlastnú kópiu údajov, tieto údaje musia byť aktualizované. Tento mechanizmus zvyšuje náklady na vývoj aj údržbu systému, alebo
* bude údaje online preberať, čo sa môže prejavovať menším komfortom pre používateľa.

### Aké sú výhody centrálneho modulu integrácie údajov?

* Ako producent údajov ich poskytnem iba centrálnemu modulu, ak niekto chce využiť moje údaje, rieši integráciu s centrálnym komponentom nie s referenčným registrom.
* Ako konzument údajov nemusím riešiť integráciu s množstvom systémov ale iba s centrálnym komponentom.
* Údaje, ktoré sú otvorené môže producent publikovať priamo cez centrálny komponent.
* Centrálny komponent môže poskytovať analytické dáta
* Poskytovanie informácií o využívaní osobných údajov pre „Moje dáta“ zabezpečí centrálny komponent, keďže on vie kto si čo pýtal.
* Bude CSRU riešiť bezpečnosť údajov?

### Aké sú možné limity centrálneho modulu integrácie údajov?

* Niektoré formy integrácie nie sú podporované na úrovni centrálneho modulu a je lepšie zabezpečiť ich na úrovni referenčného registra/ov, alebo na strane samotného producenta,
* Niektoré údaje zo svojej povahy vyžadujú špecifické požiadavky na integráciu, zdieľanie a opätovné využitie (napríklad osobné, alebo priestorové údaje),
* Centrálne riešenie zvyšuje aj riziko pádu infraštruktúry pri nedostupnosti centrálneho uzla,
* ?

# Kvalita údajov

Čo je kvalita údajov? Chceme, aby dáta boli správne, aby povinné údaje boli vyplnené, aby dáta boli aktuálne, aby údaje dodržiavali dohodnutý formát a podobne. V nasledujúcej časti vysvetlíme jednotlivé atribúty dátovej kvality, ktoré je odporúčané sledovať.

## Atribúty a kritériá kvality

### Referenčná integrita

Týka sa údajov, ktoré boli v predchádzajúcej časti označené ako referencované. Príkladom je identita majiteľa v registri vozidiel. Tieto údaje sú spravované referenčným registrom, ktorý sa stará o ich dátovú kvalitu. Vyzerá to, že ako konzument sa nemusím o nič starať. Cieľom je, že všetky referencované údaje musíme **stotožniť** s údajmi v referenčnom registri, musíme zabezpečiť, aby boli vždy zhodné s údajmi v referenčnom registri. Takto dosiahneme správnosť údajov vo svojom IS a navyše prispejeme k dodržaniu princípu jedenkrát a dosť, lebo referencované údaje sa budú automaticky meniť podľa referenčného registra.

Poznámka: naplnenie princípu 1 krát a dosť môže znamenať, že pri zmene referenčného údaju, informačný systém, ktorý tento údaj referencuje (používa), automaticky nie len upraví údaj, ale ešte aj spustí prípadné ďalšie súvisiace procesy (napr. vydanie nového dokladu). Do domény dátovej kvality však patrí iba zabezpečenie konzistentnosti údajov.

Pri tomto atribúte dátovej kvality je cieľ definovaný priamo zákonom. Referenčná integrita, referencovanie údajov z referenčných registrov je povinné podľa zákona č. 305/2013 § XXXXX. To znamená, ISVS musí dosiahnuť 100% referenčnú integritu.

### Strojová spracovateľnosť

Ako ľahko sú dáta použiteľné strojmi. Pri referenčných údajoch sa predpokladá, že musia byť poskytované vo forme, ktorá je strojovo spracovateľná, predpokladá sa ich poskytovanie cez API.

Klasifikovať strojovú spracovateľnosť údajov je zaujímavé pri otvorených údajoch (tzv[. hviezdičková konvencia](http://5stardata.info/en/)), čo je obsahom iných dokumentov.

### Unikátnosť

Dve právnické osoby nemôžu mať to isté IČO, dve autá nemôžu mať rovnaké EČV (evidenčné číslo vozidla), dve fyzické osoby by nemali mať rovnaké rodné číslo, ... jednoducho povedané, nemôžem mať duplicitný jednoznačný identifikátor.

Na druhú stranu, právnická osoba môže vlastniť niekoľko vozidiel, preto v registri vozidiel sa môže ičo nachádzať aj viackrát.

Ak objekty evidencie definujú jednoznačný referencovateľný identifikátor, tak tento musí byť unikátny (jednoznačný). V takomto prípade používame ako kritérium počet duplicít, ktoré systém obsahuje (napr. počet duplicitných rodných čísel), prípadne počet duplicít voči celkovému počtu záznamov (napr. že 5% záznamov je duplicitných). Cieľom je dosiahnuť hodnotu 0 (alebo 0%).

Požiadavka na unikátnosť je nie len pri definovaní referencovateľných identifikátoroch. Napr. v registri daňových subjektov nemajú dva rôzne subjekty odkazovať na to isté ičo (alebo iný identifikátor osoby). Aj v týchto prípadoch je cieľom je dosiahnuť pri duplicitách hodnotu 0 (0%).

### Aktuálnosť

Údaje musia byť aktuálne, časovo príslušné. Nemalo by sa stať, že OVM vydá rozhodnutie, ale do ISVS príslušné údaje zapisuje iba raz za týždeň a teda údaje v registri niekoľko dní nezodpovedajú skutočnosti.

Ďalšie zdržanie môže byť napríklad medzi zaevidovaním údaju v zdrojovom registri a jeho následnom spracovaní v referenčnom registri.

Problémy s aktuálnosťou údajov môže byť komplikované strojovo identifikovať. Často sú spôsobované komplikovanými procesmi, ktorými sú dáta zapisované do ISVS. Pokiaľ údaj v ISVS nie je, je ťažké zistiť, že nie je aktuálny, nemusí byť jasné, s čím sa má porovnať.

### Kompletnosť, Konzistentnosť, Správnosť

Kompletnosť – či sú všetky povinné údaje vyplnené.  
Konzistentnosť – či sú splnené základné logické podmienky, napr. či je dátum narodenia pred dátumom úmrtia, minimálna hodnota musí byť menšia ako maximálna a podobne.  
Správnosť – či formát údajov zodpovedá definovaným pravidlám, napríklad počet desatinných miest, formát dátumu, maximálna dĺžka reťazca. Súčasťou správnosti, je aj používanie centrálnych číselníkov (napr. číselník obcí).

Minimálnu množinu pravidiel, ktoré je treba dodržať, definuje výnos Ministerstva financií Slovenskej republiky č. 55/2014 Z. z. zo 4. marca 2014 o štandardoch pre informačné systémy verejnej správy v znení neskorších predpisov. Predpisuje údaje napr. o adrese, osobe a podobne. Predpokladá sa, že ISVS definuje ďalšie pravidlá špecifické pre jeho dáta.

V prípade referenčného registra sa očakáva, že kompletnosť bude dosahovať aspoň 97%, pre konzumentov 70%. Pri konzistentnosti a správnosti je cieľom dosiahnuť 100%.

### Presnosť

Tomuto kritériu moc nerozumiem ... . Centrálny model bude spravovať dátová kancelária, definujú ho štandardy, tu sa asi nemá zmysel toto kritérium, ak niečo nie je správne, treba to upraviť. A pre špecifické objekty evidencie v jednotlivých ISVS, to by ten čo to kontroluje musel rozumieť danej doméne. Ťažko si viem predstaviť, ako sa to bude vyhodnocovať a zdvôvodňovať.

Napríklad polohová alebo tematická presnosť v prípade priestorových údajov[[1]](#footnote-1).

## Musia mať údaje 100% kvalitu?

Údaje, hlavne v referenčných registroch, majú zabezpečiť právnu istotu. Ak máme v systéme nesprávne údaje, hrozí, že urobíme nesprávne rozhodnutia, chyby sa môžu šíriť do ďalších systémov. Odďaľovanie vyhlásenia registra za referenčný z dôvodu, že údaje nie sú 100% kvalitné, bráni systematickému používaniu jeho dát a brzdí princíp 1 krát a dosť a znemožňuje zvyšovanie kvality služieb pre občanov.

Ak vyhlásime za referenčný register s nízkou kvalitou dát, môže sa stať, že na základe týchto údajov bude urobené nesprávne rozhodnutie a bude spôsobená škoda. Na druhú stranu, ak je potrebné pre získanie údaju osloviť inštitúciu a čakať 30 dní na odpoveď, tak táto dlhá čakacia doba môže rovnako spôsobovať škody. Navyše, táto čakacia doba je „systémová chyba“, každý žiadateľ je vystavený riziku škody spojeného s touto čakacou dobou. Máme proti sebe dva možné prístupy:

* Čakať, kým kvalita dát dosiahne 100% a potom údaje zverejniť, čo so sebou nesie škody spojené s čakacou dobou, nemožnosťou automatizovania procesov, nemožnosťou integrácie služieb a z toho vyplývajúcou nízkou kvalitou služieb pre občana.
* Zverejniť údaje s istým percentom chýb, kde hrozí, že na základe nesprávnych údajov bude prijaté nesprávne rozhodnutie a spôsobí škodu. Pre takéto prípady, musia byť pripravené postupy, ako sa s týmto vysporiadať.

Je potrebné rozhodnúť, ako nastaviť kritériá kvality tak, aby sa dosiahol praktický kompromis medzi týmito dvoma pohľadmi.

### Rozdeľuj a panuj

Pokiaľ kvalita údajov nedosahuje očakávania vlastníka dát, jedna z možností ako zrýchliť dosiahnutie hodnôt umožňujúcich vyhlásenie údajov za referenčné, je zamerať sa iba na podmnožinu dát. To znamená, že vyhlásime za referenčné nie všetky údaje, ale iba ich dôležitú časť. Často platí, že vyhlásenie aj menšej skupiny údajov za referenčné, môže priniesť veľký pozitívny dopad.

**Historické údaje – za referenčné vyhlásime iba aktuálne, nie historické údaje**. Napríklad povieme, že v registri právnických osôb, sa nachádzajú všetky osoby, ktoré boli aktívne 1.1.2010 a neskôr. To znamená, že firma, ktorá vznikla v roku 2005 a zanikla v roku 2008 sa v registri nenachádza. Firma, ktorá vznikla v roku 2005 a zanikla v roku 2015 sa v registri nachádza, rovnako ako firma ktorá vznikla v roku 2012 bez ohľadu na to, do kedy existovala.   
Môžeme povedať aj reštriktívnejšie pravidlo, napr. že v registri vozidiel sa nachádzajú iba vozidlá zaregistrované od 1.1.2015. To znamená, že vozidlo zaregistrované v roku 2014 sa v registri nenachádza, napriek tomu, že sa stále aktívne používa.   
Pokiaľ je to účelné, môže vlastník dát postupne pridávať do registra historické údaje a posúvať hraničných dátum viac do minulosti.

**Rozdelenie údajov na skupiny.** Napríklad povieme, že register úpadcov (zoznam konkurzov a reštrukturalizácií) je referenčný iba pre konkurzy právnických osôb, nie pre konkurzy fyzických osôb (osobný bankrot). Kritérií, podľa ktorých môže tieto množiny vznikať závisia od konkrétneho registra, ale často je možné určiť množinu údajov, ktorej zverejnenie má výrazný prínos.

### Aká má byť kvalita údajov?

* Vlastník dát identifikuje skupinu údajov, kde vie dosiahnuť vysokú kvalitu údajov
* Kvalita údajov, ktoré budú vyhlásené za referenčné by mala byť 100%. V praxi to znamená, že
  + nemali by sme vedieť, že niektoré údaje nie sú správne
  + nemali by existovať systémové chyby, ktoré spôsobujú vznik nových chybných údajov
  + je úlohou vlastníka dát definovať kritéria dátovej kvality, napr. pripustiť, že niektorý menej dôležitý údaj nie je vždy vyplnený
  + transparentne informovať o kvalite údajov

# Dátová politika

Dátová politika definuje zámery, pravidlá, procesy, role, ktoré sa týkajú dát. Zodpovedá za ňu gestor dátovej politiky, čo je rola pracovníka OVM. Hovoriť by mala o tom, ako a kde vznikajú údaje, ako sa spracovávajú, definovať kvalitu údajov, bezpečnosť údajov, používanie údajov.

Z pohľadu dátovej kvality je dôležité definovať, aké kritériá chceme sledovať, ako ich budeme sledovať a vyhodnocovať (najlepšie automatizovane), aké sú aktuálne hodnoty týchto kritérií, aké hodnoty chceme dosiahnuť, návrhy postupov ako zvyšovať dátovú kvalitu. ), vrátane mechanizmov na ich kontrolu. Tieto kritéria okrem prirodzeného tlaku na zvyšovanie kvality umožnia aj lepšiu porovnateľnosť medzi podobnými údajovými produktmi.

## Princípy dátovej kvality

P1. Riadenie kvality údajov je súčasťou manažérskych funkcií OVM.

P2. OVM má vypracovanú dátovú politiku pre údaje spracovávané v jeho IS. Súčasťou dátovej politiky sú aj definované kritériá dátovej kvality pre spracovávané údaje, rovnako ako aj procesy, ktoré pomáhajú ich dosiahnutiu, monitorovaniu a udržaniu.

P3. OVM pravidelne vyhodnocuje dátovú kvalitu ním spracovaných údajov a proaktívne opravuje prípadné nedostatky v dátovej kvalite.

P4. OVM proaktívne hľadá príčinu prípadných nedostatkov v dátovej kvalite a eliminuje ich výskyt.

## Príklad kritérií kvality

Nasledujúca tabuľka ukazuje príklad definovania kritérií kvality pre fiktívny register vozidiel.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Údaj** | **Referenčný**  **Údaj** | **Kritérium** | **Súčasná**  **hodnota** | **Cieľová**  **hodnota** |
| Evidenčné číslo vozidla | Áno | Unikátnosť – musí byť unikátne | 100% | 100% |
| Typ vozidla | Nie  Mal by byť od 1.8.2018 | Správnosť – hodnoty zodpovedajú číselníku „Typ vozidla“ | 70% | 80% |
| Vlastník vozidla | Áno | Referenčná integrita – Pomer počtu vozidiel, kde atribút vlastník vozidla obsahuje referencovateľný identifikátor fyzickej alebo právnickej osoby ku počtu všetkých vozidiel | 80% | 90% |
| Kompletnosť – povinný údaj, pomer počtu vozidiel s vyplneným vlastníkom ku počtu všetkých vozidiel | 90% | 90% |
| Meno vlastníka vozidla | Nie, referencovaný údaj |  |  |  |
| Dátum zaradenia do evidencie | Nie  Cieľ aby 1.8.2018 bol referenčný | Kompletnosť – povinný údaj | 95% | 100% |
| Správnosť – formát dátumu DD.MM.YYYY | 100% | 100% |
| Konzistentnosť – dátum zaradenia pre dátum vyradenia | 100% | 100% |
| Dátum vyradenia z evidencie | Nie  Cieľ aby 1.8.2018 bol referenčný |  |  |  |

## Ako zvyšovať kvalitu údajov?

1. Školenia, poradenstvo z centrálnej úrovne (dátová kancelária)
2. Využitie služieb centrálneho modulu integráciu údajov.
   1. Profilovanie dát – na základe definovaných kritérií kvality údajov sa využije centrálny nástroj na zistenie aktuálnej kvality dát, identifikovanie problémov
   2. Stotožňovanie – na základe definovaných pravidiel nástroj stotožní záznamy s referenčnými údajmi. Tam kde to nie je možné (napr. nie je jednoznačne určených identifikátor), ponúkne možnosti
3. Využitie centrálneho modulu ako platformy, tzn. využiť centrálny nástroj (talend). Rozdiel oproti využitiu služieb je v tom, že v tomto prípade centrum poskytne iba nástroj, samotné čistenie dát vykoná ich vlastník.
4. Úprava procesov tak, aby zabraňovali vzniku chýb v údajoch, tzn. riešiť aby dáta od nejakého okamihu mali vyššiu kvalitu. Ináč povedané, na základe identifikovaných problémov v dátach, hľadať ich príčiny a tie odstrániť. Nerieši to už existujúce dáta (historický dlh v kvalite dát), ale umožní to vyhlásiť údaje za referenčné od dátumu implementácie týchto procesov.
5. Po stotožnení referenčných údajov, upraviť vlastné kópie dát tak, aby boli v súlade s referenčnými.
6. Zvážiť kritériá kvality, či nie sú zbytočne prísne. Pre niektoré atribúty, ktoré nie sú kritické je možné nastaviť toleranciu, aké chyby pripúšťame. Preferovať použitie štandardizovaných kritérií.
7. Jasne upozorňovať na dáta, kde kvalitu nevieme zaručiť, napríklad vložením „neznámy“ alebo „nedefinovaný“.
8. Pre komplikovanejšie závislosti a kontroly vyvinúť komponenty, ktoré ich riešia.
9. Využitie umelej inteligencie na rozpoznávanie chýb v dátach a ich automatické opravy.

# Povinnosti

## Čo musí dodržať každý ISVS?

1. Podľa § 52 ods.1 zákona č. 305/2013 Z. z. je každý ISVS povinný referencovať údaje, ktoré sú vyhlásené za referenčné, do jedného roka od vyhlásenia referenčného registra. Napr. všetky fyzické osoby, právnické osoby musia byť referencované na referenčný register.
2. OVM musí definovať svoju dátovú politiku, ktorou smeruje k zvyšovaniu kvality dát a dodržiavať ju.
3. OVM urobí potrebné organizačné zmeny v súlade s dátovou politikou. Napríklad vymenuje riadiaceho pracovníka- gestora zodpovedného za dátovú politiku a dátovú kvalitu, ustanoví výkonnú rolu dátovej kvality (Data Steward) a podobne.
4. OVM podpíše dohodu o integračnom zámere s modulom dátovej integrácie. OVM je povinný používať tento centrálny komponent na získanie referenčných údajov podľa § 10 ods. 11 písm. c) zákona č. 305/2013 Z. z.
5. OVM poskytuje centrálnej dátovej kancelárii základné údaje o kvalite svojich dát.
6. Údaje, ktoré sú využívané inými IS by mali byť v čo najkratšom čase vyhlásené za referenčné, aby ich využívanie malo aj právne účinky.

## Čo musí dodržať referenčný register?

1. Referenčný register je konzumentom referenčných údajov iných registrov, preto musí dodržiavať všetky pravidlá z predchádzajúcej časti.
2. Podľa § 10 ods. 11 písm. c) je referenčný register povinný poskytovať tieto údaje pomocou modulu dátovej integrácie.

## Ako postupovať pri vyhlásení referenčného registra?

TODO

# Governance z pohľadu údajov, integrácie a kvality údajov

V strategickej priorite Manažment údajov je definované, že preferovaný spôsob riadenia kvality údajov je hierarchický, tzn. centrálna metodika, parametre, procesy a monitoring, decentralizovaná zodpovednosť za kvalitu údajov. Preto je nutné mať fungujúcu dátovú kanceláriu, ktorá definuje pravidlá, udržiavať centrálny model. Aj keď dátová kancelária plní úloha garanta globálneho procesu, musí byť nie len kontrolným orgánom, ktorý predstavuje potenciálne úzke hrdlo, ale hlavne byť poradným orgánom, ktorý iniciatívne napomáha zrýchleniu postupov, vyhlasovaniu referenčných registrov.

1. Dátová kancelária schvaľuje vyhlásenie údajov za referenčné.
2. OVM v spolupráci s dátovou kancelária navrhne spoločný ontologický model, prípadne rozšírenie existujúceho modelu (príp. metadát, číselníkov a podobne), pričom dátová kancelária dbá na jeho konzistenciu a súlade medzi doménami jednotlivých OVM.
3. Dátová kancelária zbiera informácie o kvalite dát jednotlivých OVM, konsoliduje tieto údaje a zverejňuje ich.
4. Dátová kancelária aktívne podporuje vyhlasovanie referenčných registrov, aby bolo možné naplniť princíp 1 krát a dosť. Preto dátová kancelária udržiava plán vyhlasovanie referenčných údajov.

# Kto to bude platiť?

Národný projekt: Dátová integrácia – k tomu je štúdia, na základe reformného zámeru *Koncepčné budovanie digitálnej a inovatívnej VS*

* sprístupnenie infraštruktúry pre realizáciu princípu „Jeden krát a dosť“
* sprístupňovanie otvorených údajov pre „rozhodujúcu masu“ OVM

# Definície pojmov a skratky

Referenčná integrita – atribút dátovej kvality. Údaje v objekte evidencie sú referencované s referenčnými údajmi v podobe, v akej sú evidované v referenčných registroch. Referenčné údaje z objektu evidencie sú stotožnené so subjektom evidencie.

Ďalšie atribúty

# Relevantné dokumenty a zdroje

|  |  |
| --- | --- |
| Dokument | Popis |
| **Zákon 305/2013 Z. z. o elektronickej podobe výkonu pôsobnosti orgánov verejnej moci a o zmene a doplnení niektorých zákonov (zákon o e-Governmente)**  <https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2013/305/20180101> |  |
| **Národná koncepcia informatizácie verejnej správy (2016)**  Schválená: September 2016  <http://www.informatizacia.sk/narodna-koncepcia-informatizacie-verejnej-spravy--2016-> |  |
| **Strategický dokument: Strategická priorita Manažment údajov (pracovná skupina lepšie dáta)**  Schválený: Február 2017  <http://www.informatizacia.sk/ext_dok-uppvii_sp_manazment_udajov_vfinal/24196c> |  |
| **Štúdia uskutočniteľnosti: Dátová integrácia: sprístupnenie údajovej základne VS vrátane otvorených údajov prostredníctvom platformy dátovej integrácie**  Schválená: November 2017  <https://metais.finance.gov.sk/studia/detail/6a3eda58-d61a-33e8-4034-33afda4eb51a?tab=basicForm> |  |
| **Metodické usmernenie o postupe zaraďovania referenčných údajov do zoznamu č. MF/012027/2016-171**  Schválené: Marec 2016  <http://www.informatizacia.sk/metodicke-postupy/22980s> |  |
| **Metodické usmernenie Ministerstva financií Slovenskej republiky**  **č. MF/022908/2015-171 pre riadenie kvality údajov**  <http://informatizacia.sk/ext_dok-metodicke-usmernenie...c...riadenie-kvality.../22354c> |  |
| **Zoznam referenčných registrov**  Priebežne aktualizovaný  <https://metais.finance.gov.sk/refregisters/list?page=1&count=20> |  |
| **Akčný plán** |  |
|  |  |
| **Datasety** |  |
| **LOD** |  |
| **Metodika k časovému harmonogramu vytvárania a sprístupňovania datasetov**  <http://informatizacia.sk/ext_dok-metodika_k_casovemu_harmonogramu_vytvarania_a_spristupnovania_datasetov/26256c> |  |

1. <http://inspire.ec.europa.eu/documents/INSPIRE_/JRC83209_Online_Data_quality_in_INSPIRE.pdf> [↑](#footnote-ref-1)