

Výpočet KPI dátovej kvality pre jednotlivé inštitúcie verejnej správy

Projekt

Zlepšenie využívania údajov vo verejnej správe

Zmluva o dielo č. 321/2018

Výstup č.2



ÚRAD PODPRESEDU VLÁDY SR
PRE INVESTÍCIE
A INFORMATIZÁCIU

Tento projekt je podporený z Európskeho sociálneho fondu.

Zoznam skratiek

RU	Register úpadcov
RA	Register adries
RPO	Register právnických osôb
N/A	Not Applicable - Neaplikovateľné
KPI	Key Performance Indicator – kľúčový ukazovateľ výkonnosti
ID	Identifikačné číslo
UPVII	Úrad podpredsedu vlády pre investície a informatizáciu Slovenskej republiky
RFO	Register fyzických osôb

Obsah

1	Manažérske zhrnutie pre výpočet KPI dátovej kvality	3
2	Karty merania dátovej kvality	4
2.1	Karta merania pre Register úpadcov	4
2.2	Karta merania pre Register adries	5
2.3	Karta merania pre Register právnických osôb	6
3	Tabuľka KPI a jej použitie v meraní	10
4	Prvé merané KPI: ukazovateľ Sledovanie konzistentnosti	17
5	Zmerané KPI pre profiláciu dát	18
5.1	Výsledky merania dátovej kvality pre KPI v rámci profilácie	18
5.2	Zhrnutie merania dátovej kvality pre KPI v rámci profilácie	19
6	KPI merania dátovej kvality so znalosťou biznis pravidiel	21
6.1	Vybrané biznis pravidlá pre vybrané registre	21
6.2	KPI otázky a podmienky merania pre vybrané biznis pravidlá	27
6.3	Výsledok merania KPI ukazovateľov pre meranie dátovej kvality vybraných biznis pravidiel pre vybrané registre	29
6.4	Zhodnotenie merania pre vybrané biznis pravidlá	30
7	Finálne merané KPI: ukazovateľ Dodržiavanie biznis pravidla	35
8	Záverečné zhrnutie pre výpočet KPI dátovej kvality	37
9	Zoznamy	38
9.1	Zoznam tabuliek	38
10	Prílohy	39
10.1	Príloha A: Výsledky merania dátovej kvality pre KPI v rámci profilácie.xlsx	39

Tento projekt je podporený z Európskeho sociálneho fondu. 2

1 Manažérske zhrnutie pre výpočet KPI dátovej kvality

Výpočet KPI dátovej kvality je ústredným pilierom merania aj riadenia dátovej kvality.

Kľúčové ukazovatele výkonnosti (angl. key performance indicators, skr. KPI) dátovej kvality sú praktické ukazovatele, ktoré kvantifikujú a objektivizujú dátovú kvalitu vo väzbe na príslušný globálny cieľ či kritický faktor úspechu.

Celková dátová kvalita registra je zložená z výkonností jeho jednotlivých častí, entít a atribútov. Preto je veľmi dôležité správne rozloženie zvolených ukazovateľov dátovej kvality na konkrétne dataseťy. Vhodne zvolené KPI umožní dostatočne presne monitorovať dátovú kvalitu konkrétneho datasetu s ohľadom na tvorbu výstupov a pridanej hodnoty.

Štandardizovanie metodiky výpočtu KPI, určenie dátových zdrojov nevyhnutných na jeho vyčíslenie sú rovnako dôležité ako samotné určenie KPI. V tomto kroku je vhodné zohľadniť možnosti vyhodnocovania dátovej kvality, ktoré ponúka existujúci informačný systém. Prípadne je užitočné zvážiť možnosti extrahovania dát z informačných systémov a ich následné automatizované spracovanie v špeciálnom riešení pre meranie a riadenie dátovej kvality.

Prepojenie KPI dátovej kvality na ohodnotenie zodpovedných osôb pôsobí motivujúco v prípade, ak majú konkrétni pracovníci potrebné kompetencie a reálnu možnosť svojim správaním KPI ukazovateľ ovplyvniť a zároveň sú o jeho hodnote a vývoji pravidelne informovaní. Preto pravidelné systematické meranie dátovej kvality je dôležitým predpokladom jej postupného cieleného zlepšovania.

2 Karty merania dátovej kvality

Každé meranie dátovej kvality je definované kartou merania. Táto karta merania obsahuje všetky potrebné informácie na identifikáciu objektu merania, ale aj dodatočných dokumentov, ktoré sú nevyhnutné pre správne meranie dátovej kvality. Nižšie sú vyplnené tabuľky pre merania vybraných registrov:

- Register úpadcov,
- Register adries,
- Register právnických osôb.

Informácie v tabuľkách boli dopĺňané aj v priebehu merania až do úplného konca. V celom meraní dátovej kvality je potrebné zdôrazniť absenciu dokumentovaných aktuálnych biznis pravidiel. Práve biznis pravidlá sú z pohľadu merania KPI najdôležitejšie. Aktuálne biznis pravidlá totiž odzrkadľujú aj aktuálny stav informačných systémov, ktoré s databázou dát pracujú. V prípade referenčných registrov ide aj o informačné systémy zdrojových databáz.

2.1 Karta merania pre Register úpadcov

V tabuľke nižšie sú popísané najdôležitejšie informácie, vzťahujúce sa na objekt merania: Register úpadcov.

Identifikačné číslo karty merania	N/A
Názov merania dátovej kvality	Prvé meranie dátovej kvality Registra úpadcov
Definícia účelu merania	Prvé využitie metodiky merania dátovej kvality vo verejnej správe
Názov datasetu	Register úpadcov – CodeDb
Zdroj datasetu	Databáza z informačného systému Register úpadcov: Údaje v Registri úpadcov vznikajú kombináciou dát z informačných systémov súdov, Obchodného vestníka a Registra správco
Poskytovateľ datasetu	Úrad podpredsedu vlády SR pre investície a informatizáciu v spolupráci s Ministerstvom Spravodlivosti SR
Stručná definícia datasetu	Informačný systém Register úpadcov obsahuje informácie o konkurzných a reštrukturalizačných konaniach podľa § 10a ods. 2 zákona č. 7/2005 Z. z. o konkurze a reštrukturalizácii a o zmene a doplnení niektorých zákonov v znení neskorších predpisov.
Výber vzorky záznamov v datasete	7,2%: 7 tabuliek databázy CoreDb: core.konanie, core.kraj, core.majetok, core.osoba, core.pohľadavka, core.spravca, core.sud (z celkového množstva 97 tabuliek). Tabuľky boli vybraté podľa identifikácií databázových tabuliek v rámci nájdených neformálnych biznis pravidiel.

Originálny formát datasetu	SQL – dátová časť exportu
Dátum a čas exportu datasetu	2019-04-05
Dátum a čas aktuálnosti datasetu	2019-04-05
Dátum prvého dňa merania	2019-06-03
Dátum posledného dňa merania	2019-06-28
OVM zodpovedný za dataset	Ministerstvo Spravodlivosti SR
Zodpovedný za meranie	KPMG Projektový manažér
Osoby vykonávajúce meranie	KPMG Projektový tím
Aktuálne dátové štandardy pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálny dátový model pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálne metadáta pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálne biznis pravidlá pre dataset	Nie sú formálne definované, dokumentované, zdieľané a schválene v plnom rozsahu

Tabuľka 1: Karta merania pre Register úpadcov

2.2 Karta merania pre Register adries

V tabuľke nižšie sú popísané najdôležitejšie informácie, vzťahujúce sa na objekt merania: Register adries.

Identifikačné číslo karty merania	N/A
Názov merania dátovej kvality	Prvé meranie dátovej kvality Registra adries
Definícia účelu merania	Prvé využitie metodiky merania dátovej kvality vo verejnej správe
Názov datasetu	Register adries
Zdroj datasetu	Databáza z informačného systému Register adries: Do registra adries zapisujú údaje obce a mestá, v ktorých originálnej kompetencii je určovanie súpisných a orientačných čísiel ako aj určovanie názvov ulíc a iných verejných priestranstiev. Údaje o územnom členení - kraj, okres, obec, časť obce - zapisuje do registra Ministerstvo vnútra Slovenskej republiky.
Poskytovateľ datasetu	Úrad podpredsedu vlády SR pre investície a informatizáciu v spolupráci s Ministerstvom Vnútra SR

Stručná definícia datasetu	Register adries ako informačný systém verejnej správy vznikol v súlade s architektúrou integrovaného informačného systému verejnej správy navrhnutou Národnou koncepciou informatizácie verejnej správy ako jeden zo 4 základných registrov. Informačný systém registra adries, ktorého úprava je obsiahnutá v zákone č. 125/2015 Z. z. o registri adries a o zmene a doplnení niektorých zákonov, je dátovo konzistentným, centralizovaným zdrojom dát o adresách všetkých fyzických budov na území Slovenskej republiky, ktorým bolo určené súpisné číslo.
Výber vzorky záznamov v datasete	100%: 8 tabuliek databázy: 01_Kraj, 02_Okres, 03_Obce, 04_Časť Obcí, 05_Ulice, 06_Budovy_Súpisné čísla, 07_Vchody_Orientačné čísla, 08_Byty (z celkového množstva 8 tabuliek)
Originálny formát datasetu	XML
Dátum a čas exportu datasetu	2019-03
Dátum a čas aktuálnosti datasetu	2019-03
Dátum prvého dňa merania	2019-06-03
Dátum posledného dňa merania	2019-06-28
OVM zodpovedný za dataset	Ministerstvo Vnútra SR
Zodpovedný za meranie	KPMG Projektový manažér
Osoby vykonávajúce meranie	KPMG Projektový tím
Aktuálne dátové štandardy pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálny dátový model pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálne metadáta pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálne biznis pravidlá pre dataset	Nie sú formálne definované, dokumentované, zdieľané a schválené v plnom rozsahu

Tabuľka 2: Karta merania pre Register adries

2.3 Karta merania pre Register právnických osôb

V tabuľke nižšie sú popísané najdôležitejšie informácie, vzťahujúce sa na objekt merania: Register právnických osôb.

Identifikačné číslo karty merania	N/A
Názov merania dátovej kvality	Prvé meranie dátovej kvality Registra právnických osôb
Definícia účelu merania	Prvé využitie metodiky merania dátovej kvality vo verejnej správe
Názov datasetu	Register právnických osôb
Zdroj datasetu	<p>Údaje v Registri PO pochádzajú predovšetkým zo všetkých evidencií, registrov, zoznamov (exemplifikatívne uvedených vyššie), ktoré sú akýmiś zdrojovými registrami, alebo sú získané od príslušných orgánov verejnej moci, ktoré vydávajú doklady a oprávnenia na výkon podnikateľskej činnosti.</p> <p>Okrem údajov z Obchodného a Živnostenského registra, evidencií Ministerstva vnútra a Ministerstva spravodlivosti SR a okresných úradov, sa ostatné registre, evidencie a zoznamy integrujú až do 2 rokov od účinnosti Zákona, t. j. až do 1.11.2017. (poznámka ku zdroju datasetu RPO*)</p>
Poskytovateľ datasetu	Úrad podpredsedu vlády SR pre investície a informatizáciu v spolupráci so Štatistickým úradom SR
Stručná definícia datasetu	Register PO je informačným systémom verejnej správy, ktorý obsahuje údaje stanovené zákonom: i) o právnických osobách; ii) o podnikateľoch; iii) o orgánoch verejnej moci.
Výber vzorky záznamov v datasete	7 tabuliek databázy: rpo.Subject_Address, rpo.Subject_Detail, rpo.Subject_Has_Activities, rpo_Subject_LegalForms, rpo_Subjects, rpo_cl_ESA_Codes, rpo_cl_MainActivities (celkové množstvo tabuliek nebolo možné identifikovať z dôvodu nevhodného formátu vstupných dát). Poskytnutých bolo 819 462 XML súborov, kde každý z nich predstavoval informácie o jednej právnickej osobe v rámci konkrétneho zdrojového registra. V niektorých prípadoch boli v plnom rozsahu vybrané informácie z dvoch najväčších zdrojových registrov (obchodný register: 302 538 XML súborov – 100% zdrojového registra, a živnostenský register: 407 212 XML súborov – 100% zdrojového registra).
Originálny formát datasetu	XML
Dátum a čas exportu datasetu	2019-04-25
Dátum a čas aktuálnosti datasetu	2019-04-25
Dátum prvého dňa merania	2019-06-03
Dátum posledného dňa merania	2019-06-28
OVM zodpovedný za dataset	Štatistický úrad SR
Zodpovedný za meranie	KPMG Projektový manažér
Osoby vykonávajúce meranie	KPMG Projektový tím

Aktuálne dátové štandardy pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálny dátový model pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálne metadáta pre dataset	Neboli dostupné v potrebnej kvalite
Aktuálne biznis pravidlá pre dataset	Nie sú formálne definované, dokumentované, zdieľané a schválene v plnom rozsahu

Tabuľka 3: Karta merania pre Register právnických osôb

***Poznámka k zdroju datasetu RPO:**

Register právnických osôb obsahuje informácie z veľkého množstva zdrojových registrov a ďalších databáz informačných systémov. Nižšie je uvedený zoznam väčšiny evidovaných zdrojových registrov v rámci RPO:

- Živnostenský register
- Obchodný register
- Register občianskych združení;
- Register finančných agentov a finančných poradcov
- Register spoločenstiev vlastníkov bytov a nebytových priestorov
- Register taxislužby
- Register odborových a zamestnávateľských organizácií
- Register neziskových organizácií poskytujúcich všeobecnoprospešné služby
- Register prevádzkovateľov cestnej dopravy
- Register fyzických osôb v športe
- Register záujmových združení právnických osôb
- Register nadácií
- Register neinvestičných fondov
- Zoznam správcov
- Register neziskových organizácií poskytujúcich všeobecne prospešné služby
- Register právnických osôb, ktoré odvodzujú právnu subjektivitu od právnej subjektivity cirkví a náboženských spoločností
- Bratislavský samosprávny kraj - Register povolení na prevádzkovanie zdravotníckych zariadení
- Register združení obcí
- Banskobystrický samosprávny kraj - Register povolení na prevádzkovanie zdravotníckych zariadení

- Trenčiansky samosprávny kraj - Register povolení na prevádzkovanie zdravotníckych zariadení
- Žilinský samosprávny kraj - Register povolení na prevádzkovanie zdravotníckych zariadení
- Nitriansky samosprávny kraj - Register povolení na prevádzkovanie zdravotníckych zariadení
- Zoznam exekútorov
- Prešovský samosprávny kraj - Register povolení na prevádzkovanie zdravotníckych zariadení
- Zoznam prekladateľov
- Trnavský samosprávny kraj - Register povolení na prevádzkovanie zdravotníckych zariadení
- Register súkromných veterinárnych lekárov
- Register mediátorov
- Register politických strán a politických hnutí
- Register organizácií s medzinárodným prvkom
- Nitriansky samosprávny kraj - Register povolení na poskytovanie lekárenskej starostlivosti
- Trnavský samosprávny kraj - Register povolení na poskytovanie lekárenskej starostlivosti
- Trenčiansky samosprávny kraj - Register povolení na poskytovanie lekárenskej starostlivosti
- Banskobystrický samosprávny kraj - Register povolení na poskytovanie lekárenskej starostlivosti
- Žilinský samosprávny kraj - Register povolení na poskytovanie lekárenskej starostlivosti
- Prešovský samosprávny kraj - Register povolení na poskytovanie lekárenskej starostlivosti
- Register cirkví a náboženských spoločností
- Bratislavský samosprávny kraj - Register povolení na poskytovanie lekárenskej starostlivosti
- Ministerstvo zdravotníctva SR - Register povolení na prevádzkovanie zdravotníckych zariadení
- Register európskych zoskupení územnej spolupráce

3 Tabuľka KPI a jej použitie v meraní

Z pohľadu merania dátovej kvality je dôležitá znalosť parametrov a ukazovateľov. Avšak KPI pre meranie dátovej kvality sú kľúčové z pohľadu reprezentácie výsledkov merania. A samotné prahové hodnoty merania sa viažu práve na tieto KPI.

Jeden ukazovateľ môže obsahovať aj viacero KPI. Ich počet je možné v budúcnosti dopĺňať podľa potreby reprezentácie výsledkov.

Ako je možné určiť, ktoré KPI je pre meranie dátovej kvality dôležité? Dôležité sú všetky, avšak ich meranie v rámci jednotlivých atribútov tabuliek datasetu môže byť rôzne. Teda odpoveď na to, či je potrebné zmerať KPI pre konkrétny atribút, je skrytá v biznis pravidle. Pre zjednodušenie hľadania odpovedí, v tabuľke nižšie sú definované KPI otázky k jednotlivým KPI. Pokiaľ existuje kladná odpoveď (KPI podmienka je splnená) na KPI otázku v rámci biznis pravidla, s určitosťou existuje aj možnosť merania relevantného KPI pre atribút alebo skupinu atribútov.

Výpočty KPI hodnôt je možné vykonávať vo viacerých systémoch ako napríklad: Talend, SAS, SAP, Informatica, IBM a ďalšie. Pre túto situáciu bola profilácia dát vykonaná v Talende a pokročilejšie analýzy pomocou SQL dotazov.

Ukazovateľ	Typ merania	Počítané v Talende	KPI ID	KPI	KPI popis	KPI otázka (platí pre atribút alebo skupinu atribútov, odpoveďou je biznis pravidlo)	KPI podmienka merania	KPI meranie
Syntaktická presnosť hodnoty	Pokročilejšia analýza	Áno	1.1a	N/A POZOR: KPI na meranie dátovej kvality ukazovateľa Syntaktická presnosť hodnoty má využitie v procese opravovania dát na identifikáciu presne správnej alebo najpravdepodobnejšie správnej hodnoty pri adeptoch na opravu.	Za hodnotu atribútu, ktorý má definované biznis pravidlo definujúce očakávané hodnoty napríklad číselník (zoznam hodnôt), kombinácia hodnôt z číselníka (zoznam hodnôt), hodnoty podľa pravidiel gramatiky, veľké malé písmená, intervalové hodnoty napr. od do alebo komplexnejšie intervaly a pod. = definícia povolených hodnôt v biznis pravidle.	Aké sú presne očakávané hodnoty atribútu? Napr. číselník (zoznam hodnôt), kombinácia hodnôt z číselníka (zoznam hodnôt), hodnoty podľa pravidiel gramatiky, veľké malé písmená, intervalové hodnoty napr. od do alebo komplexnejšie intervaly a pod. POZOR: pre atribút nesmie existovať zdroj pravdy.	Ak existujú definované očakávané hodnoty, tak sa môže merať KPI 1.1a alebo 1.2a	Nie
Syntaktická presnosť atribútu	Pokročilejšia analýza	Áno	1.2a	Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje hodnoty v súlade s definovaným pravidlom povolujúcim určené hodnoty.	Za atribút, ktorý má definované biznis pravidlo definujúce očakávané hodnoty napríklad číselník (zoznam hodnôt), kombinácia hodnôt z číselníka (zoznam hodnôt), hodnoty podľa pravidiel gramatiky, veľké malé písmená, intervalové hodnoty napr. od do alebo komplexnejšie intervaly a pod. = definícia povolených hodnôt v biznis pravidle.	Aký je zdroj pravdy pre atribút a kde sa nachádza zdroj pravdy? Zdrojom pravdy sa myslí iný dátový zdroj, ktorý sa najviac približuje pravdivým hodnotám - realite. POZOR: Ak je odpoveď na existenciu zdroja pravdy kladná, meranie syntaktickej presnosti atribútu a syntaktickej presnosti hodnoty nemá prebiehať v určenom atribúte ale v zdroji pravdy.	Ak existuje definovaný zdroj pravdy, tak sa môže merať KPI 1.3a	Áno
Sémantická presnosť atribútu	Pokročilejšia analýza	Áno	1.3a	Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje hodnotu v súlade s definovaným biznis pravidlom odkazujúcim na zdroj pravdy.	Za atribút, ktorý má definované biznis pravidlo s určeným zdrojom pravdy. Takýto atribút preberá pravidlá syntaktickej presnosti z biznis pravidiel v zdroji pravdy a teda merá syntaktickú presnosť atribútu a syntaktickú presnosť hodnoty v určenom atribúte nemá zmysel. Syntaktická presnosť hodnoty a syntaktická presnosť atribútu má byť potom meraná v zdroji pravdy.			Áno

Tento projekt je podporený z Európskeho sociálneho fondu.

11

Ukazovateľ	Typ merania	Počítané v Talende	KPI ID	KPI	KPI popis	KPI otázka (platí pre atribút alebo skupinu atribútov, odpoveďou je biznis pravidlo)	KPI podmienka merania	KPI meranie
Sledovanie konzistentnosti	Prvé odborné posúdenie	Nie	2.1a	Percentuálny podiel atribútov, ktoré majú definované biznis pravidlá.	Za skupinu atribútov v celom meranom datase, ktoré majú mať definované biznis pravidlo. POZOR: atribút bez biznis pravidla nemôže mať meranú dátovú kvalitu.	N/A	N/A	Áno
Dodržiavanie biznis pravidla	Posledné odborné posúdenie	Nie	2.2a	Percentuálny podiel atribútov, ktorých hodnoty plne dodržia definované biznis pravidlá.	Za skupinu atribútov, ktoré majú mať definované biznis pravidlo. POZOR: atribút bez biznis pravidla nemôže mať meranú dátovú kvalitu.	N/A	N/A	Áno
	Pokročilejšia analýza	Áno	2.2b	Percentuálny podiel atribútov, ktorých hodnoty plne dodržia vzťahy medzi atribútmi.	Za skupinu atribútov, ktoré majú mať definovanú procesnú stránku biznis pravidla. Procesné a chronologické väzby medzi atribútmi tzn. naprogramovaná biznis logika informačných systémov, ktoré naplňajú a menia dáta jednotlivých atribútov.	Aká je procesná stránka atribútu resp. aké sú vzťahy medzi atribútom a relevantnými atribútmi?	Ak existujú definované väzby medzi atribútmi, tak sa môže merať KPI 2.2b	Áno
Dodržiavanie formátu atribútu	Profiling	Áno	3.1a	Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje hodnotu v požadovanom formáte.	Za atribút, ktorý má definované biznis pravidlo definujúce formát hodnoty.	Aký je formát hodnoty v atribúte?	Ak existuje definovaný formát, tak sa môže merať KPI 3.1a	Áno
Design v súlade so štandardom	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Rozlišovanie 'null' a prázdnej hodnoty	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie

Tento projekt je podporený z Európskeho sociálneho fondu.

12

Ukazovateľ	Typ merania	Počítané v Talende	KPI ID	KPI	KPI popis	KPI otázka (platí pre atribút alebo skupinu atribútov, odpoveďou je biznis pravidlo)	KPI podmienka merania	KPI meranie
Vyplnenosť povinného údajá	Profiling	Áno	4.2a	Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje akúkoľvek hodnotu okrem 'null' a prázdnej hodnoty.	Za atribút, ktorý má definované biznis pravidlo definujúce povinne vyplnenú hodnotu. POZN.: Výpočet je silne závislý na definovaní očakávaných hodnôt vstupu. Pri nejasnom zadefinovaní významu 'null' a prázdnej hodnoty dôjde k skresleným výsledkom merania dátovej kvality.	Je atribút povinný? POZOR: metadáta nemusia obsahovať informácie o všetkých povinných atribútoch z pohľadu biznisu. Povinnosť je niekedy ošetrená v samotných vstupoch do databázy, napríklad cez webové formuláre.	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a	Áno
Vyplnenosť nepovinného údajá	Profiling	Áno	4.3a	Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje akúkoľvek hodnotu vrátane prázdnej hodnoty a okrem 'null' hodnoty. POZOR: Toto KPI je ako doplnok k meraniu dátovej kvality. Nesmie sa na neho viazať prahová hodnota. Úroveň dátovej kvality je len informatívna hodnota keďže hodnoty atribútu z biznis pohľadu nie sú povinné.	Za atribút, ktorý má definované biznis pravidlo definujúce nepovinne vyplnenú hodnotu. POZN.: Výpočet je silne závislý na definovaní očakávaných hodnôt vstupu. Pri nejasnom zadefinovaní významu 'null' a prázdnej hodnoty dôjde k skresleným výsledkom merania dátovej kvality.			Áno
Unikátnosť identifikátora	Odborné posúdenie / Pokročilejšia analýza	Áno	N/A	N/A	N/A	N/A	N/A	Nie
Unikátnosť hodnoty	Profiling	Áno	5.2a	Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje rovnakú hodnotu v rámci atribútu len jeden krát. Pozn.: Percentuálny podiel distinktných záznamov v atribúte tabuľky sa vyráta ako súčet unikátnych a multiplicitných záznamov.	Za atribút, ktorý má definované biznis pravidlo definujúce jedinečnú hodnotu vo všetkých záznamoch v atribúte.	Majú byť všetky hodnoty atribútu unikátne v rámci atribútu? POZOR: metadáta nemusia obsahovať informácie o všetkých pravidlách unikátnosti z pohľadu biznisu. Unikátnosť je niekedy ošetrená v samotných vstupoch do	Ak existuje definovaná unikátnosť hodnoty, tak sa môže merať KPI 5.2a a 5.2b	Áno

Tento projekt je podporený z Európskeho sociálneho fondu.

13

Ukazovateľ	Typ merania	Počítané v Talende	KPI ID	KPI	KPI popis	KPI otázka (platí pre atribút alebo skupinu atribútov, odpoveďou je biznis pravidlo)	KPI podmienka merania	KPI meranie
			5.2b	<p>Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje rovnakú hodnotu v rámci atribútu viac ako jedenkrát. (Jedná sa o multiplicitu hodnoty = duplicita, triplicita a pod.)</p> <p>Pozn.: Percentuálny podiel distinktných záznamov v atribúte tabuľky sa vyráta ako súčet unikátnych a multiplicitných záznamov.</p>		databázy, napríklad cez programové skripty.		Áno
Unikátnosť záznamov	Pokročilejšia analýza	Áno	5.3a	<p>Percentuálny podiel záznamov za skupinu atribútov, ktorý obsahuje rovnakú kombináciu hodnôt v rámci skupiny atribútov len jedenkrát.</p> <p>Pozn.: Percentuálny podiel distinktných záznamov za skupinu atribútov sa vyráta ako súčet unikátnych a multiplicitných záznamov.</p>	Za skupinu atribútov, ktoré majú definované biznis pravidlo definujúce jedinečnú kombináciu hodnôt v rámci skupiny atribútov.	Majú byť všetky kombinácie hodnôt skupiny atribútov unikátne v rámci skupiny atribútov? POZOR: metadáta nemusia obsahovať informácie o všetkých pravidlách unikátnosti z pohľadu biznisu. Unikátnosť je niekedy ošetrená v samotných vstupoch do databázy, napríklad cez programové skripty.	Ak existuje informácia o unikátnosti záznamu za skupinu atribútov, tak sa môže merať KPI 5.3a a 5.3b	Áno
			5.3b	<p>Percentuálny podiel záznamov za skupinu atribútov, ktorý obsahuje rovnakú kombináciu hodnôt v rámci skupiny atribútov, viac ako jedenkrát. (Jedná sa o multiplicitu záznamu = duplicita, triplicita a pod.)</p> <p>Pozn.: Percentuálny podiel distinktných záznamov za skupinu atribútov sa vyráta ako súčet unikátnych a multiplicitných záznamov.</p>				Áno
Rýchlosť aktualizácie	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie

Tento projekt je podporený z Európskeho sociálneho fondu.

14

Ukazovateľ	Typ merania	Počítané v Talende	KPI ID	KPI	KPI popis	KPI otázka (platí pre atribút alebo skupinu atribútov, odpoveďou je biznis pravidlo)	KPI podmienka merania	KPI meranie
Aktuálnosť atribútu	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
5★ Open Data stupnica	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Zrozumiteľnosť (interpretovateľnosť)	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Transformovateľnosť	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Kompletnosť referenčného identifikátora	Profiling	Áno	8.1a	Percentuálny podiel záznamov v atribúte tabuľky, ktorý je zároveň referenčným identifikátorom a obsahuje akúkoľvek hodnotu okrem 'null' a prázdnej hodnoty.	Za atribút, ktorý má definované biznis pravidlo definujúce povinne vyplnenú hodnotu a zároveň je označený za referenčný identifikátor.	Je atribút referenčný identifikátor? Je atribút povinný? (Druhá otázka je identická s ukazovateľom 4.2) POZOR: metadáta nemusia obsahovať informácie o všetkých povinných atribútoch z pohľadu biznisu. Povinnosť je niekedy ošetrovaná v samotných vstupoch do databázy, napríklad cez webové formuláre.	Ak existuje potvrdenie, že atribút je referenčný identifikátor a zároveň je povinný, tak sa môže merať KPI 8.1a	Áno
Jedinečnosť referenčného identifikátora	Odborné posúdenie / Pokročilejšia analýza	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Použiteľnosť referencovateľn	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie

Tento projekt je podporený z Európskeho sociálneho fondu.

15

Ukazovateľ	Typ merania	Počítané v Talende	KPI ID	KPI	KPI popis	KPI otázka (platí pre atribút alebo skupinu atribútov, odpoveďou je biznis pravidlo)	KPI podmienka merania	KPI meranie
ého identifikátora								
Stabilita referencovateľného identifikátora	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Formát referencovateľného identifikátora	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Interoperabilita referencovateľného identifikátora	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Využívanie referencovaných objektov	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie
Podiel referencovaných objektov	Odborné posúdenie	Nie	N/A	N/A	N/A	N/A	N/A	Nie

Tabuľka 4: Zoznam KPI a ich použitie v meraní

Tento projekt je podporený z Európskeho sociálneho fondu.

16

4 Prvé merané KPI: ukazovateľ Sledovanie konzistentnosti

Z pohľadu celkového merania dátovej kvality datasetu je presne určené KPI, ktorým je odporúčané začínať. Ide o KPI 2.1a. Je naviazané na ukazovateľ Sledovanie konzistentnosti. Práve výsledok tohto KPI určuje ďalšie smerovanie v meraní a výsledná hodnota je definovaná ako percentuálny podiel atribútov, ktoré majú definované biznis pravidlá. KPI 2.1a je merané odborným posúdením a prakticky analyzuje zoznam biznis pravidiel, ktorý je popísaný v dokumente: Aktuálne biznis pravidlá pre dataset. V kartách merania je zapísaný aktuálny stav spomínaného dokumentu: **Nie sú formálne definované, dokumentované, zdieľané a schválené v plnom rozsahu.** Tento stav reflektuje výsledok merania KPI, ktorý je zapísaný v tabuľke nižšie.

Register	Parameter	Ukazovateľ	KPI 2.1a	Zmeraná hodnota KPI 2.1a
Register úpadcov - RU	Konzistentnosť	Sledovanie konzistentnosti	Percentuálny podiel atribútov, ktoré majú definované biznis pravidlá.	0 %
Register adries - RA				0 %
Register právnických osôb - RPO				0 %

Tabuľka 5: Výsledok prvého meraného KPI: ukazovateľ Sledovanie konzistentnosti (celá vzorka datasetu pre vybrané registre)

Poznámka: Farebné označenie zmeranej hodnoty KPI 2.1a je určené prahovými hodnotami KPI. Odporúčané prahové hodnoty sú nasledovné:

- Červená: pod 50 %
- Oranžová: 50-85 %
- Zelená: nad 85 %

V praxi nie je potrebné aby mali všetky atribúty definované biznis pravidlá. Je potrebné rozlišovať tie atribúty, u ktorých nie je potreba merať dátovú kvalitu. Pokiaľ tieto „pre potreby merania nepotrebné“ atribúty nemajú definované biznis pravidlá, je to akceptovateľný stav. Preto je pri meraní 2.1a potrebné rozlišovať, ktoré atribúty definujú množinu „potrebných na meranie“ a práve z tejto množiny sa má počítať KPI 2.1a.

Hodnotenie prvého merania KPI: ukazovateľ Sledovanie konzistentnosti (celá vzorka datasetu pre vybrané registre)

Hodnota KPI 2.1a 0% je indikátorom toho, že biznis pravidlá pre vybrané registre nie sú formálne definované, dokumentované, zdieľané a schválené. Bez dokumentovaného biznis pravidla je meranie dátovej kvality atribútu tabuľky datasetu nejednoznačné a nepresné.

5 Zmerané KPI pre profiláciu dát

5.1 Výsledky merania dátovej kvality pre KPI v rámci profilácie

V Talende bolo vykonané základné profilovanie dát pre vybrané registre na definovanej vzorke dát podľa kariet merania. Toto meranie zahŕňa výpočty pre:

A. Základnú profiláciu pre stĺpce

- a. Row Count – celkový počet záznamov
- b. Null Count – počet záznamov s hodnotou Null
- c. Distinct Count – počet odlišných hodnôt v záznamoch
- d. Unique Count – počet jedinečných hodnôt v záznamoch
- e. Duplicate Count – počet hodnôt, ktoré sa vyskytujú v záznamoch aspoň 2-krát (tento viacnásobný výskyt sa nazýva multiplicita a atria sem duplicity, triplicity a podobne)
- f. Blank Count – počet záznamov s prázdnu hodnotou (nevyplnenou hodnotou)

V rámci základnej profilácie sa pri výpočte distinct, unique a duplicate count hodnota Null považuje za bežnú hodnotu, tj. nie je vylúčená pri výpočte.

B. Frekvenciu vzoru (Pattern analysis)

C. Frekvencia hodnoty (Value frequency)

Práve tieto hodnoty vykonané základným profilovaním dát boli ďalej využité na výpočet šiestich KPI pre päť ukazovateľov dátovej kvality:

1. **Ukazovateľ dodržanie formátu atribútu: KPI 3.1a** Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje hodnotu v požadovanom formáte. Pri výpočte sa nepočítajú hodnoty Null.
 - ✓ Vypočíta sa ako $(\text{frekvencia vzoru/počet riadkov}) \cdot 100$, aby sa určil správny vzor atribútu je dôležité poznať biznis pravidlo. V tomto prípade sa vyberal formát, ktorý zodpovedal najpravdepodobnejšiemu výskytu vzoru, pretože biznis pravidlo nebolo známe.
2. **Ukazovateľ vyplnenosť povinného údaja: KPI 4.2a** Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje akúkoľvek hodnotu okrem 'null' a prázdnej hodnoty.
 - ✓ Vypočíta sa ako $(\text{Počet riadkov} - \text{Počet Null hodnôt} - \text{Počet prázdnych hodnôt}) / (\text{Počet riadkov}) \cdot 100$. Aby sa určili správne atribúty pre výpočet vyplnenosti povinných údajov, je dôležité poznať biznis pravidlo. V tomto prípade sa vybrali všetky atribúty.

3. **Ukazovateľ vyplnenosť nepovinného údajaja: KPI 4.3a** Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje akúkoľvek hodnotu vrátane prázdnej hodnoty a okrem 'null' hodnoty.
 - ✓ Vypočíta sa ako $(\text{Počet riadkov} - \text{Počet Null hodnôt}) / (\text{Počet riadkov}) * 100$. Aby sa určili správne atribúty pre výpočet vyplnenosti nepovinných údajov, je dôležité poznať biznis pravidlo. V tomto prípade sa vybrali všetky atribúty.
4. **Ukazovateľ unikátnosť hodnoty: KPI 5.2a** Percentuálny podiel záznamov v atribúte tabuľky, ktorý neobsahuje rovnakú hodnotu v rámci atribútu viac ako jeden krát.
 - ✓ Vypočíta sa ako $(\text{Počet unikátnych hodnôt}) / (\text{Počet riadkov}) * 100$. Aby sa určili správne atribúty pre výpočet unikátnosti hodnôt údajov, je dôležité poznať biznis pravidlo. V tomto prípade sa vybrali všetky atribúty.
5. **Ukazovateľ unikátnosť hodnoty: KPI 5.2b** Percentuálny podiel záznamov v atribúte tabuľky, ktorý obsahuje rovnakú hodnotu v rámci atribútu viac ako jeden krát.
 - ✓ Vypočíta sa ako $(\text{Počet multiplicitných hodnôt}) / (\text{Počet riadkov}) * 100$. Aby sa určili správne atribúty pre výpočet multiplicity hodnoty údajov, je dôležité poznať biznis pravidlo. V tomto prípade sa vybrali všetky atribúty.
6. **Ukazovateľ Kompletnosť referenčného identifikátora: KPI 8.1a** Percentuálny podiel záznamov v atribúte tabuľky, ktorý je zároveň referenčným identifikátorom a obsahuje akúkoľvek hodnotu aj 'null' okrem prázdnej hodnoty.
 - ✓ Vypočíta sa ako $(\text{Počet unikátnych hodnôt}) / (\text{Počet riadkov}) * 100$ pre vybraný identifikátor. Aby sa určili správny identifikátor pre výpočet kompletnosti referenčného identifikátora, je dôležité poznať biznis pravidlo. V tomto prípade sa vybrali identifikátori, ktoré zodpovedali a boli považované za identifikátory tabuliek.

Pozn.: počet odlišných (distinct) hodnôt sa rovná súčtu unikátnych a multiplicitných (duplicate) hodnôt.

Výsledky merania dátovej kvality pre KPI v rámci profilácie sa nachádzajú v prílohe: Výsledky merania dátovej kvality pre KPI v rámci profilácie.xlsx

5.2 Zhrnutie merania dátovej kvality pre KPI v rámci profilácie

Meranie dátovej kvality pre KPI v rámci profilácie je najjednoduchšia forma merania dátovej kvality. Napriek tomu je v tomto meraní veľmi dôležitá znalosť biznis pravidiel, ktoré určujú aplikovanie KPI meraní na jednotlivé atribúty tabuliek datasetu. Keďže faktom je absencia týchto pravidiel, bol zvolený nasledovný postup:

- **Odhad formátu hodnôt v atribúte:** podľa najčastejšieho výskytu formátu v atribúte bol zvolený očakávaný a teda aj najpravdepodobnejší(e) formát(y) hodnôt.
- **Povinnosť vyplnenia:** všetky atribúty tabuliek z vybranej vzorky datasetu boli označené za povinne vyplnené.
- **Nepovinnosť vyplnenia:** všetky atribúty tabuliek z vybranej vzorky datasetu boli označené za nepovinne vyplnené.
- **Unikátnosť hodnôt:** hodnoty všetkých atribútov tabuliek z vybranej vzorky datasetu boli označené za povinne unikátne.
- **Multiplicita hodnôt:** hodnoty všetkých atribútov tabuliek z vybranej vzorky datasetu boli označené za povinne multiplicitné (hodnoty sa majú nachádzať v atribúte viac ako jedenkrát).
- **Kompletnosť referenčného identifikátora:** na základe názvoslovnia atribútov bol realizovaný odhad určenia referenčných identifikátorov.

Pozn.: V rámci profilácie dát (pre RA) boli vybrané atribúty, pre ktoré malo zmysel merať jednotlivé KPI. Toto určenie bolo komunikované s predstaviteľmi RA.

Dôsledkom určeného postupu je vznik značnej rezervy v presnosti merania dátovej kvality pre vybrané registre v rámci profilácie. Nie každý atribút potrebuje mať definovaný formát, povinnosť vyplnenia, unikátnosť hodnoty, multiplicitu hodnoty alebo určenie referenčného identifikátora. Biznis pravidlá pomáhajú určiť kritéria dátovej kvality pre jednotlivé atribúty. Mali by byť teda formálne definované, dokumentované, zdieľané a schválené. Na základe biznis pravidiel je potom možné určiť na ktorých atribútoch je potrebné merať konkrétne KPI. Práve definované biznis pravidlá viditeľne zvyšujú kvalitu každého merania dátovej kvality.

6 KPI merania dátovej kvality so znalosťou biznis pravidiel

6.1 Vybrané biznis pravidlá pre vybrané registre

Vo verejnej správe neexistuje formálna, validovaná a schválená dokumentácia biznisových pravidiel, ktorá by určovala kritéria merania dátovej kvality. Preto vznikla iniciatíva definovania vybraných biznis pravidiel, aby bolo možné zmerať dátovú kvalitu v maximálnej možnej miere. Vzhľadom na náročnosť úlohy nebolo možné definovať všetky biznis pravidlá. Táto aktivita vyžaduje nielen diskusiu, ktorá potvrdí ich správnosť, ale aj existenciu aktuálnych dokumentácií pre dátové modely, dátové štandardy a metadata (dáta o dátach). Veľké množstvo biznis pravidiel má obsahovať aj informácie o vzťahoch medzi jednotlivými atribútmi a na získanie tejto informácie je potrebné poznať architektúru informačných systémov nielen referenčných registrov, ale aj samotných zdrojových systémov.

Z aktivity neformálneho získania biznis pravidiel vznikol zoznam 25 pravidiel, ktoré je potrebné zmerať a určiť základný pohľad na dátovú kvalitu vybraných registrov. Tieto biznis pravidla sú uvedené v tabuľke nižšie.

Tabuľka obsahuje informácie o registri, ku ktorému sa vzťahuje. Informácie dopĺňujú ID biznis pravidla a samotný názov, respektíve popis biznis pravidla. Každé biznis pravidlo identifikuje množina zdrojových tabuliek alebo atribútov, s ktorými biznis pravidlo pracuje. Cieľová tabuľka alebo atribút v tabuľke lokalizuje dátovú oblasť v datasete, na ktorú sa zameriavajú relevantné KPI. Práve tieto KPI je neskôr potrebné zmerať.

Register	ID biznis pravidla	Biznis pravidlo	Zdroj – tabuľka(y) v registri, atribút(y) v tabuľke	Cieľová tabuľka / atribút v tabuľke	Poznámka
Register úpadcov - RU	RU-1	Adresa súdu musí byť spárovaná s RA.	Core.sud	Core.sud	Neexistujú identifikátory na referenčný register adres
	RU-2	Právnická osoba ako dlžník, pridelená ku konaniu, má byť riadne stotožnená s RPO.	Core.osoba(dtype = pravnicka osoba)	Core.osoba	Ak je dtype právnická osoba, tak IPO nesmie byť NULL
	RU-3	Fyzická osoba ako dlžník, pridelená ku konaniu, má byť riadne stotožnená s RFO.	Core.osoba(dtype=fyzicka osoba)	Core.osoba	Ak je dtype fyzická osoba, tak IFO nesmie byť NULL
	RU-4	Každé aktívne konanie (stav začaty_proces_konania) má mať prideleného len aktívneho spravcu.	Core.spravca(ID spravca); Core.konanie(ID spravca)	Core.konanie	Stav konania = začaty_proces_konania. Stav spravcu = aktívny
	RU-5	Konanie typu konkurz v stave (začaty_proces / skončeny_proces) musí obsahovať aspoň jednu pohľadávku. Toto platí len pre konania typu konkurz, ktoré majú dátum návrhu po 1.7.2016.	Core.konanie (datum_podania >1.7.2016, dtype=konkurz, stav_konania_ru=zacaty_proces_konania or skonceny_proces)	Core.konanie	V databáze by mala byť tabuľka Uspokojenie, ktorá previaže Pohľadávku a Konanie
Register adres - RA	RA-1	Adresné body (zemepisná šírka) sa musia nachádzať v rámci intervalu medzi najzápadnejším/najvýchodnejším bodom Slovenskej republiky.	07_Building_number (AxisB)	07_Building_number	Najzápadnejší bod: stred rieky Morava pri obci Záhorská Ves 48° 22' 50.24" N, 16° 50' 0.74" E Najvýchodnejší bod: vrch Kremenec 49° 5' 16.81" N, 22° 33' 56.7" E
	RA-2	Zemepisná šírka je povinným atribútom	07_Building_number (AxisB)	07_Building_number	GPS súradnice sú povinnými údajmi
	RA-3	Adresné body (zemepisná dĺžka) sa musia nachádzať v rámci intervalu	07_Building_number (AxisA)	07_Building_number	Najsevernejší bod: pri vrchu Beskydok

Tento projekt je podporený z Európskeho sociálneho fondu.

22

Register	ID biznis pravidla	Biznis pravidlo	Zdroj – tabuľka(y) v registri, atribút(y) v tabuľke	Cieľová tabuľka / atribút v tabuľke	Poznámka
		medzi najsevernejším/najjužnejším bodom Slovenskej republiky.			49° 36' 49.43" N, 19° 28' 2.52" E Najjužnejší bod: Patince 47° 44' 1.29" N, 18° 17' 20.93" E
	RA-4	Zemepisná dĺžka je povinným atribútom.	07_Building_number (AxisA)	07_Building_number	GPS súradnice sú povinnými údajmi
	RA-5	Obec musí obsahovať minimálne 2 typy budov.	06_Property_registration_number (BuildingTypeCode_ItemName); 03_Municipality	06_Property_registration_number	Pozor na budovy typu "iná budova"
	RA-6	Typ budovy je povinný údaj.	06_Property_registration_number (BuildingTypeCode_ItemName); 03_Municipality	06_Property_registration_number	Typ budovy je povinný údaj
	RA-7	Orientačné číslo nesmie obsahovať 0, prázdnu hodnotu. V prípade, že má hodnotu "NULL", tak budova obsahuje tzv. virtuálny vchod.	06_Property_registration_number; 07_Building_number	06_Property_registration_number	Orientačné číslo je vchod
	RA-8	Súpisné číslo nesmie obsahovať 0, prázdnu hodnotu ani "NULL".	06_Property_registration_number; 07_Building_number	06_Property_registration_number	Súpisné číslo je budova
	RA-9	Nesmie existovať kombinácia súpisné číslo + ulica (musí byť len orientačné číslo + ulica).	05_Street_name; 06_Property_registration_number	N/A	Neexistuje identifikátor na ulice
Register právnických osôb - RPO	RPO-1	IČO musí byť unikátne v rámci jedného zdrojového systému - obchodný register.	Názvoslovie xlm súborov dodaných od zástupcov RPO.	Názov xml súboru	Unikátne v názvoch XML súborov, keďže každý XML súbor patrí k jednej právnickej osobe - hlavnej entite. Tento postup bol zvolený za predpokladu že IČO v názve XML je identické s informáciou o IČO v samotnom XML. Pre

Tento projekt je podporený z Európskeho sociálneho fondu.

23

Register	ID biznis pravidla	Biznis pravidlo	Zdroj – tabuľka(y) v registri, atribút(y) v tabuľke	Cieľová tabuľka / atribút v tabuľke	Poznámka
					komplikácie s parsovaním XML bol zvolený tento postup
	RPO-2	IČO musí byť unikátne v rámci jedného zdrojového systému - živnostenský register.	Názvoslovie xlm súborov dodaných od zástupcov RPO.	Názov xml súboru	Unikátne v názvoch XML súborov, keďže každý XML súbor patrí k jednej právnickej osobe - hlavnej entite. Tento postup bol zvolený za predpokladu že IČO v názve XML je identické s informáciou o IČO v samotnom XML. Pre komplikácie s parsovaním XML bol zvolený tento postup
	RPO-3	IČO je v správnom tvare a formáte v rámci jedného zdrojového systému - obchodný register.	Názvoslovie xlm súborov dodaných od zástupcov RPO.	Názov xml súboru	Povolený formát je: 99999999 v názvoch XML súborov, keďže každý XML súbor patrí k jednej právnickej osobe - hlavnej entite. Tento postup bol zvolený za predpokladu že IČO v názve XML je identické s informáciou o IČO v samotnom XML. Pre komplikácie s parsovaním XML bol zvolený tento postup
	RPO-4	IČO je v správnom tvare a formáte v rámci jedného zdrojového systému - živnostenský register.	Názvoslovie xlm súborov dodaných od zástupcov RPO.	Názov xml súboru	Povolený formát je: 99999999 v názvoch XML súborov, keďže každý XML súbor patrí k jednej právnickej osobe - hlavnej entite. Tento postup bol zvolený za predpokladu že IČO v názve XML je identické s informáciou o IČO v samotnom XML. Pre

Tento projekt je podporený z Európskeho sociálneho fondu.

24

Register	ID biznis pravidla	Biznis pravidlo	Zdroj – tabuľka(y) v registri, atribút(y) v tabuľke	Cieľová tabuľka / atribút v tabuľke	Poznámka
					komplikácie s parsovaním XML bol zvolený tento postup
	RPO-5	IČO musí byť vyplnené v rámci jedného zdrojového systému - obchodný register.	Názvoslovie xlm súborov dodaných od zástupcov RPO.	Názov xml súboru	V názvosloviu pre časť IČO je hodnota "neuvedene" chápaná ako "NULL" alebo "prázdna hodnota"
	RPO-6	IČO musí byť vyplnené v rámci jedného zdrojového systému - živnostenský register.	Názvoslovie xlm súborov dodaných od zástupcov RPO.	Názov xml súboru	V názvosloviu pre časť IČO je hodnota "neuvedene" chápaná ako "NULL" alebo "prázdna hodnota"
	RPO-7	Každý záznam obsahuje jedinečný identifikátor IPO.	rpo.Subject_Address, rpo.Subject_Detail, rpo_Subject_LegalForms, rpo.Subjects	rpo.Subject_Address (atribút [IPO]), rpo.Subject_Detail (atribút [IPO]), rpo_Subject_LegalForms (atribút [IPO]), rpo.Subjects (atribút [IPO])	Identifikátor musí byť jedinečný
	RPO-8	Každý záznam obsahuje informáciu z akého je zdrojového systému.	rpo.Subject_Address, rpo.Subject_Detail, rpo.Subject_Has_Activities, rpo.Subjects	rpo.Subject_Address (atribút [SourceSystem]), rpo.Subject_Detail (atribút [SourceSystem]), rpo.Subject_Has_Activities (atribút [SourceSystem]), rpo.Subjects (atribút [SourceSystem])	Platí pre všetky tabuľky
	RPO-9	PSČ je v správnom tvare a formáte.	rpo.Subject_Address	rpo.Subject_Address (atribút [Address_PostalCode])	Povolené formáty sú: 99999 a 999 99
	RPO-10	Rodné číslo pre štatutára právnickej osoby pre zdrojový systém (obchodný register) musí byť vyplnené.	rpo.Statutory_Bodies	rpo.Statutory_Bodies	Jedna právnická osoba môže obsahovať viac štatutárov. Jeden štatutár môže obsahovať len jedno rodné číslo.

Tento projekt je podporený z Európskeho sociálneho fondu.

25

Register	ID biznis pravidla	Biznis pravidlo	Zdroj – tabuľka(y) v registri, atribút(y) v tabuľke	Cieľová tabuľka / atribút v tabuľke	Poznámka
	RPO-11	IFO pre štatutára právnickej osoby pre zdrojový systém (obchodný register) musí byť vyplnené.	rpo.Statutory_Bodies	rpo.Statutory_Bodies	Jedna právnická osoba môže obsahovať viac štatutárov. Jeden štatutár môže obsahovať len jednu IFO hodnotu.

Tabuľka 6: Vybrané biznis pravidlá pre vybrané registre

Tento projekt je podporený z Európskeho sociálneho fondu.

26

6.2 KPI otázky a podmienky merania pre vybrané biznis pravidlá

V predošlej kapitole bol identifikovaný neformálny zoznam biznis pravidiel, ktoré je potrebné zmerať. Merané KPI za jednotlivé biznis pravidlá je možné určiť pomocou KPI otázky. Zoznam všetkých definovaných KPI otázok je uvedený v zozname (**Tabuľka 4: Zoznam KPI a ich použitie v meraní**), kde sú jednotlivé KPI napárované na ukazovatele merania dátovej kvality. Pokiaľ je splnená KPI podmienka merania v jednotlivých odpovediach na KPI otázky, automaticky to znamená možnosť merania relevantného KPI. Zoznam splnených KPI podmienok, napárovaný na jednotlivé biznis pravidlá je zobrazený v tabuľke nižšie. Platí pravidlo, kde 1 biznis pravidlo odpovedá vždy len na 1 KPI otázku, tzn. 1 biznis pravidlo = 1 KPI.

ID biznis pravidla	ID KPI	— KPI otázka (platí pre atribút alebo skupinu atribútov, odpoveďou je biznis pravidlo)	KPI podmienka merania
RU-1	N/A	N/A	N/A
RU-2	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RU-3	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RU-4	2.2b	Aká je procesná stránka atribútu resp. aké sú vzťahy medzi atribútom a relevantnými atribútmi?	Ak existujú definované väzby medzi atribútmi, tak sa môže merať KPI 2.2b
RU-5	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RA-1	1.2a	Aké sú presne očakávané hodnoty atribútu?	Ak existujú definované očakávané hodnoty, tak sa môže merať KPI 1.1a alebo 1.2a
RA-2	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RA-3	1.2a	Aké sú presne očakávané hodnoty atribútu?	Ak existujú definované očakávané hodnoty, tak sa môže merať KPI 1.1a alebo 1.2a
RA-4	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RA-5	2.2b	Aká je procesná stránka atribútu resp. aké sú vzťahy medzi atribútom a relevantnými atribútmi?	Ak existujú definované väzby medzi atribútmi, tak sa môže merať KPI 2.2b
RA-6	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RA-7	1.2a	Aké sú presne očakávané hodnoty atribútu?	Ak existujú definované očakávané hodnoty, tak sa môže merať KPI 1.1a alebo 1.2a

Tento projekt je podporený z Európskeho sociálneho fondu.

27

ID biznis pravidla	ID KPI	— KPI otázka (platí pre atribút alebo skupinu atribútov, odpoveďou je biznis pravidlo)	KPI podmienka merania
RA-8	1.2a	Aké sú presne očakávané hodnoty atribútu?	Ak existujú definované očakávané hodnoty, tak sa môže merať KPI 1.1a alebo 1.2a
RA-9	N/A	N/A	N/A
RPO-1	5.2a	Majú byť všetky hodnoty atribútu unikátne v rámci atribútu?	Ak existuje definovaná unikátnosť hodnoty, tak sa môže merať KPI 5.2a a 5.2b
RPO-2	5.2a	Majú byť všetky hodnoty atribútu unikátne v rámci atribútu?	Ak existuje definovaná unikátnosť hodnoty, tak sa môže merať KPI 5.2a a 5.2b
RPO-3	3.1a	Aký je formát hodnoty v atribúte?	Ak existuje definovaný formát, tak sa môže merať KPI 3.1a
RPO-4	3.1a	Aký je formát hodnoty v atribúte?	Ak existuje definovaný formát, tak sa môže merať KPI 3.1a
RPO-5	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RPO-6	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RPO-7	5.2a	Majú byť všetky hodnoty atribútu unikátne v rámci atribútu?	Ak existuje definovaná unikátnosť hodnoty, tak sa môže merať KPI 5.2a a 5.2b
RPO-8	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RPO-9	3.1a	Aký je formát hodnoty v atribúte?	Ak existuje definovaný formát, tak sa môže merať KPI 3.1a
RPO-10	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a
RPO-11	4.2a	Je atribút povinný?	Ak existuje definovaná povinnosť vyplnenia, tak sa môže merať KPI 4.2a alebo 4.3a

Tabuľka 7: KPI otázky a odpovede pre definované biznis pravidlá

Poznámka: Vyplnená hodnota "N/A" indikuje, že biznis pravidlo nedokáže splniť KPI podmienku pre žiadne definované KPI. Fialové podfarbenie v tabuľke indikuje potrebu pokročilejšej analýzy v nástroji merania dátovej kvality. Oranžové podfarbenie indikuje potrebu profilácie v nástroji merania dátovej kvality.

6.3 Výsledok merania KPI ukazovateľov pre meranie dátovej kvality vybraných biznis pravidiel pre vybrané registre

Na základe informácií z predošlej kapitoly boli identifikované kľúčové KPI otázky, ktoré spĺňajú KPI podmienky vo vybraných biznis pravidlách. Napáovanie KPI otázok na relevantné KPI určilo jednotlivé KPI merania, ktoré bolo potrebné vykonať. Výsledky z tohto merania sú zobrazené v tabuľke nižšie. Každé biznis pravidlo môže obsahovať informácie na splnenie KPI podmienok aj pre viacero KPI. V takom prípade je k biznis pravidlu pridelený KPI výsledok viac ako jedenkrát.

ID biznis pravidla	KPI - 1.2a [%]	KPI - 1.3a [%]	KPI - 2.2b [%]	KPI - 3.1a [%]	KPI - 4.2a [%]	KPI - 4.3a [%]	KPI - 5.2a [%]	KPI - 5.2b [%]	KPI - 5.3a [%]	KPI - 5.3b [%]	KPI - 8.1a [%]
RU-1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RU-2	N/A	N/A	N/A	N/A	61,94	N/A	N/A	N/A	N/A	N/A	N/A
RU-3	N/A	N/A	N/A	N/A	50,60	N/A	N/A	N/A	N/A	N/A	N/A
RU-4	N/A	N/A	96,43	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RU-5	N/A	N/A	N/A	N/A	88,24	N/A	N/A	N/A	N/A	N/A	N/A
RA-1	99,99	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RA-2	N/A	N/A	N/A	N/A	73,48	N/A	N/A	N/A	N/A	N/A	N/A
RA-3	99,99	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RA-4	N/A	N/A	N/A	N/A	73,48	N/A	N/A	N/A	N/A	N/A	N/A
RA-5	N/A	N/A	99,83	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RA-6	N/A	N/A	N/A	N/A	99,73	N/A	N/A	N/A	N/A	N/A	N/A
RA-7	99,11	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RA-8	99,82	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RA-9	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RPO-1	N/A	N/A	N/A	N/A	N/A	N/A	99,45	N/A	N/A	N/A	N/A
RPO-2	N/A	N/A	N/A	N/A	N/A	N/A	99,99	N/A	N/A	N/A	N/A
RPO-3	N/A	N/A	N/A	100	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RPO-4	N/A	N/A	N/A	100	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RPO-5	N/A	N/A	N/A	N/A	99,47	N/A	N/A	N/A	N/A	N/A	N/A
RPO-6	N/A	N/A	N/A	N/A	99,99	N/A	N/A	N/A	N/A	N/A	N/A
RPO-7	N/A	N/A	N/A	N/A	N/A	N/A	99,93	N/A	N/A	N/A	N/A
RPO-8	N/A	N/A	N/A	N/A	100	N/A	N/A	N/A	N/A	N/A	N/A
RPO-9	N/A	N/A	N/A	97,73	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Tento projekt je podporený z Európskeho sociálneho fondu.

29

ID biznis pravidla	KPI - 1.2a [%]	KPI - 1.3a [%]	KPI - 2.2b [%]	KPI - 3.1a [%]	KPI - 4.2a [%]	KPI - 4.3a [%]	KPI - 5.2a [%]	KPI - 5.2b [%]	KPI - 5.3a [%]	KPI - 5.3b [%]	KPI - 8.1a [%]
RPO-10	N/A	N/A	N/A	N/A	89,55	N/A	N/A	N/A	N/A	N/A	N/A
RPO-11	N/A	N/A	N/A	N/A	41,49	N/A	N/A	N/A	N/A	N/A	N/A

Tabuľka 8: Výsledok merania KPI ukazovateľov pre meranie dátovej kvality vybraných biznis pravidiel

Poznámka: Vyplnená hodnota "N/A" indikuje, že biznis pravidlo nedokáže splniť KPI podmienku pre žiadne z definovaných KPI. Teda KPI meranie v tomto prípade nebolo možné vykonať. Fialové podfarbenie v tabuľke indikuje potrebu pokročilejšej analýzy v nástroji merania dátovej kvality. Oranžové podfarbenie indikuje potrebu profilácie v nástroji merania dátovej kvality.

6.4 Zhodnotenie merania pre vybrané biznis pravidlá

Po určení vybraných biznis pravidiel v rámci vybraných registrov, boli identifikované KPI otázky, ktoré spĺňali KPI podmienku merania. Nájdenie KPI otázok zároveň umožnilo identifikáciu relevantného KPI, ktoré bolo možné zmerať. Z výsledkov jednotlivých meraní, pridelených k vybraným biznis pravidlám, vyplývajú nasledovné závery:

- **Biznis pravidlo RU-1:** Adresa súdu musí byť spárovaná s RA.
 - **Záver RU-1:** Za spôsob merania bolo určené odborné posúdenie dátového modelu, pretože meranie pomocou nástroja merania nebolo možné vykonať. V tabuľke Core.sud sa nenachádza identifikátor na referenčný register adries.
- **Biznis pravidlo RU-2:** Právnická osoba ako dlžník, pridelená ku konaniu, má byť riadne stotožnená s RPO.
 - **Záver RU-2:** Všetkých právnických osôb je 18718. Z týchto záznamov má vyplnený identifikátor právnickej osoby (IPO) 11594 osôb, to je 61,94 % z celkového počtu.
- **Biznis pravidlo RU-3:** Fyzická osoba ako dlžník, pridelená ku konaniu, má byť riadne stotožnená s RFO.
 - **Záver RU-3:** Všetkých fyzických osôb (vrátane podnikateľov) je 20283. Žiadna z týchto osôb zatiaľ nemá priradený identifikátor fyzickej osoby (IFO). V súčasnosti IFO slúži ako rodné číslo. S evidovaným rodným číslom je 10264 osôb, to je 50,60 % všetkých záznamov.
- **Biznis pravidlo RU-4:** Každé aktívne konanie (stav začaty_proces_konania) má mať prideleného len aktívneho správcu.
 - **Záver RU-4:** Aktívnych konaní alebo začatých (a neskončených) procesov konania je 5168. Z týchto konaní nemá prideleného správcu alebo

Tento projekt je podporený z Európskeho sociálneho fondu.

30

pridelený správca nie je aktívny 185 konaní, to je 3,57 %. 96,43 % konaní teda spĺňa biznis pravidlo.

- **Biznis pravidlo RU-5:** Konanie typu konkurz v stave (začaty_proces / skončený_proces) musí obsahovať aspoň jednu pohľadávku. Toto platí len pre konania typu konkurz, ktoré majú dátum návrhu po 1.7.2016.
 - **Záver RU-5:** Začatých, alebo skončených konkurzov v období od 1.7.2016 vrátane je 2469. Tabuľka 'konanie' bola previazaná s tabuľkou 'uspokojenie', ktorá odkazuje na pohľadávky k daným konaniam. Zo všetkých konkurzov s danými parametrami nemá priradenú pohľadávku 315 konkurzov, to je 12,76 %. 88,24 % konaní typu konkurz teda spĺňa biznis pravidlo.
- **Biznis pravidlo RA-1:** Adresné body (zemepisná šírka) sa musia nachádzať v rámci intervalu medzi najzápadnejším/najvýchodnejším bodom Slovenskej republiky.
 - **Záver RA-1:** Máme 1 605 270 záznamov, z ktorých adresné body pokrývajú 1 179 601 záznamov. V prípade adresného bodu "Zemepisná šírka" je 76 záznamov mimo interval (0,006%), pričom 52 záznamov má pravdepodobne len vymenené hodnoty so zemepisnou dĺžkou (0,004%).
- **Biznis pravidlo RA-2:** Zemepisná šírka je povinným atribútom.
 - **Záver RA-2:** Z 1 605 270 záznamov zemepisnej šírky 425 669 obsahuje hodnotu NULL (26,52%).
- **Biznis pravidlo RA-3:** Adresné body (zemepisná dĺžka) sa musia nachádzať v rámci intervalu medzi najsevernejším/najjužnejším bodom Slovenskej republiky.
 - **Záver RA-3:** Máme 1 605 270 záznamov, z ktorých adresné body pokrývajú 1 179 601 záznamov. V prípade adresného bodu "Zemepisná dĺžka" je 62 záznamov mimo interval, čo je (0,005%), pričom 52 záznamov má pravdepodobne len vymenené hodnoty so zemepisnou šírkou (0,004%).
- **Biznis pravidlo RA-4:** Zemepisná dĺžka je povinným atribútom.
 - **Záver RA-4:** Z 1 605 270 záznamov zemepisná dĺžky 425 669 záznamov obsahuje hodnotu NULL (26,52%).
- **Biznis pravidlo RA-5:** Obec musí obsahovať minimálne 2 typy budov.
 - **Záver RA-5:** "Pri spájaní tabuliek nedošlo k 100% spojeniu, pretože v tabuľke 06_Property_registration_number atribút municipalityIdentifier (FK) obsahoval aj hodnoty väčšie ako 3032 a menšie ako 92, ktoré su mimo rozsah hodnôt v atribúte objectId (PK) v tabuľke 03_Municipality.
Z celkového počtu 2929 obcí bolo identifikovaných 5 obcí, ktoré majú len 1 typ budovy (väčšinou ""iná budova""). Medzi týchto 5 obcí patria: Juskova Voľa, Nižný Slavkov, Ruská Voľa, Ondavské Matiašovce, Javorina.

Až 99,83% obcí obsahuje minimálne 2 typy budov. Pravdepodobne ďalších 59 obcí, ktoré majú len 2 typy budov (väčšinou ""iná budova""), by mali mať preverenú presnosť resp. vierohodnosť určenia typov budov."

- **Biznis pravidlo RA-6:** Typ budovy je povinný údaj.
 - **Záver RA-6:** Celkový počet záznamov (pridelení budov) je 1 564 570. Počet záznamov s NULL pre typ budovy je 4295. Tento atribút nesmie obsahovať NULL, ani prázdnu hodnotu.
- **Biznis pravidlo RA-7:** Orientačné číslo nesmie obsahovať 0, prázdnu hodnotu. V prípade, že má hodnotu "NULL", tak budova obsahuje tzv. virtuálny vchod.
 - **Záver RA-7:** Existuje 1 605 270 záznamov orientačných čísiel a z toho 14 357 teda 0,89% nespĺňa biznis pravidlo, ktoré hovorí, že orientačné číslo nesmie byť 0, prázdna hodnota. Záznamy obsahujú 14 357 záznamov s hodnotou 0 a neobsahujú záznamy s prázdnu hodnotou. Existuje 919 365 virtuálnych vchodov.
- **Biznis pravidlo RA-8:** Súpisné číslo nesmie obsahovať 0, prázdnu hodnotu ani "NULL".
 - **Záver RA-8:** Existuje 1 564 570 záznamov súpisných čísiel a z toho 2 753 teda 0,18% nespĺňa biznis pravidlo, ktoré hovorí, že súpisné číslo nesmie byť 0, prázdna hodnota a ani NULL. Záznamy obsahujú 90 záznamov s hodnotou 0 a 2 663 záznamov s hodnotou NULL. Záznamy neobsahujú žiadnu prázdnu hodnotu.
- **Biznis pravidlo RA-9:** Nesmie existovať kombinácia súpisné číslo + ulica (musí byť len orientačné číslo + ulica).
 - **Záver RA-9:** Za spôsob merania bolo určené odborné posúdenie dátového modelu, pretože meranie pomocou nástroja merania nebolo možné vykonať. V tabuľke Budovy (Súpisné číslo) sa nenachádza identifikátor ulice.
- **Biznis pravidlo RPO-1:** IČO musí byť unikátne v rámci jedného zdrojového systému - obchodný register.
 - **Záver RPO-1:** Všetkých evidovaných subjektov je 302 538. V rámci všetkých záznamov bolo nájdených 300 881 unikátnych (neboli zarátane hodnoty „neuvedene“). To predstavuje unikátnosť na úrovni 99,45 percent.
- **Biznis pravidlo RPO-2:** IČO musí byť unikátne v rámci jedného zdrojového systému - živnostenský register.
 - **Záver RPO-2:** Všetkých evidovaných subjektov je 407 212. V rámci všetkých záznamov bolo nájdených 407 164 unikátnych (neboli zarátane hodnoty „neuvedene“). To predstavuje unikátnosť na úrovni 99,99 percent.
- **Biznis pravidlo RPO-3:** IČO je v správnom tvare a formáte v rámci jedného zdrojového systému - obchodný register.
 - **Záver RPO-3:** Všetkých evidovaných subjektov je 302 538. Z toho má 300 924 záznamov uvedenú hodnotu IČO. Neuvedená hodnota sa považuje IČO s hodnotou "Neuvedene". Jediným správnym formátom

Tento projekt je podporený z Európskeho sociálneho fondu.

32

- hodnoty atribútu je 8-miestne celé kladné číslo. Túto požiadavku spĺňa 300 924 záznamov z počtu 300 924 záznamov. To predstavuje správny formát na úrovni 100 percent.
- **Biznis pravidlo RPO-4:** IČO je v správnom tvare a formáte v rámci jedného zdrojového systému - živnostenský register.
 - **Záver RPO-4:** Všetkých evidovaných subjektov je 407 212. Z toho má 407 168 záznamov uvedenú hodnotu IČO. Neuvedená hodnota sa považuje IČO s hodnotou "Neuvedene". Jediným správnym formátom hodnoty atribútu je 8-miestne celé kladné číslo. Túto požiadavku spĺňa 407 168 záznamov z počtu 407 168 záznamov. To predstavuje správny formát na úrovni 100 percent.
 - **Biznis pravidlo RPO-5:** IČO musí byť vyplnené v rámci jedného zdrojového systému - obchodný register.
 - **Záver RPO-5:** Všetkých evidovaných subjektov je 302 538. Z toho má 1614 záznamov neuvedenú hodnotu IČO a teda 300 924 má uvedenú hodnotu. Neuvedená hodnota sa považuje IČO s hodnotou "Neuvedene". To predstavuje vyplnenosť povinného údaja na úrovni 99,47 percent..
 - **Biznis pravidlo RPO-6:** IČO musí byť vyplnené v rámci jedného zdrojového systému - živnostenský register.
 - **Záver RPO-6:** Všetkých evidovaných subjektov je 407 212. Z toho má 44 záznamov neuvedenú hodnotu IČO a teda 407 168 má uvedenú hodnotu. Neuvedená hodnota sa považuje IČO s hodnotou "Neuvedene". To predstavuje vyplnenosť povinného údaja na úrovni 99,99 percent..
 - **Biznis pravidlo RPO-7:** Každý záznam obsahuje jedinečný identifikátor IPO.
 - **Záver RPO-7:** Každá z týchto tabuliek (rpo.Subject_Address, rpo.Subject_Detail, rpo_Subject_LegalForms, rpo.Subjects) obsahuje 104 200 záznamov. Tabuľka rpo.Subject_Has_Activities obsahuje 1 272 280 záznamov. Jedinečnosť v rámci jednotlivých tabuliek:
 - rpo.Subject_Address obsahuje 104,128 unikátnych IPO hodnôt, to je 99,93 %,
 - rpo.Subject_Detail obsahuje 104,128 unikátnych IPO hodnôt, to je 99,93 %,
 - rpo_Subject_LegalForms obsahuje 104,128 unikátnych IPO hodnôt, to je 99,93 %,
 - rpo.Subjects obsahuje 104,128 unikátnych hodnôt, to je 99,93 %.Tieto 4 tabuľky obsahujú identické hodnoty atribútu 'IPO'.
Finálna hodnota je priemerom výsledkov.
 - **Biznis pravidlo RPO-8:** Každý záznam obsahuje informáciu z akého je zdrojového systému.
 - **Záver RPO-8:** Všetky záznamy vo v špecifikovaných 4 tabuľkách obsahujú správnu informáciu, z akého zdrojového systému pochádzajú. Drvivá

väčšina (viac ako 95 %) pochádza z Obchodného registra. Druhé najväčšie zastúpenie má Živnostenský register.

- **Biznis pravidlo RPO-9:** PSČ je v správnom tvare a formáte.
 - **Záver RPO-9:** Všetkých evidovaných záznamov je 104 200. PSČ musí byť 5-ciferné celé kladné číslo. Akceptované sú aj formáty, kde je posledné dvojčíslenie oddelené medzerou, teda: XXXXX alebo XXX XX. Tento formát nespĺňa 2363 záznamov, to je 2,27 %. Najčastejšími chybami boli: 6-ciferné číslo, 4-ciferné číslo, prítomnosť interpunkčných znamienok v čísle.
- **Biznis pravidlo RPO-10:** Rodné číslo pre štatutára právnickej osoby pre zdrojový systém (obchodný register) musí byť vyplnené.
 - **Záver RPO-10:** Všetkých evidovaných štatutárov je 410 472 (100%) v 302 538 právnických osobách z obchodného registra.
Rodné číslo má uvedených 367 589 štatutárov (89,55%) (kde sa nenachádza hodnota NULL, prázdna hodnota a ani hodnota 'Neuvedené').
Hodnota NULL je uvedená u 36 586 štatutárov. Prázdna hodnota nie je uvedená u žiadneho štatutára. Hodnota 'Neuvedené' je uvedená u 6297 štatutárov.
- **Biznis pravidlo RPO-11:** IFO pre štatutára právnickej osoby pre zdrojový systém (obchodný register) musí byť vyplnené.
 - **Záver RPO-11:** Všetkých evidovaných štatutárov je 410 472 (100%) v 302 538 právnických osobách z obchodného registra.
IFO má uvedených 170 315 štatutárov (41,49%) (kde sa nenachádza hodnota NULL, prázdna hodnota a ani hodnota 'Neuvedené').
Hodnota NULL je uvedená u 240 157 štatutárov. Prázdna hodnota a hodnota 'Neuvedené' nie sú uvedené u žiadneho štatutára.

7 **Finálne merané KPI: ukazovateľ Dodržiavanie biznis pravidla**

Z pohľadu celkového merania dátovej kvality je presne určené KPI, ktorým je odporúčané zakončiť meranie celého objektu merania. Ide o KPI 2.2a a je naviazané na ukazovateľ Dodržiavanie biznis pravidla. Práve výsledok tohto KPI vytvára sumarizáciu v meraní a výsledná hodnota je definovaná ako percentuálny podiel atribútov, ktorých hodnoty plne dodržiajú definované biznis pravidlá. KPI 2.2a je merané odborným posúdením a prakticky analyzuje výsledky všetkých doterajších meraní. Toto KPI je svojim spôsobom naviazané na KPI 2.1a. Kým v KPI 2.1a je vymedzená množina atribútov, ktoré majú definované biznis pravidlá, v KPI 2.2a je táto množina zosumarizovaná. Výsledok merania KPI je zapísaný v tabuľkách nižšie.

Výsledok finálne meraného KPI: ukazovateľ Dodržiavanie biznis pravidla (celá vzorka datasetu pre vybrané registre)

Register	Parameter	Ukazovateľ	KPI 2.2a	Zmeraná hodnota KPI 2.2a
Register úpadcov - RU	Konzistentnosť	Dodržiavanie biznis pravidla	Percentuálny podiel atribútov, ktorých hodnoty plne dodržiajú definované biznis pravidlá.	0 %
Register adries - RA				0 %
Register právnických osôb - RPO				0 %

Tabuľka 9: Finálne merané KPI: ukazovateľ Dodržiavanie biznis pravidla (celá vzorka datasetu pre vybrané registre)

Poznámka: Farebné označenie zmeranej hodnoty KPI 2.2a je určené prahovými hodnotami KPI. Odporúčané prahové hodnoty sú nasledovné:

- Červená: pod 50 %
- Oranžová: 50-85 %
- Zelená: nad 85 %

Hodnotenie finálne meraného KPI: ukazovateľ Dodržiavanie biznis pravidla (celá vzorka datasetu pre vybrané registre)

KPI 2.2a sa vzťahuje k výsledku KPI 2.1a. V praxi to znamená, že ak neexistujú biznis pravidlá, ktoré by boli zároveň formálne definované, dokumentované, zdieľané a schválené, hodnota KPI 2.2a je automaticky 0%. Tento výsledok je rovnaký pre všetky vybrané registre.

Výsledok finálne meraného KPI: ukazovateľ Dodržiavanie biznis pravidla (vybrané biznis pravidlá pre vybrané registre)

Register	Parameter	Ukazovateľ	KPI 2.2a	Zmeraná hodnota KPI 2.2a
Register úpadcov - RU	Konzistentnosť	Dodržiavanie biznis pravidla	Percentuálny podiel atribútov, ktorých hodnoty plne dodržiajú definované biznis pravidlá.	0 %
Register adries - RA				0 %
Register právnických osôb - RPO				27 %

**Tabuľka 10: Finálne merané KPI: ukazovateľ Dodržiavanie biznis pravidla
(vybrané biznis pravidlá pre vybrané registre)**

KPI 2.2a indikuje dodržiavanie biznis pravidiel na 100% pre každý atribút. V budúcnosti sa môže definovať nové KPI pre ukazovateľ dodržiavanie biznis pravidla s nižšou prahovou hodnotou ako 100%. Určenie prahových hodnôt by malo vychádzať z biznis potrieb.

Poznámka: Farebné označenie zmeranej hodnoty KPI 2.2a je určené prahovými hodnotami KPI. Odporúčané prahové hodnoty sú nasledovné:

- Červená: pod 50 %
- Oranžová: 50-85 %
- Zelená: nad 85 %

Hodnotenie finálne meraného KPI: ukazovateľ Dodržiavanie biznis pravidla (vybrané biznis pravidlá pre vybrané registre)

V prípade, ak by bolo všetkých 25 vybraných biznis pravidiel (meraných v predošlých kapitolách) formálne definovaných, zdokumentovaných, zdieľaných a schválených, tak hodnotu KPI 2.2a je možné objektívne zmerať. Výsledky zobrazujú nasledovný stav:

- Register úpadcov RU: žiadne z definovaných biznis pravidiel nie je v plnom rozsahu dodržiavané. KPI 2.2a je 0%, lebo žiadna z hodnôt KPI nedosahuje 100%.
- Register adries RA: žiadne z definovaných biznis pravidiel nie je v plnom rozsahu dodržiavané. KPI 2.2a je 0%, lebo žiadna z hodnôt KPI nedosahuje 100%.
- Register právnických osôb RPO: tri z jedenástich definovaných biznis pravidiel sú v plnom rozsahu dodržiavané. KPI 2.2a je 27%, lebo len tri z jedenást' hodnôt KPI dosahuje 100%.

8 **Záverečné zhrnutie pre výpočet KPI dátovej kvality**

Meranie vybraných dátových ukazovateľov pre vybrané registre a datasety vo verejnej správe, popísané v tomto dokumente, má za svoj hlavný cieľ pomôcť zavedeniu kultúry, ktorá je založená na objektívnych faktoch a kvantitatívnych merateľných metrikách, ktoré sa dajú koncepčne a systematicky monitorovať a vyhodnocovať. Cieľom je maximálna pravdivosť súčasného stavu dátovej kvality a návod i motivácia pre tých, ktorí chcú zlepšovať nielen dátovú kvalitu, ale aj kvalitu verejných služieb, ktoré sú na dátovej kvalite závislé.

Druhým cieľom tejto iniciatívy je zavedenie jednotnej konzistentnej metodiky na meranie a vyhodnocovanie dátovej kvality. Aby každý rezort vo verejnej správe mal jednotnú použiteľnú metodickú podporu, ktorá umožní konzistentné porovnávanie výsledkov merania a ľahšiu integráciu medzirezortných výsledkov.

Tretím cieľom tohto výstupu je vysvetliť dôležitosť znalosti a dodržiavania biznis pravidiel v procese merania dátovej kvality a tiež základné princípy ich vytvárania a vážne dôsledky ich neprítomnosti tam, kde sú potrebné pre objektívne kontextové vyhodnotenie.

Štvrtým cieľom bol výpočet samotných KPI dátovej kvality.

Piatym cieľom bolo vytvorenie základov diagnostiky dátovej kvality pre ich efektívnejšiu terapiu a proaktívnu prevenciu pre zlepšenie celkovej kvality dát a riadenia ich životného cyklu. Formalizácia odporúčaní pre zlepšenie dátovej kvality vo verejnej správe bude nasledujúcim výstupom tejto aktivity.

A šiestym zámerom tejto aktivity je príprava používateľsky prívetivej prezentácie vo verejnej správe. Tiež sa na základe týchto vstupov pripraví dataset parametrov dátovej kvality vo forme otvorených dát.

S istotou je možné povedať, že táto iniciatíva merania dátovej kvality prispeje k naštartovaniu objektívneho monitoringu a vyhodnocovaniu stavu aj trendov v kvalite informácii, tak aby sa postupne zvýšila kvalita komunikácie, rozhodovania, služieb občanom a tým aj celková kvalita života na Slovensku.

9 Zoznamy

9.1 Zoznam tabuliek

Tabuľka 1: Karta merania pre Register úpadcov	5
Tabuľka 2: Karta merania pre Register adries	6
Tabuľka 3: Karta merania pre Register právnických osôb	8
Tabuľka 4: Zoznam KPI a ich použitie v meraní	16
Tabuľka 5: Výsledok prvého meraného KPI: ukazovateľ Sledovanie konzistentnosti (celá vzorka datasetu pre vybrané registre).....	17
Tabuľka 6: Vybrané biznis pravidlá pre vybrané registre	26
Tabuľka 7: KPI otázky a odpovede pre definované biznis pravidlá	28
Tabuľka 8: Výsledok merania KPI ukazovateľov pre meranie dátovej kvality vybraných biznis pravidiel.....	30
Tabuľka 9: Finálne merané KPI: ukazovateľ Dodržiavanie biznis pravidla (celá vzorka datasetu pre vybrané registre).....	35
Tabuľka 10: Finálne merané KPI: ukazovateľ Dodržiavanie biznis pravidla (vybrané biznis pravidlá pre vybrané registre).....	36

10 Prílohy

10.1 Príloha A: Výsledky merania dátovej kvality pre KPI v rámci profilácie.xlsx