



# Výstup č. 1.2.1: Opatrenia pre zvýšenie dátovej kvality

Zmluva o dielo č. 445/2022

*Projekt:*

**Zlepšenie využívania údajov vo verejnej  
správe**

*ITMS kód projektu:*

**314011S979**

## Kontrola a schválenie dokumentu

### História revízií

Verzia	Autor	Dátum	Revízia
1.0	Mário Finka	18.08.2023	

### Revízia

	Recenzent	Dátum revízie
1		
2		
3		
4		
5		

### Schválenie

	Schvaľovateľ	Podpis	Dátum schválenia
1			
2			
3			
4			
5			

## Slovník a skratky

<b>AI</b>	Umelá inteligencia (Artificial Intelligence)
<b>API</b>	Rozhranie využívané pri vývoji webových alebo mobilných aplikácií (Application Programming Interface)
<b>biznis pravidlo</b>	Formalizovaná požiadavka pre hodnoty údajov v atribúte v registri (databáze, zozname)
<b>CIP</b>	Projekt Centrálna integračná platforma
<b>CMÚ</b>	Centrálny model údajov verejnej správy
<b>CSRÚ</b>	Centrálna správa referenčných údajov
<b>dataset</b>	Ucelená a samostatne použiteľná skupina súvisiacich údajov vytvorených a udržiavaných na určitý účel a uložených spoločne podľa rovnakej schémy
<b>DBSCAN</b>	Algoritmus klastrovania založený na hustote (Density-Based Clustering Algorithm)
<b>DI</b>	Projekt Dátová integrácia
<b>DK</b>	Dátová kvalita
<b>DQM</b>	Data Quality Monitor
<b>ESA 2010</b>	Európsky systém národných a regionálnych účtov
<b>EŠS</b>	Európsky štatistický systém
<b>EÚ</b>	Európska únia
<b>GPT</b>	Počítačový model hĺbkového učenia vytvorený na spracovanie ľudského jazyka vyvinutý spoločnosťou OpenAI (Generative Pre-trained Transformers)
<b>IČO</b>	Identifikačné číslo organizácie
<b>IFO</b>	Identifikátor fyzickej osoby
<b>IP</b>	Inšpektorát práce
<b>IS</b>	Informačný systém
<b>ISO</b>	Medzinárodná organizácia pre štandardizáciu (International Organization for Standardization)
<b>ISOP</b>	Informačný systém ochrany práce

<b>ISVS</b>	Informačný systém verejnej správy
<b>JSON-LD</b>	JavaScript Object Notation for Linked Data
<b>LLM</b>	Rozsiahly jazykový model (Large Language Model)
<b>Metodika merania DK</b>	Metodika merania dátovej kvality vo verejnej správe <sup>1</sup>
<b>MIRRI</b>	Ministerstvo investícií, regionálneho rozvoja a informatizácie Slovenskej republiky
<b>ML</b>	Strojové učenie (Machine Learning)
<b>MPSVaR SR</b>	Ministerstvo práce, sociálnych vecí a rodiny Slovenskej republiky
<b>MV SR</b>	Ministerstvo vnútra Slovenskej republiky
<b>NIP</b>	Národný inšpektorát práce
<b>NLP</b>	Spracovanie prirodzeného jazyka (Natural Language Processing)
<b>OE</b>	Objekt evidencie
<b>OVM</b>	Orgán verejnej moci
<b>PaaS</b>	Platforma ako služba
<b>PDF</b>	Formát súboru vyvinutý firmou Adobe na ukladanie dokumentov nezávisle na softvéri a hardvéri, na ktorom boli vytvorené (Portable Document Format)
<b>PSČ</b>	Poštové smerovacie číslo
<b>RA</b>	Register adries
<b>RDF</b>	Resource Description Framework
<b>REGOB</b>	Registra obyvateľov Slovenskej republiky
<b>RFO</b>	Register fyzických osôb
<b>RNZ</b>	Register nelegálneho zamestávania
<b>RPCA</b>	Robustná analýza hlavných komponentov (Robust Principal Component Analysis)
<b>RPO</b>	Register právnických osôb
<b>SQL</b>	Štruktúrovaný dotazovací jazyk (Structured Query Language)

---

<sup>1</sup> [KPMG KIDS Report v5.2a \(datalab.digital\)](#)

<b>ŠRO</b>	Štatistický register organizácií vedený Štatistickým úradom SR podľa zákona č. 540/2001 Z.z. o štátnej štatistike v znení neskorších predpisov
<b>ŠÚ SR</b>	Štatistický úrad Slovenskej republiky
<b>ÚPSVR</b>	Úrad práce sociálnych vecí a rodiny Slovenskej republiky
<b>Zákon o údajoch</b>	Pripravovaný zákon o správe, prístupňovaní a používaní vybraných kategórií údajov evidovaných v informačných systémoch verejnej správy a o zmene a doplnení niektorých zákonov <sup>2</sup>

---

<sup>2</sup> [1.-Zakon-o-udajoch\\_28052021.docx \(live.com\)](#)

## Obsah

1	Úvod a zhrnutie	1
2	Vyhodnotenie súčasnej dátovej kvality	2
2.1	Register právnických osôb (RPO)	2
2.1.1	Charakteristika RPO	2
2.1.2	Procesy riadenia kvality dát RPO	3
2.1.3	Kvalita dát v RPO	4
2.1.4	Zhrnutie zistení	15
2.2	Register adries (RA)	16
2.2.1	Charakteristika RA	16
2.2.2	Procesy riadenia kvality dát RA	17
2.2.3	Kvalita dát v RA	18
2.2.4	Zhrnutie zistení	51
2.3	Register nelegálneho zamestnávania (RNZ)	57
2.3.1	Charakteristika RNZ	58
2.3.2	Procesy riadenia kvality dát RNZ	59
2.3.3	Kvalita dát v RNZ	59
2.3.4	Zhrnutie zistení	65
2.4	Prehľad zistení z výstupu č. 1.1.1 Štandardizácia dátovej kvality	69
3	Formulácia odporúčaní pre zlepšenie dátovej kvality	72
3.1	Centrálne dostupné služby riadenia kvality dát	72
3.2	Väzby na analýzu súčasného stavu	73
3.3	Odporúčania pre RPO	75
3.4	Odporúčania pre RA	76
3.5	Odporúčania pre RNZ	77
4	Zovšeobecnenie odporúčaní - Príručka kvality pre OVM	78
4.1	Úvod, kontext a východiská	79
4.2	Údaje vo verejnej správe a ich kvalita	81
4.2.1	Vymedzenie dátovej kvality vo verejnej správe	81
4.2.2	Závazok k dátovej kvalite	83
4.3	Riadenie dátovej kvality	84
4.3.1	Organizačné zabezpečenie riadenia dátovej kvality	85
4.3.2	Procesné zabezpečenie riadenia dátovej kvality	87
4.3.3	Technické zabezpečenie riadenia dátovej kvality a dátové štandardy	94
4.4	Koordinácia a dohľad nad riadením dátovej kvality OVM	95
4.4.1	Koordináčne nástroje zabezpečujúce dohľad nad riadením dátovej kvality OVM	96
5	Návrh postupov a metód pre čistenie údajov	100
5.1	Osvedčené postupy v širšom kontexte správy dát	100
5.2	Najčastejšie používané postupy a metódy čistenia údajov	101
5.3	Návrh postupov čistenia údajov hodnotených registrov	104
5.3.1	RPO	104

5.3.2	RA	104
5.3.3	RNZ	105
6	Návrh algoritmov pre zlepšenie dátovej kvality	106
6.1	Modelové prípady použitia pokročilých algoritmických metód	107
6.1.1	Study Fetch – Zefektívnenie čistenia údajov pomocou veľkých jazykových modelov (LLM)	107
6.1.2	Uber – Zabezpečenie kvality údajov pomocou strojového učenia	109
6.1.3	Netflix – Detekcia anomálií a ďalšie so strojovým učením	109
6.2	Algoritmy pre pokročilé čistenie dát - Talend a OpenMetadata	111
6.2.1	Tradičné algoritmické metódy v Talende	111
6.2.2	Pokročilé metódy a algoritmy strojového učenia v Talende	112
7	Naplánovanie zlepšenia dátovej kvality	114
8	Dohľad nad realizáciou odporúčaní pre zlepšenie dátovej kvality	118
9	Register tabuliek a obrázkov	121

# 1 Úvod a zhrnutie

Tento dokument bol pripravený v rámci projektu „Zlepšenie využívania údajov vo verejnej správe“, ktorý má ambíciu transformovať fungovanie inštitúcií verejnej správy tak, aby dokázali maximálne efektívne spravovať a zdieľať údaje, využívať údaje pre lepšie rozhodovanie na základe faktov a dôkazov a lepším využívaním dát zvýšiť efektivitu a adresnosť služieb. Vznikol ako súčasť realizácie aktivity číslo 1 Manažment kvality údajov, ktorej zámerom je zaviesť manažment kvality údajov do procesov vybraných inštitúcií verejnej správy. Cieľom tohto výstupu je identifikovať odporúčania pre zvýšenie dátovej kvality vo verejnej správe a navrhnúť časový rámec ich realizácie.

Prvá kapitola výstupu poskytuje pohľad na súčasný stav dátovej kvality posúdením dátovej kvality a jej riadenia v súvislosti s tromi vybranými registrami (Register právnických osôb, Register adries a Register nelegálneho zamestnávania). Výsledkom tohto posúdenia sú zistenia, ktoré vyplynuli z analytických stretnutí so správcami registrov ako aj z vykonaných meraní dátovej kvality. V závere kapitoly sú pre vytvorenie komplexnejšieho pohľadu a chápanie širších súvislostí doplnené zistenia z posúdenia dátovej kvality vo všeobecnosti, ktorých zdrojom je výstup č. 1.1.1 Štandardizácia dátovej kvality.

Nasleduje kapitola, ktorej cieľom je formulácia odporúčaní pre zlepšenie dátovej kvality založených na zisteniach uvedených v prvej kapitole. Súčasťou tejto kapitoly je aj stručný prehľad služieb dátovej kvality, ktoré sú detailne rozpracované vo výstupe č. 1.2.2 Služby dátovej kvality, a ktorých využitie orgánmi verejnej moci by malo napomáhať k napĺňaniu odporúčaní pre zlepšenie dátovej kvality uvedenými v druhej kapitole. Navrhovaný plán realizácie týchto odporúčaní je predmetom kapitoly 7.

Súčasťou tohto dokumentu je aj kapitola zameraná na príručku dátovej kvality OVM v podobe navrhovaného rámca pre riadenie dátovej kvality a implementovanie procesov riadenia kvality do každodenných činností OVM. Táto príručka predpokladá a vyžaduje jednotné vymedzenie pojmu dátová kvalita vo verejnej správe ako aj proaktívny prístup relevantných subjektov riadenia dátovej kvality v podobe jednotlivých OVM a MIRRI ako koordinátora riadenia dátovej kvality.

Ďalšie dve kapitoly sú venované návrhom postupov a metód čistenia, vrátane návrhu algoritmov pre zlepšenie dátovej kvality. Obsahujú osvedčené a najčastejšie používané metódy čistenia údajov doplnené o informáciu k potrebe/využitelnosti alebo uchopeniu oblasti čistenia údajov v prípade troch registrov už skôr posudzovaných v tomto dokumente.

V záverečnej kapitole je pozornosť venovaná dohľadu nad realizáciou opatrení na zlepšovanie kvality načrtnutím modelu dohľadu vykonávaného zo strany koordinátora dátovej kvality (MIRRI) voči OVM. Okrem nástrojov dohľadu, ktoré sú vo veľkej miere popísané už v kapitole zameranej na príručku kvality pre OVM, je cieľom tejto kapitoly aj zdôrazniť potrebu využitia dohľadu spôsobom, ktorý motivuje OVM k zlepšovaniu dátovej kvality.



## 2 Vyhodnotenie súčasnej dátovej kvality

Táto časť popisuje vyhodnotenie súčasnej dátovej kvality vo verejnej správe. Pri hodnotení boli samostatne posudzované registre rozličných veľkostí, konkrétne Register právnických osôb (RPO), Register adries (RA) a Register nelegálneho zamestnávania (RNZ). RPO a RA sú v súčasnosti dva najčastejšie využívané referenčné registre na Slovensku. RNZ by sa mal stať referenčným v blízkej dobe. V rámci vyhodnotenia boli najprv uskutočnené analytické stretnutia so správcami jednotlivých registrov a na základe získaných informácií boli v spolupráci so správcami registrov a Dátovej kancelárie pripravené podklady na meranie dátovej kvality RA a RNZ. Z organizačných dôvodov nebolo možné uskutočniť meranie DK na RPO, ale pre účely tohto dokumentu boli použité výstupy z meraní, ktoré ŠÚ SR zverejňoval do minulého roka.

Na základe uskutočnených meraní a analytických stretnutí bola vyhodnotená dátová kvalita a riadenie dátovej kvality vo všetkých troch registroch. Záverom sú doplnené aj zistenia z hodnotenia situácie v dátovej kvalite vo verejnej správe vo všeobecnosti (súčasť výstupu č. 1.1.1 Štandardizácia dátovej kvality).

### 2.1 Register právnických osôb (RPO)

Register právnických osôb je integrovaným registrom údajov o právnických osobách, podnikateľoch a orgánoch verejnej moci, ktoré sú v súčasnosti registrované alebo evidované vo viac ako 100 zdrojových registroch a evidenciách rôznych informačných systémov. RPO vznikol v súlade s Národnou koncepciou informatizácie verejnej správy a od 16. 5. 2016 je referenčným registrom.

#### 2.1.1 Charakteristika RPO

Vytvorenie a fungovanie RPO je v SR legislatívne zakotvené prostredníctvom zák. č. 272/2015 Z. z. o registri právnických osôb, podnikateľov a orgánov verejnej moci<sup>3</sup> (ďalej len „zákon o RPO“). Tento zákon upravuje zriadenie a vedenie RPO, zoznam údajov vedených v RPO (ako aj podmienky a postup pri ich zápise/zmene/výmaze a s tým spojené úlohy orgánov verejnej moci), identifikačné číslo organizácie (spôsob jeho tvorby, určovania a pridelovania) a vytvorenie a fungovanie generického registra.

RPO je informačným systémom verejnej správy, ktorého správcom je Štatistický úrad.

Údaje zapisované do RPO sú poskytované zo zdrojových registrov povinných osôb (povinnými osobami sú orgány verejnej moci, ktoré vedú príslušnú evidenciu – napríklad registrové súdy vo vzťahu k údajom obchodného registra). Hodnoty zapisovaných údajov zapisuje do RPO Štatistický úrad v takej podobe, v akej ich poskytnú povinné osoby. Zodpovednosť za to, aby hodnoty poskytnutých údajov zodpovedali hodnotám údajov zapísaných v zdrojovom registri (do momentu, kým sa dostanú do dispozície Štatistického úradu), má povinná osoba.

<sup>3</sup> [272/2015 Z.z. - Zákon o registri právnických osôb, ... - SLOV-LEX](#)

Hodnoty zapisovaných údajov sú od povinných osôb poskytované elektronicky:

- prostredníctvom ich evidencie v generickom registri (generický register je nástroj na evidovanie a poskytovanie zapisovaných údajov do RPO, prevádzkovateľom generického registra je Štatistický úrad),
- zabezpečením technickej dostupnosti hodnôt údajov a zverejnením podrobností o spôsobe a periodicite na svojom webovom sídle,
- priamo Štatistickým úradom, ktorý za povinnú osobu na základe poverenia vykonáva evidenciu do generického registra (približne 85% obcí).

Jadro RPO (veľké registre ako napríklad živnostenský register alebo obchodný register) sú na RPO integrované cez webové služby a poskytovanie dát prebieha automatizované prostredníctvom zmenových dávok. Evidovanie údajov do generického registra (napríklad dáta od zväzov, združení, komôr a pod.) prebieha na základe vloženia údajov cez importný hárok, pričom povinné osoby majú pre prístup do generického registra vytvorené prihlasovacie meno a heslo.

Údaje zapísané v RPO sú poskytované širokému spektru subjektov (napríklad orgánom štátnej správy, súdom, obciam, VÚC, členským štátom EÚ), pričom RPO je na informatívne účely dostupný aj na webovom sídle Štatistického úradu<sup>4</sup>. S využitím údajov zapísaných v RPO sú poskytované služby ako napríklad výpis z RPO alebo elektronický odpis z RPO.

## 2.1.2 Procesy riadenia kvality dát RPO

Riadenie kvality dát RPO je realizované správou informačného systému RPO, ktorý má v sebe zabudovanú automatickú identifikáciu systémových chýb. Dáta poskytované na vstupe do RPO sú podrobené automatickej systémovej kontrole podľa definovaných pravidiel, pričom identifikované chyby sú spätne poskytované povinným osobám s cieľom ich opravy.

Základom pre riadenie kvality dát RPO sú informácie z výsledkov pravidelného merania dátovej kvality vykonávané Štatistickým úradom. Meranie kvality dát RPO je vykonávané od roku 2020 (najprv štvrťročne a od roku 2022 jedenkrát ročne). V rámci merania Štatistický úrad definuje pravidlá (tzv. biznis pravidlá), ktoré sú vyhodnocované, pričom tieto pravidlá sú rozdelené podľa ich závažnosti (od závažnosti 1 po závažnosť 5) – správy z merania sú zverejňované na webovom sídle Štatistického úradu<sup>5</sup>. Na základe výsledkov z merania Štatistický úrad formuluje odporúčania a zistené nedostatky v kvalite údajov RPO sa snaží riešiť v spolupráci s jednotlivými vlastníkami zdrojových registrov.

Posledné meranie dátovej kvality bolo uskutočnené v apríli 2022 – schválená novela zákona o RPO (s účinnosťou od októbra 2023) zavádza povinnosť Štatistického úradu každoročne vypracúvať z merania kvality údajov RPO správu a zároveň ju zverejňovať na svojom webovom sídle. Táto správa má obsahovať najmä merateľné kritériá kvality údajov vopred určené Štatistickým úradom, výsledky merania kvality údajov a ich hodnotenie. V súčasnosti (podľa vyjadrenia Štatistického úradu) dochádza k prehodnocovaniu vyhodnotenia kvality a biznis pravidiel na základe doterajších skúseností.

<sup>4</sup> [RPO - Prihlasovacia obrazovka \(statistics.sk\)](https://statistics.sk)

<sup>5</sup> [Správy z merania dátovej kvality RPO \(statistics.sk\)](https://statistics.sk)

Štatistický úrad má za účelom riadenia dátovej kvality registrov novozriadené oddelenie kvality registrov, ktoré má však na starosti iba kvalitu/hodnoty tých údajov, ktoré do RPO zapisuje podľa zákona o RPO Štatistický úrad (ide o kód hlavnej činnosti a ESA 2010). Pre riadenie kvality ostatných údajov zapisovaných do RPO nie je k dispozícii formalizovaná metodika.

Vyššie spomenutá novela zákona o RPO<sup>6</sup> zavádza správne delikty vo vzťahu k poskytovaniu údajov povinnými osobami do RPO a zároveň upravuje možnosť uloženia pokuty (Štatistickým úradom) povinnej osobe, ktorá tieto údaje neposkytne v zákonom stanovenej lehote.

Vo vzťahu k riadeniu dátovej kvality je potrebné ešte uviesť, že na úrovni Štatistického úradu prebieha projekt s priamym vplyvom na dátovú kvalitu RPO - Manažment údajov RPO<sup>7</sup>, ktorý smeruje k nastaveniu procesov riadenia kvality dát v rámci celého životného cyklu údajov (čistenie údajov a dosiahnutie požadovanej kvality dát) a navrhnutiu s tým spojenej potrebnej legislatívnej úpravy.

### 2.1.3 Kvalita dát v RPO

V nasledujúcej kapitole sa nachádza zhrnutie výsledkov merania dátovej kvality na registri RPO v marci v roku 2022. Biznis pravidlá pomocou ktorých bola dátová kvalita odmeraná neboli v danej dobe začlenené do skupín parametrov dátovej kvality, ktoré by mali slúžiť ako ukazovatele na prezentovanie celkovej kvality dátových súborov v jednotlivých oblastiach. V rámci merania boli pravidlá začlenené len do kategórií závažnosti problému dátovej kvality v prípade, že je zistená dátová nekvalita určitého BP. Tento spôsob prezentovania dátovej kvality resp. nekvality je síce funkčný, avšak nedokáže všeobecne a objektívne informovať o reálnom stave dátovej kvality dátového súboru. Nižšie sa teda nachádzajú tabuľky so zatriedenými biznis pravidlami z tohto merania na základe kategorizácie po parametroch dátovej kvality, v rámci ktorých je vyhodnotená dátová kvalita RPO v čase merania po jednotlivých parametroch a následne aj celková dátová kvalita RPO po jednotlivých parametroch dátovej kvality.

#### 2.1.3.1 Presnosť

V rámci merania dátovej kvality bola vykonaná kontrola syntaktickej presnosti pomocou 3 biznis pravidiel nad 2 atribútmi nachádzajúcimi sa v RPO. V nasledujúcej tabuľke je uvedený zoznam biznis pravidiel z merania, ktoré podľa ich obsahovej povahy spadajú pod parameter dátovej kvality presnosť.

**Tabuľka 1 Biznis pravidlá pre meranie presnosti RPO**

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_015	76,03 %	23,97%	1116025	351933	Rodné číslo je 9- alebo 10- znakový reťazec zložený z číslíc.

<sup>6</sup> [nrsr.sk/web/Dynamic/DocumentPreview.aspx?DocID=532183](https://nrsr.sk/web/Dynamic/DocumentPreview.aspx?DocID=532183)

<sup>7</sup> [Manažment údajov Registra právnických osôb, podnikateľov a orgánov verejnej moci \(statistics.sk\)](https://www.statistics.sk/)

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_080	99,98 %	0,02 %	1467631	327	Rodné číslo obsahuje dátum narodenia osoby.
BP_SU_RPO_081	100 %	0 %	933890	4	"IČO začína číslicou 0, 1, 2, 3, 4 alebo 5 (s výnimkou pre IČO "Neuvedené")."

Zistenie: Register právnických osôb na základe 3 meraných biznis pravidiel dosahuje dátovú kvalitu v parametri presnosť na úrovni 92 %.

### 2.1.3.2 Správnosť

V rámci merania dátovej kvality bola vykonaná kontrola správnosti vyplnenia atribútu pomocou 23 biznis pravidiel nad nešpecifikovaným množstvom atribútov, keďže z formulácie jednotlivých biznis pravidiel nie je jednoznačne možné určiť tento počet atribútov. V nasledujúcej tabuľke je uvedený zoznam biznis pravidiel z merania, ktoré podľa ich obsahovej povahy spadajú pod parameter dátovej kvality správnosť.

**Tabuľka 2 Biznis pravidlá pre meranie správnosti RPO**

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_007	100 %	0 %	933894	0	"Ak sú pri atribúte vyplnené hodnoty „dátum platnosti od“ aj „dátum platnosti do“, tak „dátum platnosti od“ je nižší ako „dátum platnosti do“."
BP_SU_RPO_008	100 %	0 %	933893	1	"Ak sú v adrese sídla vyplnené obec a okres z príslušných číselníkov, tak daná obec patrí do daného okresu."
BP_SU_RPO_009	100 %	0 %	933893	1	"Ak sú v adrese sídla vyplnené obec a kraj z príslušných číselníkov, tak daná obec patrí do daného kraja."
BP_SU_RPO_010	100 %	0 %	933893	1	"Ak sú v adrese sídla vyplnené okres a kraj z príslušných číselníkov, tak daný okres patrí do daného kraja."
BP_SU_RPO_012	99,73 %	0,27 %	931369	2525	"Subjekt má uvedenú ako právnu formu platnú položku z číselníka právnych foriem a zároveň nemá

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
					uvedenú právnu formu „995 – Nešpecifikovaná právna forma“.
BP_SU_RPO_013	99,90 %	0,10 %	932893	1001	Dátum vzniku subjektu je vyšší ako 1.1.1910.
BP_SU_RPO_014	91,91 %	8,09 %	858326	75568	Adresa sídla obsahuje údaj o obci z príslušného číselníka.
BP_SU_RPO_031	99,99 %	0,01 %	307378	48	"Spoločnosť s ručením obmedzeným má v názve uvedenú svoju právnu formu."
BP_SU_RPO_032	99,73 %	0,27 %	6890	19	Akciová spoločnosť má v názve uvedenú svoju právnu formu.
BP_SU_RPO_033	98,51 %	1,49 %	858	13	"Verejná obchodná spoločnosť má v názve uvedenú svoju právnu formu."
BP_SU_RPO_034	99,62 %	0,38 %	1301	5	"Komanditná spoločnosť má v názve uvedenú svoju právnu formu."
BP_SU_RPO_035	100 %	0 %	359	0	"Jednoduchá spoločnosť na akcie má v názve uvedenú svoju právnu formu."
BP_SU_RPO_075	100 %	0 %	321874	31	"Ak má subjekt z Obchodného registra, ktorý nie je organizačnou zložkou, v názve uvedenú právnu formu spoločnosť s ručením obmedzeným, tak má subjekt takúto právnu formu."
BP_SU_RPO_076	100 %	0 %	321899	6	"Ak má subjekt z Obchodného registra, ktorý nie je organizačnou zložkou, v názve uvedenú právnu formu akciová spoločnosť, tak má subjekt takúto právnu formu."
BP_SU_RPO_077	100 %	0 %	321904	1	"Ak má subjekt z Obchodného registra, ktorý nie je organizačnou zložkou, v názve uvedenú právnu formu verejná obchodná spoločnosť, tak má subjekt takúto právnu formu."

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_078	100 %	0 %	321905	0	"Ak má subjekt z Obchodného registra, ktorý nie je organizačnou zložkou, v názve uvedenú právnu formu
BP_SU_RPO_079	100 %	0 %	321905	0	"Ak má subjekt z Obchodného registra, ktorý nie je organizačnou zložkou, v názve uvedenú právnu formu jednoduchá spoločnosť na akcie, tak má subjekt takúto právnu formu."
BP_SU_RPO_083	99,88 %	0,12 %	932755	1139	"Ak je subjekt v likvidácii, tak má v obchodnom mene dodatok „v likvidácii“."
BP_SU_RPO_084	99,90 %	0,10 %	932930	964	"Ak je subjekt v konkurze, tak má v obchodnom mene dodatok „v konkurze“."
BP_SU_RPO_099	98,40 %	1,60 %	918917	14977	"Ak sú v adrese sídla vyplnené atribúty obec a PSČ, tak dané PSČ je priradené danej obci."

Zistenie: Register právnických osôb na základe 23 meraných biznis pravidiel dosahuje dátovú kvalitu v parametri správnosť na úrovni 99,39 %.

### 2.1.3.3 Kompletnosť

V rámci merania dátovej kvality bola vykonaná kontrola kompletnosti pomocou 45 biznis pravidiel (1 biznis pravidlo bolo vyradené) nad nešpecifikovaným množstvom atribútov, keďže z formulácie jednotlivých biznis pravidiel nie je jednoznačne možné určiť tento počet atribútov. V nasledujúcej tabuľke je uvedený zoznam biznis pravidiel z merania, ktoré podľa ich obsahovej povahy spadajú pod parameter dátovej kvality kompletnosť.

**Tabuľka 3 Biznis pravidlá pre meranie kompletnosti RPO**

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_001	99,99 %	0,01 %	933739	155	"IČO má v rámci jedného zdrojového registra najviac jeden záznam (s výnimkou pre IČO „Neuvedené“)."

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_003	100 %	0%	933862	32	"IČO právnickej alebo fyzickej osoby spĺňa kontrolu ADDO modulov 11 (s výnimkou pre IČO „Neuvedené“)."
BP_SU_RPO_011	98,43 %	1,57 %	919221	14673	"Subjekt má uvedený aspoň jeden aktívny predmet činnosti,
BP_SU_RPO_016	98,68 %	1,32 %	920544	13350	"Subjekt má uvedenú hodnotu pre Inštitucionálny sektor (ESA 2010)."
BP_SU_RPO_017	98,97 %	1,03 %	924202	9692	Subjekt má uvedený kód hlavnej ekonomickej činnosti.
BP_SU_RPO_018	100 %	0 %	933882	12	Pri adrese sídla je vyplnený atribút obec (názov obce).
BP_SU_RPO_019			Vyradené		Pri adrese sídla je vyplnený atribút ulica.
BP_SU_RPO_020	62,31 %	37,69 %	581892	352002	Pri adrese sídla je vyplnený atribút okres.
BP_SU_RPO_021	2,07 %	97,93 %	19299	914595	Pri adrese sídla je vyplnený atribút kraj.
BP_SU_RPO_022	99,98 %	0,02 %	933670	224	Pri adrese sídla je vyplnený atribút krajina.
BP_SU_RPO_023	97,29 %	2,71 %	908530	25364	Pri adrese sídla je vyplnený atribút PSČ.
BP_SU_RPO_024	100 %	0%	933894	0	Subjekt má vyplnený atribút pre obchodné meno/názov.
BP_SU_RPO_026	96,50 %	3,50 %	901118	32776	"Pri subjekte je uvedený aspoň jeden štatutárny orgán alebo zainteresovaná osoba."
BP_SU_RPO_027	95,79 %	4,21 %	977101	43031	"Ak je štatutárny orgán fyzická osoba, tak má uvedené rodné číslo alebo IFO."

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_028	82,10 %	17,90 %	837496	182636	Ak je štatutárny orgán fyzická osoba, tak má uvedené IFO.
BP_SU_RPO_029	38,24 %	61,76 %	167963	300441	"Ak je zainteresovaná osoba fyzická osoba, tak má uvedené rodné číslo alebo IFO."
BP_SU_RPO_030	26,94 %	73,06 %	126153	342251	"Ak je zainteresovaná osoba fyzická osoba, tak má uvedené IFO."
BP_SU_RPO_043	98,46 %	1,54 %	574	9	"Neinvestičný fond má štatutárny orgán typu „Správca“ alebo „Likvidátor“."
BP_SU_RPO_044	97,80 %	2,20 %	10796	243	"Spoločenstvo vlastníkov bytov a nebytových priestorov má štatutárny orgán typu „Predseda“ alebo „Likvidátor“."
BP_SU_RPO_045	100 %	0 %	8	0	Samosprávny kraj má štatutárny orgán typu „Predseda“.
BP_SU_RPO_046	85,62 %	14,38 %	2505	421	"Obec (mesto) má štatutárny orgán typu „Starosta“ alebo „Primátor“."
BP_SU_RPO_047	98,40 %	1,60 %	857	14	"Verejná obchodná spoločnosť má zainteresovanú osobu typu „Spoločník“ alebo „Spoločník v.o.s. / s.r.o.“."
BP_SU_RPO_048	99,70 %	0,30 %	1302	4	"Komanditná spoločnosť má zainteresovanú osobu typu „Komplementár“."
BP_SU_RPO_049	99,85 %	0,15 %	1304	2	"Komanditná spoločnosť má zainteresovanú osobu typu
BP_SU_RPO_050	99,90 %	0,10 %	307103	323	"Spoločnosť s ručením obmedzeným má zainteresovanú osobu typu „Spoločník“ alebo „Spoločník v.o.s. / s.r.o.“."



ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_051	84,55 %	15,45 %	8347	1526	"Ak je subjekt v likvidácii, tak má štatutárny orgán typu „Likvidátor“."
BP_SU_RPO_052	97,45 %	2,55 %	1713	45	"Ak je subjekt v konkurze, tak má zainteresovanú osobu typu „Správca konkurznej podstaty“."
BP_SU_RPO_053	33,34 %	66,66 %	1	2	"Ak je subjekt vo vyrovnaní, tak má zainteresovanú osobu"
BP_SU_RPO_054	98,15 %	1,85 %	53	1	"Ak je subjekt v reštrukturalizácii, tak má zainteresovanú osobu typu „Správca reštrukturalizačného konania“."
BP_SU_RPO_055	99,98 %	0,02 %	307357	69	"Spoločnosť s ručením obmedzeným má vyplnené základné imanie."
BP_SU_RPO_056	97,10 %	2,90 %	1268	38	Komanditná spoločnosť má vyplnené základné imanie.
BP_SU_RPO_057	99,79 %	0,21 %	6894	15	Akciová spoločnosť má vyplnené základné imanie.
BP_SU_RPO_058	92,88 %	7,12 %	1863	143	Družstvo má vyplnené základné imanie.
BP_SU_RPO_059	99,79 %	0,21 %	6894	15	Akciová spoločnosť má vyplnené akcie.
BP_SU_RPO_060	92,38 %	7,62 %	1853	153	Družstvo má vyplnené vklady.
BP_SU_RPO_061	87,61 %	12,39 %	763	108	"Verejná obchodná spoločnosť má štatutárny orgán typu „Spoločník“ alebo „Likvidátor“."
BP_SU_RPO_062	99,47 %	0,53 %	1299	7	"Komanditná spoločnosť má štatutárny orgán typu „Komplementár“ alebo „Likvidátor“."

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_063	99,19 %	0,81 %	304918	2508	"Spoločnosť s ručením obmedzeným má štatutárny orgán typu „Konateľ“ alebo „Likvidátor“."
BP_SU_RPO_064	100 %	0 %	8	0	"Európske zoskupenie hospodárskych záujmov má štatutárny orgán typu „Konateľ“ alebo „Likvidátor“."
BP_SU_RPO_065	100 %	0 %	4	0	"Európske zoskupenie územnej spolupráce má štatutárny orgán typu „Riaditeľ“ alebo „Likvidátor“."
BP_SU_RPO_066	98,40 %	1,60 %	4220	69	"Nezisková organizácia a nezisková organizácia poskytujúca všeobecne prospešné služby má štatutárny orgán typu „Riaditeľ“, „Likvidátor“, „Predseda“ alebo „Predseda predstavenstva“."
BP_SU_RPO_067	86,38 %	13,62 %	672	106	"Nadácia má štatutárny orgán typu „Správca“ alebo „Likvidátor“."
BP_SU_RPO_082	5,76 %	94,24 %	53783	880111	"Adresa sídla obsahuje údaj o identifikátore adresy z Registra adries."
BP_SU_RPO_092	100 %	0 %	933894	0	"Ak je subjekt fyzická osoba, ktorá podniká na základe živnostenského zákona, tak má uvedenú právnu formu „Podnikateľ-fyzická osoba-nezapísaný v obchodnom v registri“."
BP_SU_RPO_093	100 %	0 %	933891	3	"Ak má subjekt uvedený právny stav „Zrušenie z dôvodu zániku platnosti oprávnenia na výkon činnosti“, tak má vyplnený dátum zániku."
BP_SU_RPO_094	100 %	0 %	933892	2	"Ak má subjekt uvedený právny stav „Zaniknutá spoločnosť“,

Zistenie: Register právnických osôb na základe 45 meraných biznis pravidiel dosahuje dátovú kvalitu v parametri kompletnosť na úrovni 87,76 %.

#### 2.1.3.4 Unikátnosť

V rámci merania dátovej kvality RPO, ktoré bolo uskutočnené vybranými biznis pravidlami navrhnutými Štatistickým úradom v marci 2022 sa nenachádzali žiadne biznis pravidlá, ktoré by bolo možné začleniť pod kategóriu parametra unikátnosť. Parameter dátovej kvality unikátnosť je potrebné začleniť do merania dátovej kvality na RPO, keďže v prípade určitých atribútov má veľký význam merať, či je splnená unikátnosť záznamu.

#### 2.1.3.5 Konzistentnosť

V rámci merania dátovej kvality bola vykonaná kontrola kompletnosti pomocou 18 biznis pravidiel (7 biznis pravidiel bolo nemeranych) nad nešpecifikovaným množstvom atribútov, keďže z formulácie jednotlivých biznis pravidiel nie je jednoznačne možné určiť tento počet atribútov. V nasledujúcej tabuľke je uvedený zoznam biznis pravidiel z merania, ktoré podľa ich obsahovej povahy spadajú pod parameter dátovej kvality konzistentnosť.

**Tabuľka 4 Biznis pravidlá pre meranie konzistentnosti RPO**

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_005	100 %	0 %	933855	39	"Ak je subjekt fyzická osoba, tak jeho IČO nie je pridelené inému subjektu naprieč všetkými zdrojovými registrami."
BP_SU_RPO_006	100 %	0 %	933823	71	"Ak je subjekt právnická osoba, tak jeho IČO je evidované v práve jednom zdrojovom registri."
BP_SU_RPO_036			Nemerané		"Ak je IČO zo zdrojového registra evidované ako aktívne v RPO, tak je IČO evidované ako aktívne aj v príslušnom zdrojovom registri."
BP_SU_RPO_037			Nemerané		"Ak je IČO evidované ako aktívne v zdrojovom registri, tak je IČO evidované ako aktívne aj v RPO."
BP_SU_RPO_038			Nemerané		"Názov subjektu v RPO je zhodný s názvom subjektu v zdrojovom registri, ktorý údaje do RPO pre daný subjekt zaslal."
BP_SU_RPO_039			Nemerané		"Hodnota pre atribút ulica subjektu v RPO je zhodná s hodnotou pre tento atribút v zdrojovom registri, ktorý údaje do RPO pre daný subjekt zaslal."

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_040					"Hodnota pre atribút obec subjektu v RPO je zhodná s hodnotou pre tento atribút v zdrojovom registri, ktorý údaje do RPO pre daný subjekt zaslal."
BP_SU_RPO_041					"Zoznam názvov štatutárnych orgánov subjektu v RPO je zhodný so zoznamom názvov štatutárnych orgánov tohto subjektu v zdrojovom registri, ktorý údaje do RPO pre daný subjekt zaslal."
BP_SU_RPO_042					"Zoznam názvov zainteresovaných osôb subjektu v RPO je zhodný so zoznamom názvov zainteresovaných osôb tohto subjektu v zdrojovom registri, ktorý údaje do RPO pre daný subjekt zaslal."
BP_SU_RPO_068	100 %	0 %	307401	25	"Zdrojovým registrom pre spoločnosť s ručením obmedzeným je Obchodný register."
BP_SU_RPO_069	100 %	0 %	871	0	"Zdrojovým registrom pre verejnú obchodnú spoločnosť je Obchodný register."
BP_SU_RPO_070	100 %	0 %	1306	0	"Zdrojovým registrom pre komanditnú spoločnosť je Obchodný register."
BP_SU_RPO_071	99,99 %	0,01 %	6908	1	"Zdrojovým registrom pre akciovú spoločnosť je Obchodný register."
BP_SU_RPO_072	100 %	0 %	359	0	"Zdrojovým registrom pre jednoduchú spoločnosť na akcie je Obchodný register."
BP_SU_RPO_073	100 %	0 %	583	0	"Zdrojovým registrom pre neinvestičný fond je Register neinvestičných fondov."

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_SU_RPO_074	100 %	0 %	778	0	Zdrojovým registrom pre nadáciu je Register nadácií.
BP_SU_RPO_086	100 %	0 %	2942	0	"Zdrojovým registrom pre cirkevnú organizáciu je Register cirkví a náboženských spoločností alebo Register právnických osôb, ktoré odvodzujú právnu subjektivitu od právnej subjektivity cirkví a náboženských spoločností."
BP_SU_RPO_087	100 %	0 %	47	0	"Zdrojovým registrom pre verejnú výskumnú inštitúciu je Register verejných výskumných inštitúcií."
BP_SU_RPO_088	100 %	0 %	2006	0	Zdrojovým registrom pre družstvo je Obchodný register.
BP_SU_RPO_090	100 %	0 %	159	0	"Zdrojovým registrom pre politickú stranu, politické hnutie je Register politických strán a politických hnutí."
BP_SU_RPO_091	99,98 %	0,02 %	59853	17	"Zdrojovým registrom pre združenie (zväz, spolok, spoločnosť, klub ai.) je Register občianskych združení alebo Register odborových a zamestnávateľských organizácií."
BP_SU_RPO_095	100 %	0 %	390	0	"Zdrojovým registrom pre združenie účastníkov pozemkových
BP_SU_RPO_096	100 %	0 %	3191	0	"Zdrojovým registrom pre poľovnícku organizáciu je Register poľovníckych organizácií."
BP_SU_RPO_097	100 %	0 %	11039	0	"Zdrojovým registrom pre spoločenstvo vlastníkov bytov a nebytových priestorov je Register spoločenstiev vlastníkov bytov a nebytových priestorov."
BP_SU_RPO_098	100 %	0 %	2824	0	"Zdrojovým registrom pre pozemkové spoločenstvo je

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
--------------------	-----------	-------------	---------------	-----------------	-----------------

Register pozemkových spoločností."

Zistenie: Register právnických osôb na základe 18 meraných biznis pravidiel dosahuje dátovú kvalitu v parametri konzistentnosť na úrovni 99,99 %.

### 2.1.3.6 Referenčná integrita

V rámci merania dátovej kvality bola vykonaná kontrola syntaktickej presnosti pomocou 2 biznis pravidiel nad 1 atribútom nachádzajúcimi sa v RPO. V nasledujúcej tabuľke je uvedený zoznam biznis pravidiel z merania, ktoré podľa ich obsahovej povahy spadajú pod parameter dátovej kvality referenčná integrita.

**Tabuľka 5 Biznis pravidiel pre meranie referenčnej integrity RPO**

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
--------------------	-----------	-------------	---------------	-----------------	-----------------

BP_SU_RPO_002	99,89 %	0,11 %	932808	1086	"IČO právnickej alebo fyzickej osoby je referenčný identifikátor a jeho hodnota je 8-znakový reťazec zložený z číslíc."
BP_SU_RPO_004	99,99 %	0,01 %	933753	141	"Ak je subjekt fyzická osoba, tak má priradené práve jedno IČO naprieč všetkými zdrojovými registrami."

Zistenie: Register právnických osôb na základe 2 meraných biznis pravidiel dosahuje dátovú kvalitu v parametri referenčná integrita na úrovni 99,94 %.

### 2.1.4 Zhrnutie zistení

Napriek tomu, že Štatistický úrad zrejme ako jediný OVM spomedzi vlastníkov referenčných registrov dobrovoľne vykonával merania dátovej kvality na pravidelnej báze a zverejňoval výsledky týchto meraní v podobe správy na svojom webovom sídle, v roku 2023 takáto správa zverejnená nebola (z dôvodu personálnych zmien), z čoho možno usudzovať, že ani meranie vykonané nebolo. Pozitívnym signálom pokračovania, resp. oživenia tejto aktivity je zákonne vymedzená povinnosť Štatistického úradu spočívajúca v povinnosti každoročne takúto správu zverejňovať (na základe vyššie spomenutej novely zákona o RPO účinné od októbra 2023). Zákon iba rámcovo vymedzuje čo má byť obsahom správy z merania dátovej kvality avšak spôsob, akým má byť kvalita meraná, nie je definovaný. Zároveň nie je definovaný ani postup ako ďalej riešiť zistenia uvedené v správe a vyplývajúce z meraní.

Kvalita dát RPO je významne ovplyvnená kvalitou dát príslušných zdrojových registrov, kde za najvýznamnejšie problémy sú z pohľadu Štatistického úradu považované

hlavne nekvalitu v údajoch historických záznamov zapísaných do týchto registrov ešte pred účinnosťou zákona o RPO. Vyčistenie tejto nekvality si bude vyžadovať preskúmanie príčin priamo na úrovni zdrojových registrov a nápravu tak, aby boli tieto „opravené“ dáta premietnuté do RPO. Preto je nevyhnutná súčinnosť medzi povinnými osobami a Štatistickým úradom s cieľom zvýšenia dátovej kvality údajov v RPO – aktivity súvisiace s čistením údajov v zdrojových registroch musia byť riadené a centrálné koordinované vo vzťahu ku všetkým dotknutým OVM.

Novela zákona o RPO síce zavádza povinnosť povinných osôb poskytovať zákonom vymedzené údaje do RPO v stanovenej lehote s hrozbou udelenia pokuty v prípade nesplnenia tejto povinnosti, avšak zákon o RPO nedefinuje kvalitu poskytovaných údajov ani parametre kvality, ktoré majú byť splnené. Preto novela zákona o RPO nevyhnutne nemusí viesť k zvýšeniu kvality údajov v rámci jednotlivých parametrov kvality. Pokuty môžu byť nástrojom pre získanie údajov do RPO v určitých lehotách avšak nie v konkrétnej úrovni kvality.

Na riadenie kvality údajov poskytovaných do RPO nie je vytvorený jednotný štandardizovaný proces, podľa ktorého by OVM mali postupovať. Táto činnosť nie je centrálné koordinovaná. Štatistický úrad nemá vytvorené štandardizované procesy riadenia dátovej kvality RPO, ktoré by vychádzali z jednotného formalizovaného rámca pre vlastníkov referenčných registrov. Tu je nutné pripomenúť, že riadenie dátovej kvality referenčných registrov v súčasnosti nie je centrálné štandardizované.

## 2.2 Register adries (RA)

Register adries je informačný systém, ktorý obsahuje adresy a adresné body bytových a nebytových budov na celom území Slovenska. Keďže ide o referenčný register, údaje v ňom uvedené sú považované za všeobecne záväzné a nespochybniteľné. Správcom RA je Ministerstvo vnútra Slovenskej republiky (MV SR).

### 2.2.1 Charakteristika RA

RA je spolu s RPO najčastejšie využívaným registrom na Slovensku. Prevádzku a údržbu informačného systému RA vykonáva MV SR, zatiaľ čo obsah RA vytvárajú a aktualizujú mestá a obce v rámci preneseného výkonu štátnej správy. Adresa v RA je súbor údajov, ktoré identifikujú umiestnenie hlavného vchodu alebo vstupov do budovy, ktorým bolo určené orientačné číslo. Ich identifikácia je možná cez identifikátor adresy (množina alfanumerických znakov) určeného na účely komunikácie medzi informačnými systémami. Údaje z tejto evidencie môžu využívať štátne orgány, podnikateľská sféra a občania.

Správa registra a poskytovanie jeho informácií sa riadia nasledujúcou legislatívou:

- Zákon č. 125/2015 Z. z. o registri adries a o zmene a doplnení niektorých zákonov,
- Zákon č. 42/2022 Z. z., ktorým sa mení a dopĺňa zákon č. 125/2015 Z. z. o registri adries a o zmene a doplnení niektorých zákonov a ktorým sa mení a dopĺňa zákon Slovenskej národnej rady č. 369/1990 Zb. o obecnom zriadení v znení neskorších predpisov,
- Vyhláška MVSR č. 142/2015 Z. z. ktorou sa vykonáva zákon č. 125/2015 Z. z. o registri adries a o zmene a doplnení niektorých zákonov,

- Zákon SNR č. 369/1990 Zb. o obecnom zriadení,
- Vyhláška MVSR č. 31/2003 Z. z. ktorou sa ustanovujú podrobnosti o označovaní ulíc a iných verejných priestranstiev a o číslovaní stavieb.

RA bol pôvodne vytvorený migráciou adries trvalých pobytov z Registra obyvateľov SR (REGOB) do Informačného systému Registra adries. Pri migrácii boli údaje čiastočne skontrolované, ale aj napriek tomu bolo do RA prenesených veľké množstvo chýb. Kvalitu týchto historických údajov zvyšuje MV SR postupne. Pokiaľ ide o novo vkladané údaje, MV SR nevytvára nové záznamy, vkladajú ich zamestnanci obcí prostredníctvom webového formulára. V prípade zistenia nesprávnych údajov (v novo vložených aj v historických údajoch) požiada MV SR miestne príslušný úrad o opravu alebo aktualizáciu záznamu.

## 2.2.2 Procesy riadenia kvality dát RA

MV SR nemá schválenú metodiku riadenia dátovej kvality a s ňou súvisiace interné akty riadenia. Aj napriek tomu sa však prakticky používajú procesy riadenia dátovej kvality v súlade s Metodikou riadenia DK.

### 10-krokový proces pre zhodnotenie, zlepšenie a vytvorenie dátovej kvality

Správcovia registra vyhodnotili požiadavky na kvalitu registra, ktoré vychádzajú najmä z legislatívnych požiadaviek na jeho údaje. Na základe požiadaviek vykonávajú pravidelné merania selectami nad databázou. Posúdili dopad zistenej nekvality a vyhodnotením výsledkov merania identifikovali príčiny nekvalitných údajov a konkrétne záznamy. Vytvorili realizovateľný plán odstraňovania nedostatkov postavený na dostupných personálnych kapacitách a na definovanom procese opravy chýb v registri. Správca registra kontaktuje úrady zodpovedné za jednotlivé chybné záznamy, upozorní ich na nedostatky a požiada ich o opravu. Následne meraním skontroluje vykonanie opravy. V rámci prevencie budúcich chýb boli vo formulároch pre vkladanie a opravu údajov implementované kontroly pred odoslaním, ktoré zabezpečujú merateľné dodržiavanie požiadaviek z hľadiska presnosti, správnosti, kompletnosti a unikátnosti.

### Pripomienkovanie kvality RA verejnosťou

V rámci zvyšovania dátovej kvality bol tiež zverejnený webový formulár Pripomienkovanie kvality RA. Táto webová služba umožňuje pripomienkovanie a sprístupňuje spätnú väzbu na kvalitu informácie o adrese. Komunikácia cez Pripomienkovanie kvality RA prebieha synchrónne.

### Školenia zvyšovania dátovej kvality

MV SR pravidelne realizuje pre obce školenia zamerané nielen na používanie rozhrania pre zadávanie a aktualizáciu údajov v RA, ale aj na metodickú stránku správy RA a zvyšovanie jeho dátovej kvality. V spojení s automatizovanými kontrolami vo formulároch na zadávanie údajov prispieva táto forma osvetly k neustálemu zvyšovaniu dátovej kvality registra.

### Kampane na opravu údajov



Aj napriek tomu, že správca registra nemá k dispozícii nástroje na automatizované meranie a opravu údajov, pracuje neustále na zvyšovaní dátovej kvality. Pravidelne vykonáva kampane na čistenie referenčných údajov, najmä tých, ktoré boli do registra historicky prenesené pri jeho vytvorení. Zatiaľ čo kontroly na vstupe zabraňujú zaneseniu nekvalitných údajov do registra, opakované kampane postupne zvyšujú dátovú kvalitu už existujúcich referenčných údajov. Kampane prebiehajú minimálne dvakrát ročne. Dátový kurátor spustí nad aktuálnymi údajmi selekt, ktorý v databáze registra identifikuje nezhody na základe vybratého biznis pravidla. Vecný vlastník registra vyhodnotí výstup selektu. Zistené nezhody potom nahlási kontaktným osobám jednotlivých obcí a dohodne si s nimi termín odstránenia nezhôd. Po uplynutí termínu dátový kurátor znova spustí selekt s rovnakým biznis pravidlom a poskytne ho vecnému vlastníkovi na vyhodnotenie. Pokiaľ neboli nezhody odstránené, vecný vlastník opakovane vyzve obec k náprave. MV SR nemá k dispozícii mechanizmy na vynútenie úprav, ale perzistentnou snahou o čistenie dát postupne dosahujú stále vyššiu kvalitu referenčných údajov.

### 2.2.3 Kvalita dát v RA

V spolupráci s autormi tohto dokumentu a s Dátovou kanceláriou pripravilo MV SR zoznam biznis pravidiel a vykonali meranie v súlade s Metodikou merania DK. Meranie bolo vykonané nad datasetom umiestneným v perzistentnej vrstve CSRÚ. Dataset bol aktuálny k 29.06.2023 a obsahoval 6 476 963 záznamov. Pri vyhodnocovaní výsledkov merania pracovná skupina zistila, že pri definovaní väčšiny biznis pravidiel nebol vzatý do úvahy aspekt platnosti záznamu a väčšina biznis pravidiel bola meraná aj na údajoch, ktoré nie sú referenčné (neplatné záznamy). V dôsledku toho boli do výsledkov merania zahrnuté aj nezhody, ktoré neboli z hľadiska referenčných údajov relevantné. Z kapacitných dôvodov však nebolo možné meranie znova zopakovať. Preto pri vyhodnotení každého biznis pravidla pracovná skupina overovala aj to, či bol výsledok ovplyvnený nezhodami v neplatných údajoch.

Výsledky merania boli posudzované z hľadiska parametrov presnosť, správnosť, kompletnosť, unikátnosť, konzistentnosť a referenčná integrita. V tabuľke nižšie je uvedené priradenie meraných biznis pravidiel hodnoteným parametrom dátovej kvality.

#### Tabuľka 6 Zoznam meraných biznis pravidiel nad RA

ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_001	objectId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_002	versionId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_003	validFrom <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_004	validTo <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_005	effectiveDate <= ako aktuálny dátum	Presnosť

ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_007	ItemCode <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_008	ItemCode = SK* (podľa číselníka CL000023) pre objekty kde validTo > ako aktuálny dátum	Presnosť
BP_MVSR_RA_009	ItemName <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_010	ItemName min. počet znakov 3	Správnosť
BP_MVSR_RA_011	ItemName max. počet znakov 50	Správnosť
BP_MVSR_RA_012	ItemName REGEX ^(?:[a-zA-Z]+(?:[.\ \\-])?\s?)+\$	Správnosť
BP_MVSR_RA_013	ItemName = číselníkovej hodnote číselníka CL000023 pre objekty kde validTo > ako aktuálny dátum	Presnosť
BP_MVSR_RA_014	ItemName nesmie obsahovať dve medzery za sebou	Správnosť
BP_MVSR_RA_015	ItemName nesmie obsahovať medzeru na konci	Správnosť
BP_MVSR_RA_016	ItemName nesmie obsahovať medzeru na začiatku	Správnosť
BP_MVSR_RA_017	objectId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_018	versionId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_019	validFrom <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_020	validTo <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_021	effectiveDate <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_023	ItemCode <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_024	ItemCode = SK* (podľa číselníka CL000024) pre objekty kde validTo > ako aktuálny dátum	Presnosť
BP_MVSR_RA_025	ItemName <> NULL alebo prázdna hodnota	Kompletnosť

ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_026	ItemName min. počet znakov 3	Správnosť
BP_MVSR_RA_027	ItemName max. počet znakov 50	Správnosť
BP_MVSR_RA_028	ItemName REGEX <code>^(?:[a-zA-Z]+(?:[^\s,])?\s*(-)?)+\$</code>	Správnosť
BP_MVSR_RA_029	ItemName = číselníkovej hodnote číselníka CL000024 pre objekty kde validTo > ako aktuálny dátum	Presnosť
BP_MVSR_RA_030	ItemName nesmie obsahovať dve medzery za sebou	Správnosť
BP_MVSR_RA_031	ItemName nesmie obsahovať medzeru na konci	Správnosť
BP_MVSR_RA_032	ItemName nesmie obsahovať medzeru na začiatku	Správnosť
BP_MVSR_RA_033	regionIdentifier <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_034	regionIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 01 REGION	Konzistentnosť
BP_MVSR_RA_035	objectId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_036	versionId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_037	validFrom <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_038	validTo <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_039	effectiveDate <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_041	ItemCode <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_042	ItemCode = SK* (podľa číselníka CL000025) pre objekty kde validTo > ako aktuálny dátum	Presnosť
BP_MVSR_RA_043	ItemName <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_044	ItemName min. počet znakov 3	Správnosť

ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_045	ItemName max. počet znakov 50	Správnosť
BP_MVSR_RA_046	ItemName REGEX <code>^(?:[a-zA-Z]+(?:[.\-,.])?\s*(-)?)+\$</code>	Správnosť
BP_MVSR_RA_047	ItemName = číselníkovej hodnote číselníka CL000025 pre objekty kde validTo > ako aktuálny dátum	Presnosť
BP_MVSR_RA_048	ItemName nesmie obsahovať dve medzery za sebou	Správnosť
BP_MVSR_RA_049	ItemName nesmie obsahovať medzeru na konci	Správnosť
BP_MVSR_RA_050	ItemName nesmie obsahovať medzeru na začiatku	Správnosť
BP_MVSR_RA_051	countyIdentifier <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_052	countyIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 02 COUNTY	Konzistentnosť
BP_MVSR_RA_053	status = MINICIPALITY;CITY;CITY_DISTRICT;MILITARY_DISTRICT	Kompletnosť
BP_MVSR_RA_054	objectId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_055	versionId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_056	validFrom <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_057	validTo <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_058	effectiveDate <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_060	ItemCode <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_061	ItemCode = začína na "4" (podľa číselníka CL010141) pre objekty kde validTo > ako aktuálny dátum	Presnosť
BP_MVSR_RA_062	ItemName <> NULL alebo prázdna hodnota	Kompletnosť

ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_063	ItemName min. počet znakov 3	Správnosť
BP_MVSR_RA_064	ItemName max. počet znakov 50	Správnosť
BP_MVSR_RA_065	ItemName REGEX <code>^(?:[a-zA-Z](?:[^\s])?\s?)+\$</code>	Správnosť
BP_MVSR_RA_066	ItemName = číselníkovej hodnote číselníka CL010141 pre objekty kde validTo > ako aktuálny dátum	Presnosť
BP_MVSR_RA_067	ItemName nesmie obsahovať dve medzery za sebou	Správnosť
BP_MVSR_RA_068	ItemName nesmie obsahovať medzeru na konci	Správnosť
BP_MVSR_RA_069	ItemName nesmie obsahovať medzeru na začiatku	Správnosť
BP_MVSR_RA_070	municipalityIdentifier <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_071	municipalityIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 03 MUNICIPALITY	Konzistentnosť
BP_MVSR_RA_072	objectId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_073	versionId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_074	validFrom <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_075	validTo <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_076	effectiveDate <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_077	StreetName nesmie obsahovať dve medzery za sebou	Správnosť
BP_MVSR_RA_078	StreetName nesmie obsahovať medzeru na konci	Správnosť
BP_MVSR_RA_079	StreetName nesmie obsahovať medzeru na začiatku	Správnosť
BP_MVSR_RA_081	municipalityIdentifier <> NULL alebo prázdna hodnota	Kompletnosť

ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_082	ItemName min. počet znakov 3	Správnosť
BP_MVSR_RA_083	ItemName max. počet znakov 50	Správnosť
BP_MVSR_RA_084	ItemName REGEX <code>^(?:[a-zA-Z0-9]+(?:[.\-])?\s?)+\$</code>	Správnosť
BP_MVSR_RA_085	municipalityIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 03 MUNICIPALITY	Konzistentnosť
BP_MVSR_RA_086	districtIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 04 DISTRICT	Konzistentnosť
BP_MVSR_RA_087	objectId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_088	versionId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_089	validFrom <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_090	validTo <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_091	effectiveDate <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_092	PropertyRegistrationNumber <> 0	Presnosť
BP_MVSR_RA_093	PropertyRegistrationNumber <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_094	PropertyRegistrationNumber iba číselné znaky	Presnosť
BP_MVSR_RA_095	PropertyRegistrationNumber max 5 miestne prirodzené číslo	Presnosť
BP_MVSR_RA_097	Ak PropertyRegistrationNumber <> 0 a <> NULL musí existovať väzba na nadradený objekt MUNICIPALITY, čiže municipalityIdentifier <> NULL	Presnosť
BP_MVSR_RA_098	Ak PropertyRegistrationNumber <> 0 a <> NULL musí existovať väzba na nadradený objekt DISTRICT, čiže districtIdentifier <> NULL	Presnosť
BP_MVSR_RA_100	BuildingPurpose_ItemCode <> NULL alebo prázdna hodnota	Kompletnosť

ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_101	BuildingPurpose_ItemName <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_103	BuildingTypeCode_ItemCode <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_104	BuildingTypeCode_ItemName <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_105	municipalityIdentifier <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_106	municipalityIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 03 MUNICIPALITY	Konzistentnosť
BP_MVSR_RA_107	districtIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 04 DISTRICT	Konzistentnosť
BP_MVSR_RA_108	objectId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_109	versionId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_110	validFrom <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_111	validTo <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_112	effectiveDate <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_113	BuildingNumber <> 0 (môže však nadobúdať hodnotu NULL, v prípade ak pre dané SČ "propertyRegistrationNumber" z 06 Building nie je určené OČ a existuje väzba cez propertyRegistrationNumberIdentifier, čiže propertyRegistrationNumberIdentifier OČ <> NULL)	Kompletnosť
BP_MVSR_RA_114	BuildingNumber max 5 miestny text (alfanumerická hodnota - čísla a veľké písmená vid'. BP_MVSR_RA_115)	Kompletnosť
BP_MVSR_RA_115	BuildingNumber na 1 - 4 mieste môže byť prirodzené číslo, na 5 miestne môže byť text veľké písmeno abecedy okrem CH a DZ a bez diakritiky	Kompletnosť
BP_MVSR_RA_116	Ak BuildingNumber <> 0 a <> NULL musí existovať väzba na platný nadradený objekt ulica streetNameIdentifier, čiže streetNameIdentifier <> NULL	Presnosť

ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_117	Ak BuildingNumber <> 0 a <> NULL musí byť unikátne, nesmie obsahovať duplicity. Duplicity sú prípustné len v prípade, ak je OČ pridelené na inú ulicu (rôzny streetNameIdentifier).	Presnosť
BP_MVSR_RA_118	BuildingIndex <> NULL alebo prázdna hodnota Identifikátor adresy je referenčný identifikátor a jeho hodnota je alfanumerický reťazec zložený z číslíc a písmena A, ktoré môže byť na poslednom mieste.	Referenčná integrita
BP_MVSR_RA_119	Každá adresa má priradený práve jeden unikátny identifikátor adresy pre objekty kde validTo > ako aktuálny dátum.	Unikátnosť
BP_MVSR_RA_120	PostalCode <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_121	PostalCode = 5 miestne číslo	Presnosť
BP_MVSR_RA_122	PropertyRegistrationNumberIdentifier <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_123	PropertyRegistrationNumberIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu PropertyRegistrationNumber z 06 BUILDING	Konzistentnosť
BP_MVSR_RA_124	streetNameIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 05 STREET	Konzistentnosť
BP_MVSR_RA_125	Adresné body (zemepisná šírka) sa musia nachádzať v rámci intervalu medzi najzápadnejším/ najvýchodnejším bodom Slovenskej republiky. ADRBOD_X_MAX < 22.60151 ADRBOD_X_MIN > 16.71026	Presnosť
BP_MVSR_RA_126	Adresné body (zemepisná dĺžka) sa musia nachádzať v rámci intervalu medzi najsevernejším/ najjužnejším bodom Slovenskej republiky. ADRBOD_Y_MAX < 49.61379 ADRBOD_Y_MIN > 47.73116	Presnosť
BP_MVSR_RA_127	Zemepisná dĺžka GPS súradnica je povinným atribútom. AxisL <> NULL	Kompletnosť
BP_MVSR_RA_128	Zemepisná šírka GPS súradnica je povinným atribútom. AxisB <> NULL	Kompletnosť



ID BP	BP (biznis pravidlo)	Parameter
BP_MVSR_RA_129	objectId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_130	versionId <> NULL alebo prázdna hodnota	Kompletnosť
BP_MVSR_RA_131	validFrom <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_132	validTo <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_133	effectiveDate <= ako aktuálny dátum	Presnosť
BP_MVSR_RA_134	BuildingUnit_UnitNumber nesmie obsahovať dve medzery za sebou	Presnosť
BP_MVSR_RA_135	BuildingUnit_UnitNumber nesmie obsahovať medzeru na konci	Presnosť
BP_MVSR_RA_136	BuildingUnit_UnitNumber nesmie obsahovať medzeru na začiatku	Presnosť
BP_MVSR_RA_137	BuildingUnit_UnitNumber <> NULL alebo prázdna hodnota	Presnosť
BP_MVSR_RA_138	BuildingUnit_Floor nesmie obsahovať dve medzery za sebou	Presnosť
BP_MVSR_RA_139	BuildingUnit_Floor nesmie obsahovať medzeru na konci	Presnosť
BP_MVSR_RA_140	BuildingUnit_Floor nesmie obsahovať medzeru na začiatku	Presnosť
BP_MVSR_RA_141	BuildingUnit_Floor = celé číslo v prípade ak nie je NULL alebo prázdna hodnota	Presnosť
BP_MVSR_RA_142	BuildingUnit_Floor < 50	Presnosť
BP_MVSR_RA_143	BuildingNumberIdentifier <> NULL alebo prázdna hodnota	Presnosť
BP_MVSR_RA_144	BuildingNumberIdentifier <> 0	Presnosť
BP_MVSR_RA_145	BuildingNumberIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 07 BUILDING_NUMBER	Presnosť

### 2.2.3.1 Presnosť

V rámci merania DK z hľadiska presnosti bolo meraných 45 biznis pravidiel na tabuľkách 01 REGION, 02 COUNTY, 03 MUNICIPALITY, 04 DISTRICT, 05 STREET, 06 BUILDING, 07 BUILDING\_NUMBER a 08 BUILDING\_UNIT. V nasledujúcej tabuľke je uvedený zoznam biznis pravidiel, u ktorých bola zistená presnosť 100 %. Biznis pravidlá, u ktorých bola meraním zistená znížená presnosť, sú vyhodnotenie nižšie.

**Tabuľka 7 Biznis pravidlá s overenou presnosťou 100 %**

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_MVSR_RA_003	100 %	0 %	20	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_005	100 %	0 %	20	0	effectiveDate <= ako aktuálny dátum
BP_MVSR_RA_008	100 %	0 %	8	0	ItemCode = SK* (podľa číselníka CL000023) pre objekty kde validTo > ako aktuálny dátum
BP_MVSR_RA_019	100 %	0 %	83	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_021	100 %	0 %	83	0	effectiveDate <= ako aktuálny dátum
BP_MVSR_RA_037	100 %	0 %	3 039	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_039	100 %	0 %	3 039	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_056	100 %	0 %	7 240	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_074	100 %	0 %	52 203	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_089	100 %	0 %	3 470 288	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_091	100 %	0 %	3 470 185	2	effectiveDate <= ako aktuálny dátum
BP_MVSR_RA_097	100 %	0 %	3 303 170	0	Ak PropertyRegistrationNumber <> 0 a <> NULL musí existovať väzba na nadradený objekt MUNICIPALITY, čiže municipalityIdentifier <> NULL

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_MVSR_RA_110	100 %	0 %	6 476 963	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_112	100 %	0 %	6 476 963	0	effectiveDate <= ako aktuálny dátum
BP_MVSR_RA_131	100 %	0 %	166 943	0	validFrom <= ako aktuálny dátum
BP_MVSR_RA_133	100 %	0 %	166 943	0	effectiveDate <= ako aktuálny dátum
BP_MVSR_RA_135	100 %	0 %	150 945	2	BuildingUnit_UnitNumber nesmie obsahovať medzeru na konci
BP_MVSR_RA_136	100 %	0 %	150 946	1	BuildingUnit_UnitNumber nesmie obsahovať medzeru na začiatku
BP_MVSR_RA_137	100 %	0 %	150 947	0	BuildingUnit_UnitNumber <> NULL alebo prázdna hodnota
BP_MVSR_RA_143	100 %	0 %	150 947	0	BuildingNumberIdentifier <> NULL alebo prázdna hodnota
BP_MVSR_RA_144	100 %	0 %	150 947	0	BuildingNumberIdentifier <> 0

#### BP\_MVSR\_RA\_004 – validTo => ako aktuálny dátum

Kontrola bola vykonaná na 20 záznamoch tabuľky 02 REGION.

#### Tabuľka 8 Výsledok merania pravidla BP\_MVSR\_RA\_004

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
40 %	60 %	8	12

Zistenie: Presnosť atribútu „validTo“ v tabuľke REGION je 60 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_020** – validTo => ako aktuálny dátum

Kontrola bola vykonaná na 83 záznamoch tabuľky 02 COUNTY.

**Tabuľka 9 Výsledok merania pravidla BP\_MVSR\_RA\_020**

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
95,18 %	4,82 %	79	4

Zistenie: Presnosť atribútu „validTo“ v tabuľke 02 COUNTY je 95,18 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_038** – validTo => ako aktuálny dátum

Kontrola bola vykonaná na 83 záznamoch tabuľky 03 MUNICIPALITY.

**Tabuľka 10 Výsledok merania pravidla BP\_MVSR\_RA\_038**

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
96,64 %	3,36 %	2 937	102

Zistenie: Presnosť atribútu „validTo“ v tabuľke 03 MUNICIPALITY je 96,64 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_057** – validTo => ako aktuálny dátum

Kontrola bola vykonaná na 7 240 záznamoch tabuľky 04 DISTRICT.

**Tabuľka 11 Výsledok merania pravidla BP\_MVSR\_RA\_057**

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
49,54 %	50,46 %	3 587	3 653

Zistenie: Presnosť atribútu „validTo“ v tabuľke 04 DISTRICT je 49,54 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_061** – ItemCode = začína na "4" (podľa číselníka CL010141)

Kontrola bola vykonaná na 3 587 záznamoch tabuľky 04 DISTRICT.

#### Tabuľka 12 Výsledok merania pravidla BP\_MVSR\_RA\_061

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
50,96 %	49,04 %	1828	1759

Zistenie: Presnosť atribútu „ItemCode“ v tabuľke 04 DISTRICT je 50,96 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_075** – validTo <= ako aktuálny dátum

Kontrola bola vykonaná na 52 203 záznamoch tabuľky 05 STREET.

#### Tabuľka 13 Výsledok merania pravidla BP\_MVSR\_RA\_075

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
57,48 %	42,52 %	30006	22197

Zistenie: Presnosť atribútu „validTo“ v tabuľke 05 STREET je 57,48 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_090** – validTo <= ako aktuálny dátum

Kontrola bola vykonaná na 3 470 290 záznamoch tabuľky 06 BUILDING.

#### Tabuľka 14 Výsledok merania pravidla BP\_MVSR\_RA\_090

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
45,08 %	54,92 %	1564564	1905726

Zistenie: Presnosť atribútu „validTo“ v tabuľke 06 BUILDING je 45,08 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

#### BP\_MVSR\_RA\_092 – PropertyRegistrationNumber <> 0

Kontrola bola vykonaná na 2 000 581 záznamoch tabuľky 06 BUILDING.

#### Tabuľka 15 Výsledok merania pravidla BP\_MVSR\_RA\_092

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,64 %	0,36 %	1993410	7171

Zistenie: Presnosť atribútu „PropertyRegistrationNumber“ v tabuľke 06 BUILDING je 99,64 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

#### BP\_MVSR\_RA\_094 – PropertyRegistrationNumber iba číselné znaky

Kontrola bola vykonaná na 2 000 581 záznamoch tabuľky 06 BUILDING.

#### Tabuľka 16 Výsledok merania pravidla BP\_MVSR\_RA\_094

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,65 %	0,35 %	1993668	6913

Zistenie: Presnosť atribútu „PropertyRegistrationNumber“ v tabuľke 06 BUILDING je 99,65 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

#### BP\_MVSR\_RA\_095 – PropertyRegistrationNumber max 5 miestne prirodzené číslo

Kontrola bola vykonaná na 2 000 581 záznamoch tabuľky 06 BUILDING.

#### Tabuľka 17 Výsledok merania pravidla BP\_MVSR\_RA\_095

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,65 %	0,35 %	1 993 664	6 917

Zistenie: Presnosť atribútu „PropertyRegistrationNumber“ v tabuľke 06 BUILDING je 99,65 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_098** – Ak PropertyRegistrationNumber <> 0 a <> NULL musí existovať väzba na nadradený objekt DISTRICT, čiže districtIdentifier <> NULL

Kontrola bola vykonaná na 3 303 170 záznamoch tabuľky 06 BUILDING.

#### Tabuľka 18 Výsledok merania pravidla BP\_MVSR\_RA\_098

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
42,93 %	57,07 %	1418023	1885147

Zistenie: Presnosť atribútu „PropertyRegistrationNumber“ v tabuľke 06 BUILDING je 42,93 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_111** – validTo <= ako aktuálny dátum

Kontrola bola vykonaná na 6 476 963 záznamoch tabuľky 07 BUILDING\_NUMBER.

#### Tabuľka 19 Výsledok merania pravidla BP\_MVSR\_RA\_111

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
41,82 %	58,18 %	2708446	3768517

Zistenie: Presnosť atribútu „validTo“ v tabuľke 07 BUILDING\_NUMBER je 41,82 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_116** – Ak BuildingNumber <> 0 a <> NULL musí existovať väzba na platný nadradený objekt ulica streetNameIdentifier, čiže streetNameIdentifier <> NULL

Kontrola bola vykonaná na 634 708 záznamoch tabuľky 07 BUILDING\_NUMBER.

#### Tabuľka 20 Výsledok merania pravidla BP\_MVSR\_RA\_116

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,96 %	0,04 %	634432	276

Zistenie: Presnosť atribútu „BuildingNumber“ v tabuľke 07 BUILDING\_NUMBER je 99,96 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

### BP\_MVSR\_RA\_121 – PostalCode = 5 miestne číslo

Kontrola bola vykonaná na 2 708 446 záznamoch tabuľky 07 BUILDING\_NUMBER.

#### Tabuľka 21 Výsledok merania pravidla BP\_MVSR\_RA\_121

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
54,93 %	45,07 %	1487642	1 220 804

Zistenie: Presnosť atribútu „PostalCode“ v tabuľke 07 BUILDING\_NUMBER je 54,93 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

### BP\_MVSR\_RA\_125 – Adresné body (zemepisná šírka) sa musia nachádzať v rámci intervalu medzi najzápadnejším/ najvýchodnejším bodom Slovenskej republiky.

Kontrola bola vykonaná na 2 708 446 záznamoch tabuľky 07 BUILDING\_NUMBER.

#### Tabuľka 22 Výsledok merania pravidla BP\_MVSR\_RA\_125

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
83,25 %	16,75 %	2254777	453669

Zistenie: Presnosť atribútu „Adresné body (zemepisná šírka)“ v tabuľke 07 BUILDING\_NUMBER je 83,25 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

### BP\_MVSR\_RA\_126 – Adresné body (zemepisná dĺžka) sa musia nachádzať v rámci intervalu medzi najsevernejším/ najjužnejším bodom Slovenskej republiky.

Kontrola bola vykonaná na 2 708 446 záznamoch tabuľky 07 BUILDING\_NUMBER.

#### Tabuľka 23 Výsledok merania pravidla BP\_MVSR\_RA\_126

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
83,25 %	16,75 %	2254763	453683



Zistenie: Presnosť atribútu „Adresné body (zemepisná dĺžka)“ v tabuľke 07 BUILDING\_NUMBER je 83,25 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_132** – validTo <= ako aktuálny dátum

Kontrola bola vykonaná na 166 943 záznamoch tabuľky 08 BUILDING\_UNIT.

#### Tabuľka 24 Výsledok merania pravidla BP\_MVSR\_RA\_132

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
90,42 %	9,58 %	150947	15996

Zistenie: Presnosť atribútu „validTo“ v tabuľke 08 BUILDING\_UNIT je 90,42 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_134** – BuildingUnit\_UnitNumber nesmie obsahovať dve medzery za sebou

Kontrola bola vykonaná na 150 947 záznamoch tabuľky 08 BUILDING\_UNIT.

#### Tabuľka 25 Výsledok merania pravidla BP\_MVSR\_RA\_134

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,98 %	0,02%	150 917	30

Zistenie: Presnosť atribútu „BuildingUnit\_UnitNumber“ v tabuľke 08 BUILDING\_UNIT je 99,98 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_138** – BuildingUnit\_Floor nesmie obsahovať dve medzery za sebou

Kontrola bola vykonaná na 150 947 záznamoch tabuľky 08 BUILDING\_UNIT.

#### Tabuľka 26 Výsledok merania pravidla BP\_MVSR\_RA\_138

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
94,79 %	5,21 %	143089	7858

Zistenie: Presnosť atribútu „BuildingUnit\_Floor“ v tabuľke 08 BUILDING\_UNIT je 94,79 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_139** – BuildingUnit\_Floor nesmie obsahovať medzeru na konci

Kontrola bola vykonaná na 150 947 záznamoch tabuľky 08 BUILDING\_UNIT.

#### Tabuľka 27 Výsledok merania pravidla BP\_MVSR\_RA\_139

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
94,79 %	5,21 %	143089	7858

Zistenie: Presnosť atribútu „BuildingUnit\_Floor“ v tabuľke 08 BUILDING\_UNIT je 94,79 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_140** – BuildingUnit\_Floor nesmie obsahovať medzeru na začiatku

Kontrola bola vykonaná na 150 947 záznamoch tabuľky 08 BUILDING\_UNIT.

#### Tabuľka 28 Výsledok merania pravidla BP\_MVSR\_RA\_140

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
94,79 %	5,21 %	143089	7858

Zistenie: Presnosť atribútu „BuildingUnit\_Floor“ v tabuľke 08 BUILDING\_UNIT je 94,79 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_141** – BuildingUnit\_Floor = celé číslo v prípade ak nie je NULL alebo prázdna hodnota

Kontrola bola vykonaná na 150 947 záznamoch tabuľky 08 BUILDING\_UNIT.

#### Tabuľka 29 Výsledok merania pravidla BP\_MVSR\_RA\_141

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
94,79 %	5,21 %	143089	7858

Zistenie: Presnosť atribútu „BuildingUnit\_Floor“ v tabuľke 08 BUILDING\_UNIT je 94,79 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

#### BP\_MVSR\_RA\_142 – BuildingUnit\_Floor < 50

Kontrola bola vykonaná na 150 947 záznamoch tabuľky 08 BUILDING\_UNIT.

#### Tabuľka 30 Výsledok merania pravidla BP\_MVSR\_RA\_142

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
94,78 %	5,22 %	143072	7875

Zistenie: Presnosť atribútu „BuildingUnit\_Floor“ v tabuľke 08 BUILDING\_UNIT je 94,78 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

#### BP\_MVSR\_RA\_145 – BuildingNumberIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 07 BUILDING\_NUMBER

Kontrola bola vykonaná na 150 947 záznamoch tabuľky 08 BUILDING\_UNIT.

#### Tabuľka 31 Výsledok merania pravidla BP\_MVSR\_RA\_145

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
97,94 %	2,06 %	147841	3106

Zistenie: Presnosť atribútu „BuildingNumberIdentifier“ v tabuľke 08 BUILDING\_UNIT je 97,94 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

### 2.2.3.2 Správnosť

V rámci merania správnosti RA bolo meraných 28 biznis pravidiel na tabuľkách 01 REGION, 02 COUNTY, 03 MUNICIPALITY, 04 DISTRICT a 05 STREET.

V nasledujúcej tabuľke je uvedený zoznam biznis pravidiel, u ktorých bola zistená presnosť 100 %. Biznis pravidlá, u ktorých bola meraním zistená znížená presnosť, sú vyhodnotenie nižšie.

**Tabuľka 32 Biznis pravidlá s overenou správnosťou 100 %**

ID biznis pravidla	Zhoda v (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_MVSR_RA_010	100 %	0 %	20	0	ItemName min. počet znakov 3
BP_MVSR_RA_011	100 %	0 %	20	0	ItemName max. počet znakov 50
BP_MVSR_RA_014	100 %	0 %	20	0	ItemName nesmie obsahovať dve medzery za sebou
BP_MVSR_RA_015	100 %	0 %	20	0	ItemName nesmie obsahovať medzeru na konci
BP_MVSR_RA_016	100 %	0 %	20	0	ItemName nesmie obsahovať medzeru na začiatku
BP_MVSR_RA_026	100 %	0 %	83	0	ItemName min. počet znakov 3
BP_MVSR_RA_027	100 %	0 %	83	0	ItemName max. počet znakov 50
BP_MVSR_RA_030	100 %	0 %	83	0	ItemName nesmie obsahovať dve medzery za sebou
BP_MVSR_RA_031	100 %	0 %	83	0	ItemName nesmie obsahovať medzeru na konci
BP_MVSR_RA_032	100 %	0 %	83	0	ItemName nesmie obsahovať medzeru na začiatku
BP_MVSR_RA_044	100 %	0 %	3 039	0	ItemName min. počet znakov 3
BP_MVSR_RA_045	100 %	0 %	3 039	0	ItemName max. počet znakov 50
BP_MVSR_RA_048	100 %	0 %	3 039	0	ItemName nesmie obsahovať dve medzery za sebou
BP_MVSR_RA_049	100 %	0 %	3 039	0	ItemName nesmie obsahovať medzeru na konci
BP_MVSR_RA_050	100 %	0 %	3039	0	ItemName nesmie obsahovať medzeru na začiatku

BP_MVSR_RA_063	100 %	0 %	7 240	0	ItemName min. počet znakov 3
BP_MVSR_RA_064	100 %	0 %	7 240	0	ItemName max. počet znakov 50
BP_MVSR_RA_067	100 %	0 %	7 240	0	ItemName nesmie obsahovať dve medzery za sebou
BP_MVSR_RA_068	100 %	0 %	7 240	0	ItemName nesmie obsahovať medzeru na konci
BP_MVSR_RA_069	100 %	0 %	7 240	0	ItemName nesmie obsahovať medzeru na začiatku
BP_MVSR_RA_078	100 %	0 %	7 240	0	StreetName nesmie obsahovať medzeru na konci
BP_MVSR_RA_079	100 %	0 %	7 240	0	StreetName nesmie obsahovať medzeru na začiatku
BP_MVSR_RA_083	100 %	0 %	52 203	0	ItemName max. počet znakov 50

### BP\_MVSR\_RA\_012 – ItemName REGEX $^(?:[a-zA-Z]+(?:[.\'\\-,\])?\s?)+\$$

Meranie bolo vykonané 20 záznamoch v tabuľke 01 REGION. Cieľom bolo overiť, či názov položky (atribút „ItemName“) neobsahuje nepovolené znaky.

#### Tabuľka 33 Výsledok merania pravidla BP\_MVSR\_RA\_012

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
20 %	80 %	4	16

Zistenie: Správnosť atribútu „ItemName“ v tabuľke 01 REGION pre biznis pravidlo BP\_MVSR\_RA\_012 je 20 %.

### BP\_MVSR\_RA\_028 – ItemName REGEX $^(?:[a-zA-Z]+(?:[.\'\\-,\])?\s?)+\$$

Meranie bolo vykonané 83 záznamoch v tabuľke 02 COUNTY. Cieľom bolo overiť, či názov položky (atribút „ItemName“) neobsahuje nepovolené znaky.

### Tabuľka 34 Výsledok merania pravidla BP\_MVSR\_RA\_028

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
38,55 %	61,45 %	32	51

Zistenie: Správnosť atribútu „ItemName“ v tabuľke 02 COUNTY pre biznis pravidlo BP\_MVSR\_RA\_028 je 38,55 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

#### BP\_MVSR\_RA\_065 – ItemName REGEX `^(?:[a-zA-Z]+(?:[.\'\\-,\])?\s?)+$`

Meranie bolo vykonané 7 240 záznamoch v tabuľke 04 DISTRICT. Cieľom bolo overiť, či názov položky (atribút „ItemName“) neobsahuje nepovolené znaky.

### Tabuľka 35 Výsledok merania pravidla BP\_MVSR\_RA\_065

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
26,92 %	73,08 %	1 949	5 291

Zistenie: Správnosť atribútu „ItemName“ v tabuľke 04 DISTRICT pre biznis pravidlo BP\_MVSR\_RA\_065 je 29,92 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

#### BP\_MVSR\_RA\_077 – StreetName nesmie obsahovať dve medzery za sebou

Meranie bolo vykonané 30 006 záznamoch v tabuľke 05 STREET. Cieľom bolo overiť, či názov ulice (atribút „StreetName“) neobsahuje dve medzery za sebou.

### Tabuľka 36 Výsledok merania pravidla BP\_MVSR\_RA\_077

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
26,92 %	73,08 %	1 949	5 291

Zistenie: Správnosť atribútu „ItemName“ v tabuľke 05 STREET pre biznis pravidlo BP\_MVSR\_RA\_077 je 26,92 %. Zistené nezhody je potrebné opraviť.

#### BP\_MVSR\_RA\_082 – StreetName nesmie obsahovať medzeru na konci

Meranie bolo vykonané 52 203 záznamoch v tabuľke 05 STREET. Cieľom bolo overiť, či názov ulice (atribút „StreetName“) neobsahuje medzeru na konci názvu.

### Tabuľka 37 Výsledok merania pravidla BP\_MVSR\_RA\_078

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
98,9 %	1,1 %	51 631	572

Zistenie: Správnosť atribútu „ItemName“ v tabuľke 05 STREET pre biznis pravidlo BP\_MVSR\_RA\_082 je 99,9 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

### 2.2.3.3 Kompletnosť

V rámci merania dátovej kvality z hľadiska kompletnosti boli vykonaná kontrola vyplnenosti nad 41 atribútmi. V nasledujúcej tabuľke je uvedený zoznam biznis pravidiel, u ktorých bola zistená presnosť 100 %. Biznis pravidlá, u ktorých bola meraním zistená znížená presnosť, sú vyhodnotené nižšie.

### Tabuľka 38 Biznis pravidlá s overenou kompletnosťou 100 %

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_MVSR_RA_001	100 %	0 %	20	0	objectId <> NULL alebo prázdna hodnota
BP_MVSR_RA_002	100 %	0 %	20	0	versionId <> NULL alebo prázdna hodnota
BP_MVSR_RA_007	100 %	0 %	20	0	ItemCode <> NULL alebo prázdna hodnota
BP_MVSR_RA_009	100 %	0 %	20	0	ItemName <> NULL alebo prázdna hodnota
BP_MVSR_RA_017	100 %	0 %	83	0	objectId <> NULL alebo prázdna hodnota
BP_MVSR_RA_018	100 %	0 %	83	0	versionId <> NULL alebo prázdna hodnota

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_MVSR_RA_0 23	100 %	0 %	83	0	ItemCode <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 25	100 %	0 %	83	0	ItemName <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 35	100 %	0 %	3 039	0	objectId <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 36	100 %	0 %	3 039	0	versionId <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 41	100 %	0 %	3 039	0	ItemCode <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 43	100 %	0 %	3 039	0	ItemName <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 51	100 %	0 %	3 039	0	countyIdentifier <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 53	100 %	0 %	3 039	0	objectId <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 54	100 %	0 %	7 240	0	objectId <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 55	100 %	0 %	7 240	0	versionId <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 60	100 %	0 %	7 240	0	ItemCode <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 62	100 %	0 %	7 240	0	ItemName <> NULL alebo prázdna hodnota



ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_MVSR_RA_0 70	100 %	0 %	7 240	0	municipalityIdentifier <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 72	100 %	0 %	52 203	0	objectId <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 73	100 %	0 %	52 203	0	versionId <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 81	100 %	0 %	52 203	0	municipalityIdentifier <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 87	100 %	0 %	3 470 290	0	objectId <> NULL alebo prázdna hodnota
BP_MVSR_RA_0 88	100 %	0 %	3 470 290	0	versionId <> NULL alebo prázdna hodnota
BP_MVSR_RA_1 05	100 %	0 %	2 000 581	0	municipalityIdentifier <> NULL alebo prázdna hodnota
BP_MVSR_RA_1 08	100 %	0 %	6 476 963	0	objectId <> NULL alebo prázdna hodnota
BP_MVSR_RA_1 09	100 %	0 %	6 476 963	0	versionId <> NULL alebo prázdna hodnota
BP_MVSR_RA_1 15	100 %	0 %	2 708 433	13	BuildingNumber na 1 - 4 mieste môže byť prirodzené číslo, na 5 miestne môže byť text veľké písmeno abecedy okrem CH a DZ a bez diakritiky
BP_MVSR_RA_1 22	100 %	0 %	6 476 963	0	PropertyRegistrationNumberIdentifier <> NULL alebo prázdna hodnota
BP_MVSR_RA_1 29	100 %	0 %	166 943	0	objectId <> NULL alebo prázdna hodnota

ID biznis pravidla	Zhoda (%)	Nezhoda (%)	Zhoda (počet)	Nezhoda (počet)	Biznis pravidlo
BP_MVSR_RA_1 30	100 %	0 %	166 943	0	versionId <> NULL alebo prázdna hodnota

**BP\_MVSR\_RA\_033** – Identifikátor regiónu nesmie byť NULL alebo obsahovať prázdnu hodnotu. (*regionIdentifier <> NULL alebo prázdna hodnota*)

Kontrola bola vykonaná na 83 záznamoch tabuľky 02 COUNTY. Cieľom bolo zistiť, či kontrolované záznamy majú v atribúte identifikátora regiónu (*regionIdentifier*) priradený údaj.

#### Tabuľka 39 Výsledok merania pravidla BP\_MVSR\_RA\_033

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
1,2 %	98,8 %	1	82

Zistenie: Kompletnosť atribútu „regionIdentifier“ v tabuľke 02 COUNTY je 1,2 %. Zistený výsledok sa nezhoduje s údajmi v databáze registra. Meranie je nutné opakovať.

**BP\_MVSR\_RA\_093** – Evidenčné číslo nehnuteľnosti nesmie byť NULL alebo obsahovať prázdnu hodnotu.

Kontrola bola vykonaná na 2 000 581 záznamoch tabuľky 06 BUILDING. Cieľom bolo zistiť, či kontrolované záznamy majú v atribúte evidenčné číslo nehnuteľnosti (PropertyRegistrationNumber) priradený údaj.

#### Tabuľka 40 Výsledok merania pravidla BP\_MVSR\_RA\_093

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,65 %	0,35 %	1993668	6913

Zistenie: Kompletnosť atribútu „PropertyRegistrationNumber“ v tabuľke 06 BUILDING je 0,35 %. Zistené nezahody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_100** – Kód položky účel stavby nesmie byť NULL alebo obsahovať prázdnu hodnotu.

Kontrola bola vykonaná na 2 000 581 záznamoch v tabuľke 06 BUILDING (vrátane historických). Cieľom bolo zistiť, či kontrolované záznamy majú uvedený kód položky účel stavby (atribút „BuildingPurpose\_ItemCode“).

#### Tabuľka 41 Výsledok merania pravidla BP\_MVSR\_RA\_100

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
15,86 %	84,14 %	317 335	1 683 246

Zistenie: Kompletnosť atribútu „BuildingPurpose\_ItemCode“ v tabuľke 06 BUILDING je 15,86 %. Hoci v meraní boli zahrnuté aj neplatné údaje zistený počet nezhôd prevyšuje predpokladaný počet (v RA je približne 1,7 milióna platných záznamov adries).

**BP\_MVSR\_RA\_101** – Názov položky účel stavby nesmie byť NULL alebo obsahovať prázdnu hodnotu.

Kontrola bola vykonaná na 2 000 581 záznamoch adries. Cieľom bolo zistiť, či kontrolované záznamy majú v názov položky účel stavby priradený údaj.

#### Tabuľka 42 Výsledok merania pravidla BP\_MVSR\_RA\_101

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
15,86 %	84,14 %	317 335	1 683 246

Zistenie: Kompletnosť atribútu „BuildingPurpose\_ItemName“ v tabuľke 06 BUILDING je 15,86 %. V meraní boli zahrnuté aj historické údaje. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_103** – Kód položky kód typu budovy nesmie byť NULL alebo obsahovať prázdnu hodnotu.

Kontrola bola vykonaná na 2 000 581 záznamoch adries. Cieľom bolo zistiť, či kontrolované záznamy majú v kód položky kód typu budovy priradený údaj.

#### Tabuľka 43 Výsledok merania pravidla BP\_MVSR\_RA\_103

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,71 %	0,29 %	1 994 871	5 710

Zistenie: Kompletnosť atribútu „BuildingTypeCode\_ItemCode“ v tabuľke 06 BUILDING je 99,71 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_104** – Názov položky kód typu stavby nesmie byť NULL alebo obsahovať prázdnu hodnotu.

Kontrola bola vykonaná na 2 000 581 záznamoch adries. Cieľom bolo zistiť, či kontrolované záznamy majú v názov položky kód typu stavby priradený údaj.

#### Tabuľka 44 Výsledok merania pravidla BP\_MVSR\_RA\_104

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,71 %	0,29 %	1 994 871	5 710

Zistenie: Kompletnosť atribútu „BuildingPurpose\_ItemName“ v tabuľke 06 BUILDING je 99,71 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_113** – Číslo budovy musí byť väčšie alebo menšie od nuly, ale môže nadobúdať aj hodnotu NULL v prípade ak pre dané smerové číslo nie je určené orientačné číslo a existuje väzba na identifikačné číslo nehnuteľnosti, ktorá nie je rovná hodnote NULL.

Kontrola bola vykonaná na 2 708 446 záznamoch adries. Cieľom bolo zistiť, či každý meraný záznam obsahuje číslo budovy, ktoré je väčšie alebo menšie ako nula, alebo v prípade splnenia podmienky nadobúda hodnotu NULL.

#### Tabuľka 45 Výsledok merania pravidla BP\_MVSR\_RA\_113

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,98 %	0,02 %	2 707 832	614

Zistenie: Kompletnosť atribútu „BuildingNumber“ v tabuľke 07 BUILDING\_NUMBER je 99,71 %. Zistené percento nekvalitných údajov pravdepodobne predstavuje nevyplnené údaje v referenčných záznamoch, ktoré bude potrebné doplniť.

**BP\_MVSR\_RA\_120** – Poštové smerovacie číslo nesmie byť NULL alebo obsahovať prázdnu hodnotu.

Kontrola bola vykonaná na 2 708 446 záznamoch adries. Cieľom bolo zistiť, či kontrolované záznamy majú pre poštové smerovacie číslo priradený údaj.

#### Tabuľka 46 Výsledok merania pravidla BP\_MVSR\_RA\_120

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
54,93 %	45,07 %	1 487 642	1 220 804

Zistenie: Kompletnosť atribútu „PostalCode“ v tabuľke 07 BUILDING\_NUMBER je 54,93 %. Zistené percento je podstatne nižšie, než bol predpoklad a nezhoduje sa s údajmi z databázy registra. Meranie bude nutné opakovať.

**BP\_MVSR\_RA\_127** – Zemepisná dĺžka – GPS súradnica je povinný údaj a teda hodnota nesmie byť NULL.

Kontrola bola vykonaná na 2 708 446 záznamoch adries. Cieľom bolo zistiť, či kontrolované záznamy majú vyplnený povinný údaj zemepisnej dĺžky (GPS súradnice).

#### Tabuľka 47 Výsledok merania pravidla BP\_MVSR\_RA\_127

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
83,25 %	16,75 %	2 254 842	453 604

Zistenie: Kompletnosť atribútu „AxisL“ v tabuľke 07 BUILDING\_NUMBER je 83,25 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_128** – Zemepisná šírka – GPS súradnica je povinný údaj a teda hodnota nesmie byť NULL.

Kontrola bola vykonaná na 2 708 446 záznamoch adries. Cieľom bolo zistiť, či kontrolované záznamy majú vyplnený povinný údaj zemepisnej šírky (GPS súradnice).

#### Tabuľka 48 Výsledok merania pravidla BP\_MVSR\_RA\_128

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
83,25 %	16,75 %	2 254 842	453 604

Zistenie: Kompletnosť atribútu „AxisB“ v tabuľke 07 BUILDING\_NUMBER je 83,25 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

#### 2.2.3.4 Unikátnosť

**BP\_MVSR\_RA\_119** – Každá adresa má priradený práve jeden unikátny identifikátor adresy pre objekty kde validTo > ako aktuálny dátum.

Kontrola bola vykonaná na 2 708 446 záznamoch (vrátane historických). Cieľom bolo zistiť, či sú v RA duplicitne evidované záznamy (jedna adresa s viacerými unikátnymi identifikátormi adresy).

**Tabuľka 49 Výsledok merania pravidla BP\_MVSR\_RA\_119**

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
64,43 %	35,57 %	963 504	1 744 942

Zistenie: Unikátnosť záznamov v registri je 64,43 %. Vzhľadom na to, že meranie prebehlo nad všetkými záznamami indexu (vrátane historických), nie je možné posúdiť percento unikátnosti referenčných údajov RA.

#### 2.2.3.5 Konzistentnosť

Kontrola konzistentnosti RA bola vykonaná meraním na 9 biznis pravidlách. Merané bolo biznis pravidlo, že záznam podradeného objektu nesmie obsahovať hodnotu neplatného nadradeného objektu. Meranie bolo vykonané na všetkých záznamoch v tabuľkách 03 MUNICIPALITY, 04 DISTRICT, 05 STREET, 06 BUILDING a 07 BUILDING\_NUMBER, teda nielen na referenčných údajoch ale aj na historických. Nižšie sú uvedené výsledky merania.

**BP\_MVSR\_RA\_052** – countyIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 02 COUNTY

Meranie bolo vykonané na 3 039 záznamoch v tabuľke 03 MUNICIPALITY. Meranie zistilo identifikátor nadradeného objektu (atribút „countyIdentifier,“) v zázname tabuľky 03 MUNICIPALITY, vyhľadáním zistenej hodnoty v atribúte „objectId“ v tabuľke 02 COUNTY našlo záznam nadradeného objektu a overilo, či je platný.

**Tabuľka 50 Výsledok merania pravidla BP\_MVSR\_RA\_052**

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
98,91 %	1,09 %	3 006	33

Zistenie: Konzistentnosť atribútu „countyIdentifier“ v tabuľke 03 MUNICIPALITY je 98,91 %. Zistené nezahody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_071** – municipalityIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 03 MUNICIPALITY

Meranie bolo vykonané na 7 240 záznamoch v tabuľke 04 DISTRICT. Meranie zistilo identifikátor nadradeného objektu (atribút „municipalityIdentifier,“) v zázname tabuľky 04 DISTRICT, vyhľadáním zistenej hodnoty v atribúte „objectId“ v tabuľke 03 MUNICIPALITY našlo záznam nadradeného objektu a overilo, či je platný.

#### Tabuľka 51 Výsledok merania pravidla BP\_MVSR\_RA\_071

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,81 %	0,19 %	7 226	14

Zistenie: Konzistentnosť atribútu „municipalityIdentifier“ v tabuľke 04 DISTRICT je 98,91 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_085** – municipalityIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 03 MUNICIPALITY

Meranie bolo vykonané na 30 006 záznamoch v tabuľke 05 STREET. Meranie zistilo identifikátor nadradeného objektu (atribút „municipalityIdentifier,“) v zázname tabuľky 05 STREET, vyhľadáním zistenej hodnoty v atribúte „objectId“ v tabuľke 03 MUNICIPALITY našlo záznam nadradeného objektu a overilo, či je platný.

#### Tabuľka 52 Výsledok merania pravidla BP\_MVSR\_RA\_085

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
98,75 %	1,25 %	29 632	374

Zistenie: Konzistentnosť atribútu „municipalityIdentifier“ v tabuľke 05 STREET je 98,75 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_086** – districtIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 04 DISTRICT

Meranie bolo vykonané na 30 006 záznamoch v tabuľke 05 STREET. Meranie zistilo identifikátor nadradeného objektu (atribút „districtIdentifier,“) v zázname tabuľky 05 STREET, vyhľadáním zistenej hodnoty v atribúte „objectId“ v tabuľke 04 DISTRICT našlo záznam nadradeného objektu a overilo, či je platný.

### Tabuľka 53 Výsledok merania pravidla BP\_MVSR\_RA\_086

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
51,95 %	48,05 %	15 589	14 417

Zistenie: Konzistentnosť atribútu „districtIdentifier“ v tabuľke 05 STREET je 51,95 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_106** – municipalityIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 03 MUNICIPALITY

Meranie bolo vykonané na 2 000 581 záznamoch v tabuľke 06 BUILDING. Meranie zistilo identifikátor nadradeného objektu (atribút „municipalityIdentifier,“) v zázname tabuľky 06 BUILDING, vyhľadáním zistenej hodnoty v atribúte „objectId“ v tabuľke 03 MUNICIPALITY našlo záznam nadradeného objektu a overilo, či je platný.

### Tabuľka 54 Výsledok merania pravidla BP\_MVSR\_RA\_106

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,73 %	0,27 %	1 995 156	5 425

Zistenie: Konzistentnosť atribútu „municipalityIdentifier“ v tabuľke 06 BUILDING je 99,73 %. Zistené nezhody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_107** – districtIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 04 DISTRICT

Meranie bolo vykonané na 2 000 581 záznamoch v tabuľke 06 BUILDING. Meranie zistilo identifikátor nadradeného objektu (atribút „districtIdentifier,“) v zázname tabuľky 06 BUILDING, vyhľadáním zistenej hodnoty v atribúte „objectId“ v tabuľke 04 DISTRICT našlo záznam nadradeného objektu a overilo, či je platný.

### Tabuľka 55 Výsledok merania pravidla BP\_MVSR\_RA\_107

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
43,44 %	56,56 %	869 114	1 131 467



Zistenie: Konzistentnosť atribútu „districtIdentifier“ v tabuľke 06 BUILDING je 43,44 %. Zistené nezahody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_123** – PropertyRegistrationNumberIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu PropertyRegistrationNumber z 06 BUILDING

Meranie bolo vykonané na 2 708 446 záznamoch v tabuľke 07 BUILDING\_NUMBER. Meranie zistilo identifikátor nadradeného objektu (atribút „PropertyRegistrationNumberIdentifier,“) v zázname tabuľky 07 BUILDING\_NUMBER, vyhľadáním zistenej hodnoty v atribúte „PropertyRegistrationNumberIdentifier“ v tabuľke 06 BUILDING našlo záznam nadradeného objektu a overilo, či je platný.

#### Tabuľka 56 Výsledok merania pravidla BP\_MVSR\_RA\_123

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,99 %	0,01 %	2 708 305	141

Zistenie: Konzistentnosť atribútu „PropertyRegistrationNumberIdentifier“ v tabuľke 07 BUILDING\_NUMBER je 99,99 %. Zistené nezahody boli spôsobené meraním na všetkých údajoch, vrátane údajov, ktoré nie sú referenčné (neplatné údaje).

**BP\_MVSR\_RA\_124** – streetNameIdentifier nesmie obsahovať hodnotu neplatného nadradeného objektu objectId z 05 STREET

Meranie bolo vykonané na 2 708 446 záznamoch v tabuľke 07 BUILDING\_NUMBER. Meranie zistilo identifikátor nadradeného objektu (atribút „streetNameIdentifier,“) v zázname tabuľky 07 BUILDING\_NUMBER, vyhľadáním zistenej hodnoty v atribúte „objectId“ v tabuľke 05 STREET našlo záznam nadradeného objektu a overilo, či je platný.

#### Tabuľka 57 Výsledok merania pravidla BP\_MVSR\_RA\_124

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
36,90 %	63,10 %	999 296	1 709 150

Zistenie: Konzistentnosť atribútu „streetNameIdentifier“ v tabuľke 07 BUILDING\_NUMBER je 36,9 %.

### 2.2.3.6 Referenčná integrita

**BP\_MVSR\_RA\_118** – BuildingIndex <> NULL alebo prázdna hodnota

Kontrola bola vykonaná na 6 476 963 záznamoch. Meranie overovalo splnenie podmienky, že každému záznamu je pridelený jedinečný identifikátor adresy, teda referenčný identifikátor (alfanumerický reťazec zložený z číslic a písmena A, ktoré môže byť na poslednom mieste).

**Tabuľka 58 Výsledok merania pravidla BP\_MVSR\_RA\_118**

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
100 %	0 %	6 476 963	0

Zistenie: Meraním bola potvrdená 100-percentná referenčná integrita RA.

## 2.2.4 Zhrnutie zistení

Pri väčšine biznis pravidiel bolo meranie dátovej kvality RA vykonané na všetkých údajoch uložených v perzistentnej vrstve CSRÚ, teda nielen na referenčných, ale aj na neplatných údajoch, čo v mnohých prípadoch skreslilo výsledky merania. Nakoľko z kapacitných dôvodov nebolo možné meranie zopakovať, pri vyhodnotení každého biznis pravidla pracovná skupina overovala aj to, či bol výsledok ovplyvnený nezgodami v neplatných údajoch. V rámci tejto kapitoly je venovaná pozornosť zisteniam, keď bola identifikovaná nekvalita v referenčných údajoch. Tie sú nižšie vyhodnotené podľa jednotlivých parametrov dátovej kvality.

**Presnosť** – v rámci merania DK z hľadiska presnosti bolo meraných 45 biznis pravidiel na tabuľkách 01 REGION, 02 COUNTY, 03 MUNICIPALITY, 04 DISTRICT, 05 STREET, 06 BUILDING, 07 BUILDING\_NUMBER a 08 BUILDING\_UNIT. V prípade 21 biznis pravidiel bola zistená presnosť 100 %. Väčšia alebo menšia miera nepresnosti bola zistená pri 24 biznis pravidlách uvedených v tabuľke nižšie.

**Tabuľka 59 Biznis pravidiel so zistenou presnosťou nižšou než 100 %**

ID BP	Tabuľka Atribút	Zhoda (%)	Nezhoda (počet)
BP_MVSR_RA_004	01 REGION validTo	60 %	12 z 20
BP_MVSR_RA_020	02 COUNTY validTo	95,18 %	4 z 83
BP_MVSR_RA_038	03 MUNICIPALITY validTo	96,64 %	102 z 3 039
BP_MVSR_RA_057	04 DISTRICT validTo	49,54 %	3 653 zo 7 240

ID BP	Tabuľka Atribút	Zhoda (%)	Nezhoda (počet)
BP_MVSR_RA_061	04 DISTRICT ItemCode	50,96 %	1 759 z 3 587
BP_MVSR_RA_075	05 STREET validTo	57,48 %	22 197 z 52 203
BP_MVSR_RA_090	06 BUILDING validTo	45,08 %	1 905 726 z 3 470 290
BP_MVSR_RA_092	06 BUILDING PropertyRegistrationNumber	99,64 %	7 171 z 2 000 581
BP_MVSR_RA_094	06 BUILDING PropertyRegistrationNumber	99,65 %	6 913 z 2 000 581
BP_MVSR_RA_095	06 BUILDING PropertyRegistrationNumber	99,65 %	6 917 z 2 000 581
BP_MVSR_RA_098	06 BUILDING PropertyRegistrationNumber	42,93 %	1 885 147 z 3 303 170
BP_MVSR_RA_111	07 BUILDING_NUMBER validTo	41,82 %	3 768 517 z 6 476 963
BP_MVSR_RA_116	07 BUILDING_NUMBER BuildingNumber	99,96 %	276 z 634 708
BP_MVSR_RA_121	07 BUILDING_NUMBER PostalCode	54,93 %	1 220 804 z 2 208 446
BP_MVSR_RA_125	07 BUILDING_NUMBER AxisL	83,25 %	453 669 z 2 208 446
BP_MVSR_RA_126	07 BUILDING_NUMBER AxisB	83,25 %	453 683 z 2 208 446
BP_MVSR_RA_132	08 BUILDING_UNIT validTo	90,42 %	15 996 z 166 943
BP_MVSR_RA_134	08 BUILDING_UNIT BuildingUnit_UnitNumber	99,98 %	30 zo 150 947

ID BP	Tabuľka Atribút	Zhoda (%)	Nezhoda (počet)
BP_MVSR_RA_138	08 BUILDING_UNIT BuildingUnit_Floor	94,79 %	7 858 zo 150 947
BP_MVSR_RA_139	08 BUILDING_UNIT BuildingUnit_Floor	94,79 %	7 858 zo 150 947
BP_MVSR_RA_140	08 BUILDING_UNIT BuildingUnit_Floor	94,79 %	7 858 zo 150 947
BP_MVSR_RA_141	08 BUILDING_UNIT BuildingUnit_Floor	94,79 %	7 858 zo 150 947
BP_MVSR_RA_142	08 BUILDING_UNIT BuildingUnit_Floor	94,79 %	7 858 zo 150 947
BP_MVSR_RA_145	08 BUILDING_UNIT BuildingNumberIdentifier	94,79 %	3 106 zo 150 947

Ako potvrdil správca registra, všetky vyššie uvedené prípady zníženej kvality boli spôsobené skreslením spôsobeným tým, že meranie bolo vykonané aj na neplatných údajoch.

**Správnosť** – v rámci kontroly správnosti RA bolo meraných 28 biznis pravidiel v tabuľkách 01 REGION, 02 COUNTY, 03 MUNICIPALITY, 04 DISTRICT a 05 STREET. Kontrola bola zameraná na použitie nepovolených znakov, nadbytočné medzery a na dodržanie minimálneho alebo maximálneho počtu znakov. V prípade 23 pravidiel bola zistená správnosť 100 %. Pri 5 biznis pravidlách uvedených v tabuľke nižšie meranie zistilo nezhody v použití nepovolených znakov a nadbytočných medzier.

#### Tabuľka 60 Biznis pravidiel so zistenou správnosťou nižšou než 100 %

ID BP	Tabuľka Atribút	Zhoda (%)	Nezhoda (počet)
BP_MVSR_RA_012	01 REGION ItemName	20 %	16 z 20
BP_MVSR_RA_028	02 COUNTY ItemName	38,55 %	32 z 83
BP_MVSR_RA_065	04 DISTRICT ItemName	26,92 %	5 291 zo 7 240

ID BP	Tabuľka Atribút	Zhoda (%)	Nezhoda (počet)
BP_MVSR_RA_077	05 STREET ItemName	26,92 %	5 291 zo 7 240
BP_MVSR_RA_082	05 STREET ItemName	98,9 %	572 z 52 203

Správca registra potvrdil, že výsledky merania za pravidlá BP\_MVSR\_RA\_028, BP\_MVSR\_RA\_065 a BP\_MVSR\_RA\_082 boli spôsobené meraním na neplatných údajoch. Meranie pravidiel BP\_MVSR\_RA\_012 a BP\_MVSR\_RA\_077 však zistilo použitie neplatných znakov v atribúte ItemName v tabuľke 01 REGION a v tabuľke 05 STREET.

**Kompletnosť** – v rámci kontroly správnosti RA bolo meraných 41 biznis pravidiel v tabuľkách 01 REGION, 02 COUNTY, 03 MUNICIPALITY, 04 DISTRICT, 05 STREET, 06 BUILDING, 07 BUILDING\_NUMBER a 08 BUILDING\_UNIT. Kontrola bola zameraná vyplnenosť povinných atribútov. V prípade 31 pravidiel bola potvrdená kompletnosť 100 %. V tabuľke nižšie sú je uvedených 10 biznis pravidiel, ktoré zistili nedostatočnú kompletnosť údajov.

#### Tabuľka 61 Biznis pravidlá so zistenou kompletnosťou nižšou než 100 %

ID BP	Tabuľka Atribút	Zhoda (%)	Nezhoda (počet)
BP_MVSR_RA_033	02 COUNTY regionIdentifier	1,2 %	82 z 83
BP_MVSR_RA_093	06 BUILDING PropertyRegistrationNumber	99,65 %	6 913 z 2 000 581
BP_MVSR_RA_100	06 BUILDING BuildingPurpose_ItemCode	15,86 %	1 683 246 z 2 000 581
BP_MVSR_RA_101	06 BUILDING BuildingPurpose_ItemName	15,86 %	1 683 246 z 2 000 581
BP_MVSR_RA_103	06 BUILDING BuildingPurpose_ItemName	99,71 %	5 710 z 2 000 581
BP_MVSR_RA_104	06 BUILDING BuildingPurpose_ItemName	99,71 %	5 710 z 2 000 581

ID BP	Tabuľka Atribút	Zhoda (%)	Nezhoda (počet)
BP_MVSR_RA_113	07 BUILDING_NUMBER BuildingNumber	99,98 %	614 z 2 708 446
BP_MVSR_RA_120	07 BUILDING_NUMBER PostalCode	54,93 %	1 220 804 z 2 708 446
BP_MVSR_RA_127	07 BUILDING_NUMBER AxisL	83,25 %	453 604 z 2 708 446
BP_MVSR_RA_128	07 BUILDING_NUMBER AxisB	83,25 %	453 604 z 2 708 446

Správca registra potvrdil, že väčšina vyššie uvedených výsledkov bola spôsobená meraním na neplatných údajoch, ale meranie odhalilo aj niekoľko potenciálnych problémov.

Výsledok merania pravidla BP\_MVSR\_RA\_033 bol spôsobený chybou pri meraní, čo potvrdila kontrola v databáze registra RA. Toto meranie bude potrebné opakovať a zistiť presný stav DK atribútu „regionIdentifier“ v tabuľke 02 COUNTY. Rovnako bola identifikovaná chyba merania na atribúte „BuildingPurpose\_ItemCode“ v tabuľke 06 BUILDING a atribútu „PostalCode“ v tabuľke 07 BUILDING\_NUMBER, pri ktorých boli zistená nízka kompletnosť aj napriek tomu, že kompletnosť v databáze registra je takmer 100 %. Bude potrebné skontrolovať dávku prenosu medzi RA a CSRÚ a opakovaným meraním overiť kompletnosť datasetu.

V prípade výsledku merania pravidla BP\_MVSR\_RA\_113 správca registra potvrdil, že zistené percento 0,29 % nekvalitných údajov pravdepodobne predstavuje nevyplnené údaje v referenčných záznamoch, ktoré bude potrebné doplniť.

**Unikátnosť** – tento parameter bol overený meraním 1 biznis pravidla na atribúte „BuildingIndex“ v tabuľke 07 BUILDING\_NUMBER. Meranie zistilo, že unikátnosť všetkých adries v RA je 64,43 %, konkrétne 1 744 942 nezhodných adries z celkových 2 708 446 záznamov v tabuľke. Vzhľadom na to, že meranie prebehlo nad všetkými záznamami indexu (vrátane historických), nie je možné posúdiť skutočné percento duplicit v referenčných údajoch RA.

**Konzistentnosť** – kontrola konzistentnosti RA bola vykonaná meraním na 9 atribútoch v tabuľkách 03 MUNICIPALITY, 04 DISTRICT, 05 STREET, 06 BUILDING a 07 BUILDING\_NUMBER. Žiadne z nich nedosiahlo konzistentnosť 100 %. Správca registra ale potvrdil, že k skresleniu došlo na základe merania aj na neplatných údajoch.

## Tabuľka 62 Biznis pravidlá so zistenou konzistentnosťou nižšou než 100 %

ID BP	Tabuľka Atribút	Zhoda (%)	Nezhoda (počet)
BP_MVSR_RA_052	03 MUNICIPALITY countyIdentifier	98,91 %	33 z 3 039
BP_MVSR_RA_071	04 DISTRICT municipalityIdentifier	99,81 %	41 z 7 240
BP_MVSR_RA_085	05 STREET municipalityIdentifier	98,75 %	374 z 30 006
BP_MVSR_RA_086	05 STREET districtIdentifier	51,95 %	14 417 z 30 006
BP_MVSR_RA_106	06 BUILDING municipalityIdentifier	99,73 %	5 425 z 2 000 581
BP_MVSR_RA_107	06 BUILDING districtIdentifier	43,44 %	1 131 467 z 2 000 581
BP_MVSR_RA_123	07 BUILDING_NUMBER PropertyRegistrationNumberIdentifier	99,99 %	141 z 2 708 446
BP_MVSR_RA_124	07 BUILDING_NUMBER streetNameIdentifier	36,90 %	1 709 150 z 2 708 446

**Aktuálnosť** – tento kvalitatívny parameter, ktorý posudzuje procesy súvisiace s vytváraním dát, nebol predmetom merania DK, ale je ho možné posúdiť na základe informácií o procese zapisovania dát do registra, ktoré autorom dokumentu poskytol správca registra. Každý nový záznam alebo aktualizácia existujúceho záznamu je verejne k dispozícii momentom uloženia, takže z pohľadu referenčných údajov je aktuálnosť údajov vždy okamžitá. Jediným neskôr zapísaným údajom je PSČ. PSČ nie je referenčný údaj RA, ide o doplňujúci údaj adresy, ktorý je v správe Slovenskej pošty. Slovenská pošta dostáva denne v dávkach dataset nových a aktualizovaných záznamov, do ktorých dopĺňa aktuálny údaj PSČ. Doplnená dávka PSČ je následne zapísaná do RA. Tento krok trvá maximálne 5 dní, ale zvyčajne nie dlhšie než 2 dni.

**Strojová spracovateľnosť** – tento parameter nebol predmetom vykonaného merania. Môžeme však konštatovať, že v súčasnosti sú údaje RA poskytované v štruktúrovanej podobe a s jedinečným referenčným identifikátorom, čo zaraďuje RA na úroveň 4 hviezdíčiek.

**Referenčná integrita** – referenčná integrita bola zisťovaná meraním biznis pravidla BP\_MVSR\_RA\_119. Meraním na platných záznamoch bola potvrdená referenčná integrita 100 %.

Na základe stretnutí so správcom RA a výsledkov merania môžeme konštatovať, že hoci správca registra nemá v súčasnosti formálne prijatú vlastnú metodiku riadenia DK, pri správe registra vychádza z Metodiky riadenia DK z roku 2019, ktorú publikovala Dátová kancelária. Kvalitu nových a aktualizovaných referenčných údajov zabezpečujú automatizované kontroly vo formulároch na vkladanie údajov, ktoré zamedzujú vstupu nekvalitných údajov do registra. MV SR sprístupnilo verejnosti formulár na pripomienkovanie DK RA a navyše vykonáva pravidelné kampane čistenia historických údajov, ktorými postupne zvyšuje DK aj na historických údajoch. Bohužiaľ, meranie, ktoré bolo vykonané v rámci vypracovania tohto dokumentu, kontrolovalo kvalitu nielen referenčných, ale aj už neplatných údajov RA, čo skreslilo výsledky merania. Meranie navyše zistilo potenciálny nesúlad medzi databázou RA a meraným datasetom v CSRÚ. Odporúčame, aby boli vykonané nasledovné kroky:

- overiť správnosť prenosu dávok z RA do CSRÚ,
- aktualizovať biznis pravidlá tak, aby meranie overovalo DK len na referenčných údajoch,
- vykonať opakované meranie, ktoré zistí presný stav DK v RA a v rámci neho vygenerovať aj podrobný zoznam záznamov, v ktorých budú zistené nezhody,
- na základe podrobných zistení z opakovaného merania DK vykonať kampaň na čistenie údajov.

## 2.3 Register nelegálneho zamestnávania (RNZ)

Register nelegálneho zamestnávania napomáha boju proti nelegálnemu zamestnávaniu a zvyšuje transparentnosť na trhu práce. Obsahuje informácie o právnických a fyzických osobách, ktoré porušili zákaz nelegálneho zamestnávania, ako aj o konkrétnych prípadoch nelegálneho zamestnávania. Záznamy obsahujú údaje o evidovaných subjektoch, ako sú ich názov, adresa, identifikačné číslo, údaje o zastúpení, o nelegálnych pracovných činnostiach, o vykonanej kontrole a o uloženej pokute. Údaje v registri sú prístupné verejnosti, čo umožňuje zamestnancom, orgánom verejnej správy a iným zainteresovaným stranám overovať právoplatnosť zamestnania a potenciálny zápis subjektu v registri.

### 2.3.1 Charakteristika RNZ

Správa RNZ a poskytovanie jeho informácií sa riadia nasledujúcou legislatívou:

- Zákon č. 82/2005<sup>8</sup> Z. z. o nelegálnej práci a nelegálnom zamestnávaní a o zmene a doplnení niektorých zákonov v znení neskorších predpisov,
- Zákon č. 125/2006<sup>9</sup> Z. z. o inšpekcii práce a o zmene a doplnení zákona č. 82/2005 Z. z. o nelegálnej práci a nelegálnom zamestnávaní a o zmene a doplnení niektorých zákonov v znení neskorších predpisov,

<sup>8</sup> [82/2005 Z.z. - Zákon o nelegálnej práci a nelegáln... - SLOV-LEX](#)

<sup>9</sup> [125/2006 Z.z. - Zákon o inšpekcii práce a o zmene a... - SLOV-LEX](#)



- Zákon č. 5/2004<sup>10</sup> Z. z. o službách zamestnanosti a o zmene a doplnení niektorých zákonov v znení neskorších predpisov,
- Zákon č. 404/2011<sup>11</sup> Z. z. o pobyte cudzincov a o zmene a doplnení niektorých zákonov,
- Zákon 480/2002<sup>12</sup> Z. z. o azyle a o zmene a doplnení niektorých zákonov,
- Zákon č. 351/2015<sup>13</sup> Z. z. o cezhraničnej spolupráci pri vysielaní zamestnancov na výkon prác pri poskytovaní služieb a o zmene a doplnení niektorých zákonov,
- Smernica Európskeho parlamentu a Rady 2009/52/ES<sup>14</sup> z 18. júna 2009, ktorou sa stanovujú minimálne normy pre sankcie a opatrenia voči zamestnávateľom štátnych príslušníkov tretích krajín, ktorí sa neoprávnene zdržiavajú na území členských štátov sa nachádza (.PDF).

V súlade so § 6 ods. 1 pís. t) Zákona 125/2006, správcom RNZ je Národný inšpektorát práce (NIP), ktorý vedie centrálny verejne prístupný zoznam fyzických osôb a právnických osôb, ktorým bola v predchádzajúcich piatich rokoch uložená pokuta za porušenie zákazu nelegálneho zamestnávania, s uvedením ich obchodného mena, miesta podnikania fyzickej osoby alebo sídla právnickej osoby, identifikačného čísla organizácie a dátumu nadobudnutia právoplatnosti rozhodnutia o uložení pokuty. Národný inšpektorát práce zapíše fyzickú osobu alebo právnickú osobu do tohto zoznamu bezodkladne po nadobudnutí právoplatnosti rozhodnutia o uložení pokuty za porušenie zákazu nelegálneho zamestnávania.

Štruktúra súčasnej verzie RNZ je jednoduchá. Do registra vstupujú informácie zo súčasného Informačného systému ochrany práce (ISOP Národného inšpektorátu práce) a údaje z ÚPSVaR. Nové záznamy do registra zapisujú určené zamestnanci Inšpektorátov práce (IP) do lokálnej databázy. Pri zápise každého záznamu sú údaje subjektu (zamestnávateľa, t. j. právnickej alebo fyzickej osoby) na základe IČO subjektu overené voči lokálnej kópii Štatistického registra organizácií<sup>15</sup> (ŠRO). Každý vytvorený záznam ostáva neaktívny, kým nie je doplnený atribút právoplatnosti pokuty. V pravidelných týždenných intervaloch sú všetky aktívne záznamy (záznamy s uvedeným dátumom právoplatnosti pokuty) prenesené z 9 lokálnych databáz do centrálnej databázy v správe NIP. Hoci centrálna databáza uchováva všetky záznamy od jej založenia, do webového rozhrania<sup>16</sup> dostupného pre verejnosť sú zverejňované len záznamy za ostatných 5 rokov.

NIP v súčasnosti realizuje projekt implementácie nového informačného systému SAWO, v ktorom bude prevádzkovaný RNZ. Systém bude integrovaný na CSRÚ.

<sup>10</sup> [5/2004 Z.z. - Zákon o službách zamestnanosti a o ... - SLOV-LEX](#)

<sup>11</sup> [404/2011 Z.z. - Zákon o pobyte cudzincov a o zmene ... - SLOV-LEX](#)

<sup>12</sup> [480/2002 Z.z. - Zákon o azyle a o zmene a doplnení ... - SLOV-LEX](#)

<sup>13</sup> [351/2015 Z.z. - Zákon o cezhraničnej spolupráci pri... - SLOV-LEX](#)

<sup>14</sup> [CL2009L0052SK0000010.0001.3bi\\_cp 1..1 \(europa.eu\)](#)

<sup>15</sup> Register organizácií vedený Štatistickým úradom SR podľa zákona č. 540/2001 Z.z. o štátnej štatistike v znení neskorších predpisov

<sup>16</sup> [Zoznam fyzických osôb a právnických osôb, ktoré porušili zákaz nelegálneho zamestnávania \(Zákon č. 82/2005 Z. z.\) \(gov.sk\)](#)

## 2.3.2 Procesy riadenia kvality dát RNZ

NIP nemá vypracovanú metodiku riadenia dátovej kvality, ani integrované postupy riadenia dátovej kvality do procesov organizácie. Podľa vyjadrenia dátového kurátora NIP, dátová kvalita je riadená párovaním zistení IP a UPSVaR s databázou ŠRO (overenie pri vytvorení záznamu) a tým, že sa zverejnia len subjekty, ktoré „spĺňajú“ podmienky: overenie IČO (v databáze ŠRO) a dátum právoplatnosti uloženej pokuty (z databáz NIP a ÚPSVaR). Okrem overenia subjektu voči lokálnej kópii ŠRO má formulár na vkladanie nových záznamov implementované aj kontroly formátu dátumu, čísla a vyplnenosti povinných polí. Z toho dôvodu považuje NIP dátovú kvalitu RNZ za dostatočnú. Pred migráciou dát z pôvodného registra do nového informačného registra SAWO sa však rozhodli skontrolovať kvalitu svojich dát a vyriešiť prípadné nezrovnalosti. Meranie bolo vykonané v spolupráci s MIRRI a s autormi tohto dokumentu a jeho výsledky sú popísané v nasledujúcej kapitole.

## 2.3.3 Kvalita dát v RNZ

V spolupráci s autormi tohto dokumentu a s Dátovou kanceláriou pripravilo MPSVaR SR a NIP zoznam biznis pravidiel a vykonali meranie v súlade s Metodikou merania DK. Pre potrebu merania bol vytvorený dataset, ktorý bol kombináciou publikovanej verzie RNZ a záznamov z evidencie kontrol jednotlivých inšpektorátov práce. Spojenie bolo nevyhnutné pre kontrolu duplicit evidovaných pokút (BP\_NIP\_RNZ\_015). Dataset bol aktuálny k 20.07.2023 a obsahoval 3 369 záznamov.

Výsledky merania boli posudzované z hľadiska parametrov presnosť, správnosť, kompletnosť, unikátnosť a konzistentnosť. V tabuľke nižšie je uvedené priradenie meraných biznis pravidiel hodnoteným parametrom dátovej kvality.

**Tabuľka 63 Zoznam meraných biznis pravidiel nad RNZ**

ID BP	BP (biznis pravidlo)	Parameter
BP_NIP_RNZ_001	Názov subjektu je povinná hodnota	Kompletnosť
BP_NIP_RNZ_002	Kontrola názvu subjektu voči RPO na základe IČO	Presnosť
BP_NIP_RNZ_003	IČO je povinná hodnota	Kompletnosť
BP_NIP_RNZ_004	Hodnota IČO je 8 znakový reťazec zložený z číselných hodnôt	Správnosť
BP_NIP_RNZ_005	IČO začína číslicou 0, 1, 2, 3, 4, 5.	Presnosť
BP_NIP_RNZ_006	Názov obce je povinná hodnota	Kompletnosť
BP_NIP_RNZ_007	Číslo domu je povinná hodnota	Kompletnosť

ID BP	BP (biznis pravidlo)	Parameter
BP_NIP_RNZ_008	PSČ je povinná hodnota	Kompletnosť
BP_NIP_RNZ_009	Dátum právoplatnosti pokuty je požadovaný v tvare YYYY-MM-DD	Správnosť
BP_NIP_RNZ_010	Zdroj záznamu musí byť jedna z hodnôt NIP alebo UPSVaR	Presnosť
BP_NIP_RNZ_011	PSČ musí byť 5-znakový reťazec zložený z číselných hodnôt	Správnosť
BP_NIP_RNZ_012	Ak má hodnota PSČ posledné dve čísla 01, musí mať vyplnenú hodnotu ulica	Konzistentnosť
BP_NIP_RNZ_013	Dátum právoplatnosti pokuty musí byť >= ako aktuálny dátum -7 rokov	Presnosť
BP_NIP_RNZ_014	Kontrola ulice addr_psc voči RA na základe PSČ	Presnosť
BP_NIP_RNZ_015	Kontrola duplicit na základe name a date_sanction a date_control. Za duplicitu sú považované záznamy, ak majú rovnaký name a date_sanction a dátum ukončenia výkonu kontroly	Unikátnosť
BP_NIP_RNZ_016	Dátum kontroly je požadovaný v tvare YYYY-MM-DD	Správnosť
BP_NIP_RNZ_017	Dátum kontroly je povinná hodnota	Kompletnosť

### 2.3.3.1 Presnosť

#### BP\_NIP\_RNZ\_002 – Kontrola názvu subjektu voči RPO na základe IČO

Kontrola bola vykonaná na všetkých záznamoch. Pri meraní tohto biznis pravidla bol v každom zázname na základe hodnoty atribútu „ico“ zo záznamu v RNZ skontrolovaný názov subjektu v RPO. Názov nájdeného subjektu bol porovnaný s hodnotou atribútu „name“.

#### Tabuľka 64 Výsledok merania pravidla BP\_NIP\_RNZ\_002

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
97,45 %	2,55 %	3 283	86

Zistenie: Atribút „name“ je vyplnený automaticky zo ŠRO na základe hodnoty atribútu „ico“ na štatistický register. Aj napriek tomu údaje atribútu „name“ dosahujú presnosť 97,45 %. Dôvodom zrejmej chybovosti môžu byť zmeny názvu subjektu počas ostatných 6 rokov alebo nekvalita a neaktuálnosť údajov v lokálnej kópii ŠRO.

#### BP\_NIP\_RNZ\_005 – IČO začína číslou 0, 1, 2, 3, 4, 5

Kontrola bola vykonaná na všetkých záznamoch. Meranie zisťovalo syntaktickú presnosť údajov atribútu „ico“.

#### Tabuľka 65 Výsledok merania pravidla BP\_NIP\_RNZ\_005

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
100 %	0 %	3 369	0

Zistenie: Syntaktická presnosť atribútu „ico“ je 100 %.

**BP\_NIP\_RNZ\_010** – Zdroj záznamu musí byť jedna z hodnôt NIP alebo UPSVaR  
Kontrola bola vykonaná na všetkých záznamoch . Všetky záznamy sú vytvárané v lokálnych databázach na jednotlivých inšpektorátoch práce alebo na UPSVaR SR a následne sú raz týždenne prenášané do centrálnej databázy registra RNZ. Každý záznam by mal mať pri vytvorení automaticky vyplnený atribút „source“ jednou z dvoch povolených hodnôt. Meranie kontrolovalo presnosť vyplnenia tohto atribútu.

#### Tabuľka 66 Výsledok merania pravidla BP\_NIP\_RNZ\_010

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
86,23 %	13,77 %	2 905	464

Zistenie: Aj napriek tomu, že ide o atribút, ktorý je vyplnený automaticky, bola zistená chýbajúca alebo nesprávna hodnota v 464 záznamoch. Podľa správcu registra, predpokladanou príčinou je to, že dataset pre meranie nebol vytvorený exportom z databázy registra, ale prevzatím datasetu, ktorý je sprístupnený verejnosti vo webovom rozhraní. V rámci ďalších procesných krokov prípravy datasetu na meranie do neho neboli správne prenesené hodnoty tohto atribútu.

#### BP\_NIP\_RNZ\_013 – Dátum právoplatnosti pokuty nesmie byť starší ako 7 rokov

Kontrola bola vykonaná na všetkých záznamoch. Meranie kontrolovalo, či hodnota atribútu „date\_sanction“ spadá do povoleného rozsahu.

### Tabuľka 67 Výsledok merania pravidla BP\_NIP\_RNZ\_013

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
100 %	0 %	3 369	0

Zistenie: Presnosť atribútu „date\_sanction“ je 100 %.

### BP\_NIP\_RNZ\_014 – Kontrola ulice addr\_psc voči RA na základe PSČ

Kontrola bola vykonaná na všetkých záznamoch. Pri meraní boli na základe hodnoty atribútu „addr\_psc“ zistené názvy ulíc v Registri adres prislúchajúcich hodnote PSČ. Následne bolo overené, či hodnota atribútu „addr\_street“ spadá do zisteného rozsahu.

### Tabuľka 68 Výsledok merania pravidla BP\_NIP\_RNZ\_014

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
65,69 %	34,31 %	2 213	1 156

Zistenie: Aj napriek tomu, že je názov ulice automaticky vyplnený zo ŠRO na základe IČO subjektu, percentuálna presnosť tohto atribútu je len 65,69 % (1 156 nezhôd z 3 369 záznamov). Jedinou možnou príčinou je nekvalita údajov v lokálnej kópii ŠRO.

## 2.3.3.2 Správnosť

### BP\_NIP\_RNZ\_004 – Hodnota IČO je 8 znakový reťazec zložený z číselných hodnôt

Kontrola bola vykonaná na všetkých záznamoch.

### Tabuľka 69 Výsledok merania pravidla BP\_NIP\_RNZ\_004

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
100 %	0 %	3 369	0

Zistenie: Správnosť atribútu „date\_control“ je 100 %.

### BP\_NIP\_RNZ\_009 – Dátum právoplatnosti pokuty je požadovaný v tvare YYYY-MM-DD

Kontrola bola vykonaná na všetkých záznamoch. Účelom merania bolo zistiť, či formát údajov v atribúte „date\_sanction“ zodpovedá definovaným pravidlám.

#### Tabuľka 70 Výsledok merania pravidla BP\_NIP\_RNZ\_009

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
100 %	0 %	3 369	0

Zistenie: Správnosť atribútu „date\_sanction“ je 100 %.

#### BP\_NIP\_RNZ\_011 – PSČ musí byť 5-znakový reťazec zložený z číselných hodnôt

Kontrola bola vykonaná na všetkých záznamoch.

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
100 %	0 %	3 369	0

Zistenie: Správnosť atribútu „date\_control“ je 100 %.

#### BP\_NIP\_RNZ\_016 – Dátum kontroly je požadovaný v tvare YYYY-MM-DD

Kontrola bola vykonaná na všetkých záznamoch. Účelom merania bolo zistiť, či formát údajov v atribúte „date\_control“ zodpovedá definovaným pravidlám.

#### Tabuľka 71 Výsledok merania pravidla BP\_NIP\_RNZ\_016

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
100 %	0 %	3 369	0

Zistenie: Správnosť atribútu „date\_control“ je 100 %.

### 2.3.3.3 Kompletnosť

V rámci merania dátovej kvality z hľadiska kompletnosti boli vykonaná kontrola vyplnenosti atribútov „name“ (BP\_NIP\_RNZ\_001), „ico“ (BP\_NIP\_RNZ\_003), „addr\_city“ (BP\_NIP\_RNZ\_006), „addr\_number“ (BP\_NIP\_RNZ\_007), „addr\_psc“ (BP\_NIP\_RNZ\_08) a „date\_control“ (BP\_NIP\_RNZ\_0017).

**Tabuľka 72 Výsledky merania kompletnosti povinných atribútov**

Atribút	Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
name	100 %	0 %	3 369	0
ico	100 %	0 %	3 369	0
addr_city	100 %	0 %	3 369	0
addr_number	100 %	0 %	3 369	0
addr_psc	100 %	0 %	3 369	0
date_control	100 %	0 %	3 369	0

Zistenie: Kompletnosť povinných atribútov v RNZ je 100 %.

#### 2.3.3.4 Unikátnosť

**BP\_NIP\_RNZ\_015** – Kontrola duplicit na základe name a date\_sanction a date\_control. Za duplicitu sú považované záznamy, ak majú rovnaký name a date\_sanction a dátum ukončenia výkonu kontroly

Kontrola bola vykonaná na všetkých záznamoch. Cieľom bolo zistiť, či sú v RNZ duplicitne zaznamenané uvalené pokuty. Keďže subjektu mohlo byť udelených viac pokút, ktoré získali právoplatnosť v rovnaký deň, bola pri všetkých pokutách kontrolovaná kombinácia: subjekt, ktorému bola pokuta udelená („name“), dátum, kedy kontrola získala právoplatnosť („date\_sanction“) a dátum kontroly, v rámci ktorej bola pokuta udelená („date\_control“). Potenciálnou duplicitou sú záznamy, ktoré majú rovnaké hodnoty týchto atribútov.

**Tabuľka 73 Výsledok merania pravidla BP\_NIP\_RNZ\_015**

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,82 %	0,18 %	3 363	6

Zistenie: Unikátnosť záznamov v registri je 99,82 %. Bolo nájdených 6 záznamov, ktoré podľa definície biznis pravidla môžu byť navzájom duplicitné. Príčinou môže byť obmedzenie informačného systému registra (maximálny počet nelegálne zamestnaných osôb v jednom zázname).

### 2.3.3.5 Konzistentnosť

**BP\_NIP\_RNZ\_012** – Ak má hodnota PSČ posledné dve čísla 01, musí mať vyplnenú hodnotu ulica

Zhoda (%)	Nezhoda v (%)	Zhoda (počet)	Nezhoda (počet)
99,86 %	0.14 %	1 390	2

Zistenie: Konzistentnosť atribútu „addr\_street“ voči atribútu „addr\_psc“ je 99,86 %. Hodnoty oboch atribútov sú vyplňané automaticky z lokálnej kópie ŠRO na základe IČO.

### 2.3.4 Zhrnutie zistení

Výsledky merania dátovej kvality RNZ podľa definovaných biznis pravidiel sú uvedené v tabuľke nižšie.

**Tabuľka 74 Výsledok merania dátovej kvality v RNZ**

ID BP	Parameter	Zhoda (%)	Nezhoda (%)	Počet nezhôd
BP_NIP_RNZ_001	Kompletnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_002	Presnosť	97,45 %	2,55 %	86 z 3 369
BP_NIP_RNZ_003	Kompletnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_004	Správnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_005	Presnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_006	Kompletnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_007	Kompletnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_008	Kompletnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_009	Správnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_010	Presnosť	86,23 %	13,77 %	464 z 3 369



ID BP	Parameter	Zhoda (%)	Nezhoda (%)	Počet nezhôd
BP_NIP_RNZ_011	Správnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_012	Konzistentnosť	99,86 %	0,14 %	2 z 1 392
BP_NIP_RNZ_013	Presnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_014	Presnosť	65,69 %	34,31 %	1 156 z 3 369
BP_NIP_RNZ_015	Unikátnosť	99,82 %	0,18 %	6 z 3 369
BP_NIP_RNZ_016	Správnosť	100 %	0 %	0 z 3 369
BP_NIP_RNZ_017	Kompletnosť	100 %	0 %	0 z 3 369

**Správnosť** – z hľadiska správnosti je dátová kvalita RNZ 100-percentná. Kontrolou podľa biznis pravidiel zameraných na tento parameter bolo potvrdené, že všetky údaje sú evidované v správnom formáte.

**Kompletnosť** – z hľadiska kompletnosti nebolo meranie jednoznačné. Pri všetkých biznis pravidlách zameraných na vyplnenosť povinných údajov meranie potvrdilo 100-percentnú kompletnosť údajov. Pri kontrole konzistentnosti vyplnenia adresy ulice voči PSČ (atribútu „addr\_street“ voči atribútu „addr\_psc“) však bolo zistené, že v 2 záznamoch, v ktorých bola adresa predpokladaná, nebola vyplnená. Údaj adresy je však automaticky doťahovaný zo ŠRO, čo znamená, že zdrojom prípadnej (percentuálne minimálnej) nekompletnosti nie je RNZ, ale zdrojový register.

**Presnosť** – dátová kvalita z hľadiska tohto parametra bola zisťovaná meraním 5 biznis pravidiel. V troch z nich bol zistený vysoký počet nezhôd. Vo všetkých prípadoch ide o údaje, ktoré sú vyplňané automaticky.

- V prvom prípade bol kontrolovaný názov subjektu podľa IČO voči RPO. Meranie ale zistilo, že v 86 prípadoch bola zistená nezhoda. Tu mohlo dôjsť historicky k zmene názvu kontrolovaného subjektu a zistené nezhody môžu byť len odzrkadlením týchto zmien, keďže RNZ eviduje názov subjektu len v čase právoplatnosti pokuty. Druhou možnou príčinou chybovosti je neaktuálnosť zdrojovej databázy ŠRO. Kópia ŠRO, z ktorej sú údaje doťahované pri vytvorení záznamu, je aktualizovaná každé 2 mesiace, čo môže mať za následok neaktuálnosť doťahovaných údajov o kontrolovanom subjekte.
- V druhom prípade bol kontrolovaný zdroj záznamu, ktorý by mal mať len dve možnosti, NIP alebo UPSVaR. Až 13,77 % údajov (464 záznamov) nemalo ako zdroj uvedenú žiadnu z týchto možností, a to aj napriek skutočnosti, že tento atribút by mal byť vyplnený automaticky pri vytvorení záznamu. Podľa vyjadrenia správcu registra k skresleniu došlo pri exporte datasetu pre meranie a skutočnú úroveň dátovej kvality bude potrebné overiť ďalším meraním.

- Tretie merané biznis pravidlo zisťovalo, či sa podľa RA ulica v adrese kontrolovaného subjektu zhoduje s PSČ. Nezhoda bola zistená až v 34,31 % záznamov (1 156 záznamov). Keďže údaj adresy je automaticky doťahovaný zo ŠRO na základe IČO kontrolovaného subjektu a nie je ho možné meniť manuálne, chybovosť je plne spôsobená údajmi v zdrojovom registri ŠRO.

**Unikátnosť** – pri meraní unikátnosti záznamov RNZ bolo meraním nájdených 6 potenciálne duplicitných záznamov. Duplicita je predpokladá na základe toho, že záznamy o pokute majú rovnaký názov, dátum právoplatnosti kontroly a dátum ukončenia výkonu kontroly. To, či skutočné ide o duplicitné záznamy, je však potrebné overiť a ich počet je nízky. Ich príčinou môže byť skutočnosť, že pri vytvorení záznamu o pokute je evidovaný aj počet nelegálne zamestnaných osôb s maximálnym počtom 99. Pokiaľ bolo pri kontrole zistených viac, než 99 nelegálne zamestnaných osôb, systém automaticky vytvorí viac záznamov o tej istej pokute.

**Konzistentnosť** – ako už bolo uvedené pri hodnotení kompletnosti, v dvoch záznamoch nebol vyplnený atribút názvu ulice, hoci bol podľa PSČ predpokladaný. Keďže oba údaje (PSČ aj názov ulice) sú automaticky doťahované z lokálnej kópie ŠRO, obe nezahody sú spôsobené chybami v zdrojovom registri.

**Aktuálnosť** – tento kvalitatívny parameter, ktorý posudzuje procesy súvisiace s vytváraním dát, nebol predmetom merania DK, ale je ho možné posúdiť na základe informácií, ktoré autorom dokumentu poskytol správca registra.

**Strojová spracovateľnosť** – tento parameter nebol predmetom vykonaného merania. Môžeme však konštatovať, že v súčasnosti sú údaje RNZ publikované v štruktúrovanej podobe, ale nie v otvorenom formáte. Zoznam fyzických osôb a právnických osôb, ktoré porušili zákaz nelegálneho zamestnávania, so sídlom mimo územia Slovenskej republiky, je poskytovaný len vo formáte PDF. To znamená, že v súlade s Metodikou merania DK dosahuje RNZ úroveň 2 hviezdíčiek. Po implementácii jedinečného identifikátora a sprístupnení RNZ prostredníctvom CSRÚ predpokladáme nárast na úroveň 4 hviezdíčiek.

**Referenčná integrita** – hoci záznamy RNZ v súčasnosti neobsahujú jedinečný referenčný identifikátor, správca registra potvrdil úmysel o jeho implementáciu v budúcnosti.

Na základe stretnutí so správcom RNZ a výsledkov merania môžeme konštatovať, že:

- Správca registra nemá v súčasnosti implementované procesy DK a nevykonáva žiadne kontroly DK registra.
- Zo 17 meraných biznis pravidiel meranie zistilo nezahody len v prípade 5 biznis pravidiel. Zistené nezahody boli takmer vo všetkých prípadoch spôsobené neaktuálnosťou alebo nízkou kvalitou záznamov v zdrojovom registri ŠRO.
- V súčasnosti prebieha projekt implementácie nového informačného systému pre RNZ. V rámci implementácie by mali byť odstránené viaceré príčiny zistených nezahôd. Rovnako by mala byť dosiahnutá referenčná integrita a vyššia úroveň strojovej spracovateľnosti RNZ.



## 2.4 Prehľad zistení z výstupu č. 1.1.1 Štandardizácia dátovej kvality

V roku 2019 bola v rámci národného projektu „Zlepšenie využívania údajov vo verejnej správe“ (Zmluva o dielo č. 321/2018) zverejnená Metodika merania dátovej kvality vo verejnej správe<sup>17</sup> (ďalej Metodika merania DK alebo len Metodika). Uvedený národný projekt medzitým pokračuje a trvá do konca roka 2023. Zároveň boli definované odporúčania a opatrenia, ako prispieť k zvýšeniu dátovej kvality na lokálnej (rezortnej) aj na centrálnej úrovni (napr. MIRRI). Plán opatrení bol rozdelený na 4 základné oblasti a ich implementácia je súhrnne vyhodnotená v nasledujúcej tabuľke.

**Tabuľka 75 Celkový stav iniciatív DK podľa oblastí**

Oblasť	Stav implementácie opatrení
<b>Dátové štandardy</b>	Väčšina plánovaných iniciatív prebieha, ale mnohé nie sú na správnej ceste a je potrebné prehodnotiť ich ďalšie smerovanie
<b>Organizačné zabezpečenie</b>	U väčšiny iniciatív nie je zaznamenaný žiadúci pokrok alebo zaostávajú
<b>Procesné zabezpečenie</b>	U väčšiny iniciatív nie je zaznamenaný žiadúci pokrok alebo zaostávajú
<b>Technológie pre DK</b>	Čiastočný pokrok, je potrebné zlepšenie a ďalší rozvoj

Podrobné hodnotenie súčasného stavu v oblasti riadenia dátovej kvality (plnenia definovaných odporúčaní a opatrení) je súčasťou výstupu č. 1.1.1 Štandardizácia dátovej kvality. Na tomto mieste uvádzame len súhrnné zistenia za jednotlivé opatrenia. Súhrnne však môžeme konštatovať, že v prípade väčšiny opatrení ich implementácia zaostáva alebo nezaznamenala žiadny pokrok.

Legenda pre prehľadné hodnotenie použitá pre tabuľky nižšie v stĺpci termín:

<b>Zelená</b>	Opatrenie včas splnené
<b>Žltá</b>	Opatrenie splnené oneskorene alebo sa plní priebežne
<b>Oranžová</b>	Zaostáva za očakávaniami
<b>Červená</b>	Žiaden zaznamenaný pokrok

<sup>17</sup> <https://datalab.digital/wp-content/uploads/Methodika-merania-d%C3%A1tovej-kvality-vo-verejnej-spr%C3%A1ve.pdf>

Tabuľka 76 Centrálna úroveň: Vyhodnotenie realizácie opatrení

Oblasť	Opatrenie	Priorita	Výstup	Termín
Dátové štandardy	Vypracovanie jednotného dátového slovníka pre verejnú správu ako komponent centrálného dátového modelu.	1	Vypracovaný dátový slovník	31.10.2019
Dátové štandardy	Vytvorenie centrálného dátového modelu verejnej správy a definovanie kompetencií a organizačného zabezpečenia dátového modelu.	1	Vypracovaný komplexný dátový model	31.1.2020
Organizácia	Definovanie jasných kompetencií, právomocí a zodpovednosti pre pozície dátových kurátorov.	1	Kompetenčný model	30.11.2019
Organizácia	Zabezpečenie školení pre oblasť riadenia a správy dátovej kvality v rezorte a vybudovanie „call centra“ pre zabezpečenie poradenstva v oblasti dátovej kvality.	1	Zrealizované školenia pre dátových kurátorov	31.12.2019
Proces	Definovanie stratégie pre dátovú kvalitu (vrátane definovania jasných KPIs, ktoré budú následne adoptované oorganizáciami).	1	Stratégia dátovej kvality	30.11.2019
Technológie	Zabezpečenie príslušnej technologickej podpory na riadenie kvality údajov pre všetky definované oblasti vychádzajúce zo stratégie riadenia kvality údajov.	1	Licencie na riadenie kvality údajov k dispozícií	31.12.2019
Dátové štandardy	Zadefinovanie jasných pravidiel pre stotožňovanie a referencovanie údajov medzi registrami a možnosti prepájať informácie medzi registrami tak, aby poskytovali potrebné informácie pre analýzy a riadenie kvality údajov.	2	Pripravená legislatíva	30.6.2020
Organizácia	Posilnenie kompetencie Dátovej kancelárie v oblasti dátovej kvality tak, aby v súčasnosti existujúce pravidlá vedela kontrolovať a vyvodzovať prípadné konsekvencie.	2	Pripravená legislatíva	31.3.2020
Organizácia	Nastavenie pravidiel pre opravy identifikovaných nezrovnalostí v záznamoch tak, aby bol proces čo najkratší a aby bol realizovateľný v momente vzniku alebo odhalenia konzistentnosti v údajoch (viď. opravy záznamov v obchodnom registri, ktoré sú zdrojom pre RPO – chyba sa zistí v RPO).	2	Pripravená legislatíva	30.6.2020
Proces	Neustále monitorovanie dátovej kvality, jej zavedenie v praxi a v pravidelných intervaloch vyhodnocovať a aktualizovať.	2	Report z monitorovania	kontinálne
Proces	Zvedene povinnosti vykonávania pravidelného merania a vyhodnocovania dátovej kvality so zverejňovaním výsledkov vo forme open data.	2	Pripravená legislatíva	31.3.2020
Dátové štandardy	Definovanie oblastí dátovej kvality pre zabezpečenie líderstva v tejto oblasti na úrovni EU.	3	Aktualizovaná stratégia dátovej kvality	30.6.2020
Dátové štandardy	Vypracovanie a udržiavanie dátovej dokumentácie každého rezortu a zabezpečenie jej správy - Treba definovať a zaviesť jednotné dátové štandardy, jednotné formy dátových modelov, jednotné dokumentácie k metadátam, jednotné štruktúry a formy biznis pravidiel.	3	Vypracovaná dokumentácia	31.12.2020
Organizácia	Jasné nastavenie pravidiel hodnotenia a odmeňovania pre zodpovedné osoby za dátovú kvalitu reflektujú aj proaktivitu, ktorá môže viesť k poukázaniu na neakceptovateľný stav v oblasti dátovej kvality tej ktorej organizácie – pravidlo, „ak si vedel o nekvalite a nepovedal si o nej budeš sankcionovaný“.	3	Pripravené pravidlá hodnotenia	31.12.2020

## Tabuľka 77 Rezortná úroveň: Vyhodnotenie realizácie opatrení

Oblasť	Opatrenie	Priorita	Termín (v mes.)
Dátové štandardy	Definovanie presného formátu pre každý zadávaný údaj, aby bola zabezpečená konzistencia všetkých zdrojov, ktoré sú v organizácií využívané.	1	3
Dátové štandardy	Vytvorenie lokálnych (rezortných dátových modelov) a zaviesť osobnú zodpovednosť pre túto oblasť centrálnu aj na každom rezorte, plus jednotné pravidlá pre tvorbu a udržiavanie dátových modelov a dátovej architektúry.	1	9
Dátové štandardy	Zabezpečenie referencovania získavaných údajov organizáciou alebo údajov vedených v databázach organizácií na existujúce referenčné registre.	1	12
Dátové štandardy	Zadefinovanie jasných biznis pravidiel pre riadenie dátovej kvality vrátane väzby biznis pravidiel na legislatívu upravujúcu dotknuté údaje.	1	6
Dátové štandardy	Implementácia biznis pravidiel pre meranie dátovej kvality (vrátane nastavenia jednoznačných KPIs a ich prahových hodnôt ako aj stanovenie, pre ktoré atribúty / záznamy je potrebné merať dátovú kvalitu) v organizáciách a nastavenie ich riadenia.	1	12
Organizácia	Posilnenie kapacít pre organizačné zabezpečenie riadenia dátovej kvality vo forme dátových kurátorov.	1	6
Proces	Zabezpečenie merania dátovovej kvality.	1	30
Proces	Jasné definovanie povinných a nepovinných údajov pri jednotlivých záznamoch tak, aby bolo pri hodnotení údajov vidieť aká je štatistika pre povinné a nepovinné údaje.	1	6
Proces	Nastavenie jednoduchých procedúr pre transformovanie údajov do foriem využiteľných pre realizáciu procesov podporujúcich dátovú kvalitu alebo procesov tvorby reportov a analýz.	1	9
Proces	Zabezpečenie jednoduchšieho prístupu k datasetom potrebným pre meranie a riadenie dátovej kvality (napr. Prístupy do všetkých informačných systémom by mali byť na úrovni READ kedykoľvek. Export dát z databázy by mal byť pre zodpovedné osoby ľahšie možný a najlepšie vo formáte SQL).	1	9
Technológie	Implementácia a zabezpečenie nástroja na riadenie dátovej kvality (v súčasnosti existuje platforma Talend, ktorá je zabezpečená na úrovni PaaS služby v štátnom cloude).	1	9
Technológie	Nastavenie technologických pravidiel pre prácu s databázami (jasne definované pravidlá správy údajov, ktoré sa dajú podporiť existujúcimi technológiami).	1	9
Technológie	Nastavenie pravidiel, aby záznamy nemohli byť skompletované, ak neobsahujú všetky potrebné náležitosti (povinné polia) a ak nebudú dodržané definované pravidlá zápisu údajov do príslušných polí.	1	9
Dátové štandardy	Zabezpečenie súladu súčasného stavu s existujúcimi štandardami dátovej kvality.	2	15
Dátové štandardy	Vypracovanie a udržiavanie dátovej dokumentácie každého rezortu a zabezpečenie jej správy - Treba definovať a zaviesť jednotné dátové štandardy, jednotné formy dátových modelov, jednotné dokumentácie k metadátam, jednotné štruktúry a formy biznis pravidiel.	2	18
Organizácia	Striktné zadefinovanie a prípadne preformulovanie pravidiel zadávania vstupných údajov do rezortných systémov tak, aby bol v čo najväčšej možnej miere eliminovaný ľudský faktor. Znamená to technické a kapacitné posilnenie na úrovni vstupov.	2	15
Proces	Prehodnotenie a nastavenie pravidiel opráv chybných záznamov tak, aby bol proces čo najkratší s jasne nastavenými KPIs vedúcimi k motivácii na jeho výkon (napr. v prípade súdov - prideliť túto zodpovednosť na jeden konkrétny súd, ktorý bude riadne vyškolený aj motivovaný na opravu takýchto chybných alebo neúplných záznamov).	2	15
Proces	Zavedenie jasných pravidiel pre aktualizáciu záznamov a ich atribútov, ako aj implementácia procesov monitoringu a dodržiavania pravidiel.	2	18
Proces	Nastavenie procesov prípravy údajov na analytické spracovanie (forma, štruktúra potrebná pre analýzy systémoch ako napr. R, SAS, Python, ...)	2	18
Technológie	Nastavenie procesov profilovania a analyzovania údajov priamo v technologickom riešení.	2	18
Technológie	Nastavenie proaktívnych kontrol pre zdrojové systémy – kontrola bude prebiehať už pri vstupe údajov na základe definovaných technologických a procesných požiadaviek.	2	18
Technológie	Zabezpečenie technologickej podpory pre definované biznis pravidlá pri vstupe údajov do systému.	2	18
Technológie	Implementácia nástroja / algoritmov na definovanie duplicity záznamov pri vstupe údajov do systému alebo na podporu procesu deduplikácie záznamov.	2	24
Organizácia	Implementovanie zmeny v oblasti dátovej kultúry a interpretácie údajov, ktoré má organizácia vo „vlastníctve“ vrátane zavedenia biznis metadát.	3	27
Proces	Zabezpečenie lepšej osvetly pre zadávanie nepovinných údajov a vysvetlenie pridanej hodnoty zadania nepovinného údajov do systému.	3	27
Technológie	Aplikovanie systémov AI pre technologické riadenie kvality údajov.	3	27
Technológie	Identifikovanie možností a následného zabezpečenia technologickej podpory pre elimináciu faktora „človek“ vo všetkých procesoch a fázach práce s údajmi.	3	27

### 3 Formulácia odporúčaní pre zlepšenie dátovej kvality

Táto časť popisuje odporúčania na zlepšenie dátovej kvality hodnotených registrov RPO, RA a RNZ, navrhnuté na základe zistení uvedených v predchádzajúcej kapitole a všeobecné odporúčania pre zvýšenie dátovej kvality vo verejnej správe, ktoré boli podrobne popisované vo výstupe 1.1.1 Štandardizácia dátovej kvality. Navrhovaný plán ich realizácie je popísaný v kapitole 7 *Naplánovanie zlepšenia dátovej kvality* a v kapitole 8 *Dohľad nad realizáciou odporúčaní pre zlepšenie dátovej kvality* nižšie.

#### 3.1 Centrálne dostupné služby riadenia kvality dát

Pri meraní dátovej kvality na RA a RNZ bola využitá centrálna dostupná služba na riadenie dátovej kvality DQ1 – Základná analýza dátovej kvality. Táto služba dátovej kvality na základe zadaných biznis pravidiel merala dátovú kvalitu na vybraných atribútoch a následne poskytla reporty dátovej kvality, ktoré boli spracované a vyhodnotené v tomto výstupe. Aktuálne sú pre riadenie dátovej kvality dostupné aj ďalšie služby, ktoré vznikli ako súčasť projektov Dátová integrácia (DI) a Centrálna integračná platforma (CIP). Služby riadenia dátovej kvality sú určené na kontrolu a niektoré z nich aj na následné zlepšovanie dátovej kvality.

V súčasnosti sú dostupné alebo sa pripravujú tieto služby dátovej kvality:

##### **IS CSRÚ portál a IS CSRÚ portál Talend data stewardship konzola (projekt DI)**

- DQ1 – Základná analýza dátovej kvality
- DQ2 – Validácia voči objektu evidencie (OE)
- DQ3 – Asistované čistenie
- DQ4 – Automatické stotožnenie
- DQ5 – Asistované stotožnenie
- DQ6-9 – Transformácia dátového súboru
- DQ10 – Vlastná služba dátovej kvality podľa špecifických potrieb konkrétneho OVM

##### **PaaS (CIP):**

- DQ3 – Asistované čistenie
- DQ4 – Automatické stotožnenie
- DQ9 – Transformovanie datasetu
  - Využitie časti obohatenia dátového súboru
  - Využitie časti stotožnenia dátového súboru
  - Využitie na transformáciu do RDF / JSON-LD
- Nástroj na profilovanie dátových súborov OpenMetadata

Podrobnejšie popisy týchto služieb dátovej kvality (ich charakteristiky, aplikácie, funkcie, použitie a výstupy) sú dostupné vo výstupe 1.2.2 Služby dátovej kvality.



## 3.2 Väzby na analýzu súčasného stavu

Vzhľadom na zistenia z hodnotenia súčasného stavu v oblasti riadenia dátovej kvality (prehľad zistení z výstupu č. 1.1.1 Štandardizácia dátovej kvality) je potrebné uviesť aj s tým súvisiace odporúčania, ktoré majú priamy vplyv na zlepšenie dátovej kvality vo verejnej správe. Tieto odporúčania majú komplementárny charakter vo vzťahu k odporúčaniam vzťahujúcim sa k jednotlivým registrom (RPO, RA, RNZ) uvedeným v ďalších podkapitolách, pričom je potrebné ich vykladať vo vzájomnej súvislosti.

### Tabuľka 78 Všeobecné odporúčania pre zlepšenie dátovej kvality

#### Odporúčanie / Popis

#### 1 Zvýšiť počet OVM integrovaných na CMÚ

V záujme interoperability by mali byť všetky modely údajov a pravidiel definované v CMÚ a OVM by ich mali následne preberať do svojich registrov. V súčasnosti je na CMÚ napojený len RA a (postupne) CES, ale **postupne** by mali pribúdať ďalšie registre. Prioritne by nimi mali byť referenčné registre (RPO a jeho zdrojové registre, RFO). Je možné zväziť prevzatie realizácie tohto opatrenia pre referenčné registre pod zvýšený dohľad, resp. priamo pod gesciu centrálnej úrovne, správcu NSZÚ (Dátová kancelária), tzn. zamerať centrálnu úsilie na pokrok v konkrétnych kľúčových prípadoch a po dosiahnutí pokroku vytrvalo pokračovať ďalšími prioritami tak, aby sa zvýšil počet OVM, ktoré stotožňujú svoje údaje s referenčnými registrami a následne budú využívať CMÚ v plnej šírke. Pozri aj súvisiace opatrenie č. 7 nižšie.

#### 2 Zamerať sa pri navrhovaní nových projektov na osvedčenú prax v ostatných štátoch EÚ

Slovensko by sa malo inšpirovať politikami, aktivitami a riešeniami riadenia DK v iných členských štátoch EÚ, ktoré sú v oblasti dátovej kvality pred nami a vždy posúdiť, či už bola téma riešená inde a akým spôsobom. Pri navrhovaní nových projektov by mal každý žiadateľ do štúdie uskutočniteľnosti doplniť aj povinné porovnanie, ako je problém vyriešený a ako bolo riešenie realizované v iných členských štátoch EÚ.

#### 3 Zaviesť vynútiteľnú povinnosť pravidelne merať DK a jednotným spôsobom zverejňovať výsledky meraní

Metodika merania DK definuje, akým spôsobom by mali OVM postupovať pri zlepšovaní dátovej kvality svojich registrov. Súčasťou sú aj pravidelné merania kvality a vyhodnocovanie ich výsledkov. Nie je možné posúdiť, či a ako sú tieto merania vykonávané, pretože žiadne z OVM v súčasnosti nezverejňuje správy o meraniach kvality<sup>18</sup>.

#### 4 Zverejňovať pravidelné reporty dátovej kvality jednotným spôsobom

V nadväznosti na predchádzajúce odporúčanie by mali OVM pravidelne zverejňovať reporty o dátovej kvalite svojich registrov. Reporty by mali byť jednotné, aby ich bolo možné vyhodnocovať a jednoznačne určiť úroveň dátovej kvality registra.

Detailnejšie odporúčania k reportovaniu sú predmetom kapitoly 6 výstupu č. 1.1.1 Štandardizácia dátovej kvality.

<sup>18</sup> V minulosti boli zverejňované výsledky merania ŠÚ SR za RPO, podrobnejšie informácie sú uvedené v kapitole 2.1.3.



## Odporúčanie / Popis

### 5 Jasne dodefinovať roly a zodpovednosti v rámci riadenia DK na úrovni MIRRI a na úrovni OVM

Rola dátového kurátora je zatiaľ určená len uznesením vlády Slovenskej republiky č. 654 zo 14. októbra 2020<sup>19</sup> a Metodickým usmernením MIRRI k aplikácii základných princípov pri realizácii projektov IT financovaných z verejných zdrojov a zdrojov EÚ č. 00NNN/2021/SITVS zo dňa 31. marca 2021<sup>20</sup>. Žiadny z týchto dokumentov nedefinuje jeho kompetencie a povinnosti. Tie budú stanovené až v pripravovanom Zákone o údajoch. Riadenie dátovej kvality však vyžaduje aj ďalšie roly a ich zodpovednosti, ktoré by mali byť definované tak na úrovni MIRRI, ako aj na úrovni OVM, ktoré však závisia od organizačného nastavenia v konkrétnej situácii, preto nie je možné poskytnúť generalizované odporúčania.

### 6 Zjednotiť pravidlá riadenia dátovej kvality naprieč OVM

Pravidlá riadenia dátovej kvality sú síce definované, napríklad v Metodike, ale OVM ich nedodržiavajú, čo je zjavné už z toho, že nevykonávajú pravidelné merania a nezvereňujú ich výsledky. Zabezpečenie konzistentnosti riadenia dátovej kvality na Slovensku vyžaduje, aby OVM prevzali a implementovali centrálné stanovené pravidlá a na ich základe vykonávali jednotné aktivity vedúce k jednotným cieľom. Dátová kancelária ale nepovažuje za efektívne vynucovať dodržiavania usmernení prostredníctvom sankcií. Možnosťou riešenia tejto situácie by bolo zavedenie povinnosti (napríklad prostredníctvom nového Zákona o údajoch), aby na základe centrálnych stanovených metodík každé OVM vytvorilo vlastnú metodiku, ktorá bude definovať roly, procesy a aktivity v oblasti riadenia dátovej kvality prispôbené na špecifickú situáciu daného OVM tak, aby metodika nebola len formálnym dokumentom. Dátová kancelária posúdi dostatočnosť metodiky a jej súlad s centrálnymi usmerneniami. OVM následne na základe schválenej metodiky vytvorí interný riadiaci akt (alebo ekvivalentný dokument), ktorý zabezpečí jej dodržiavanie. Tým bude podporená nielen jednotnosť politík naprieč OVM, ale aj ich vhodnosť pre každé OVM a ich dodržiavanie.

### 7 Zabezpečiť stotožňovanie údajov pri vstupoch nových údajov do registrov

Príkladom toho, ako môže nestotožňovanie údajov na vstupe ovplyvniť viacero procesov riadenia dát je zápis nových subjektov do obchodného registra. Pri zápise českej spoločnosti do Obchodného registra prostredníctvom online formulára je možné, aby bola spoločnosť zapísaná bez toho, že bolo skontrolované zadávané IČO. Kontrolu musí následne vykonať príslušný krajský súd, ktorý však niekedy zápis povolí, aj keď je protiprávny. Tým dochádza k vytvoreniu záznamov, ktoré nielenže sú nekvalitné, ale aj zavádzajúce. Pokiaľ by bolo stotožnenie vykonané priamo pri vstupe, bol by záznam overený a nedošlo by k vytváraniu záznamov s neprávnyim IČO. Vo všeobecnosti, pokiaľ OVM stotožňuje údaje po zápise záznamu do registra, znižuje sa tým prácnosť pri následnej kontrole a opravách. Stotožnenie údajov pri zápise zvýši kvalitu údajov v registri a efektívnosť riadenia dátovej kvality. Preto by mala Dátová kancelária naďalej iniciovať a podporiť projekty na implementáciu stotožňovania pri vstupe dát do registrov a v prípade prioritných registrov prevziať zvýšený dohľad (pozri odporúčanie 1 vyššie).

### 8 Riadenie zmien zaradiť medzi prioritné úlohy a zodpovednosti Dátovej kancelárie

Riadenie zmien, často nazývané aj change management alebo riadenie strategickej zmeny (strategic change management), je komplexný proces, ktorý vyžaduje dôkladnú plánovaciu a komunikačnú stratégiu. Ak sa však správne uchopí a dôsledne realizuje, môže významne prispieť k zvýšeniu dátovej kvality vo verejnej správe. Tu je niekoľko kľúčových úloh riadenia zmien pri zvyšovaní dátovej kvality vo verejnej správe:

<sup>19</sup> [Detail uznesenia | Portal OV \(gov.sk\)](#)

<sup>20</sup> [Metodicke-usmernenie-009417-2021-oSBAA-1-v3.pdf \(gov.sk\)](#)

## Odporúčanie / Popis

*Identifikácia potrebných zmien* v súvislosti s dátovou kvalitou. To zahŕňa identifikáciu slabých miest v súčasných procesoch správy údajov a vyhodnotenie potenciálnych prínosov vylepšených postupov a určenie ich priority.

*Plánovanie zmien:* vytvorenie akčného plánu na podporu merania, reportovania stavu a zlepšovania DK a na zvýšenie súladu OVM s publikovanými metodikami. Plán má obsahovať jasné ciele, harmonogram a zodpovednosti. Plánovanie zmien zahŕňa aj identifikáciu, vyhodnotenie a riadenie rizík a prípravu opatrení na prekonanie možných prekážok, ktoré môžu vzniknúť pri implementácii zmien počas realizácie plánu.

*Komunikácia a zapojenie zainteresovaných strán:* Je nevyhnutné komunikovať so všetkými zainteresovanými stranami vrátane vedúcich pracovníkov, pracovníkov na úrovni správy dát a používateľov údajov. Musia byť informovaní o dôvodoch zmien, prínosoch a očakávaných vplyvoch na ich prácu. Komunikácia a zapojenie zainteresovaných strán (stakeholder management) sa týka tak centrálnej úrovne (vedenie MIRRI, Dátová kancelária a spolupracujúce útvary), ako aj úrovne OVM (opäť v troch úrovniach – vedenie rezortu, vecné útvary a technologické útvary IT/IS).

*Školenie a podpora:* akčný plán a jeho implementácia môže (a zjavne bude) vyžadovať nové nástroje, technológie alebo zručnosti pre zamestnancov. Je dôležité poskytnúť primeraný výcvik a podporu zainteresovaným osobám, aby boli schopné úspešne pracovať s novými procesmi alebo nástrojmi.

*Monitorovanie a hodnotenie:* Riadenie zmien nekončí po implementácii. Je dôležité monitorovať a hodnotiť účinnosť zmien a dátovej kvality v dlhodobom horizonte. To umožňuje identifikovať prípadné nedostatky alebo oblasti, ktoré vyžadujú ďalšie zlepšenie. Monitorovanie a hodnotenie pokroku sa teda neobmedzuje na obdobie realizácie akčného plánu ale zahŕňa aj obdobie po zavedení zmeny do rutínnej praxe.

*Udržiavanie zmien:* Udržanie výsledkov zlepšenia dátovej kvality si vyžaduje trvalé úsilie. Je dôležité mať stabilné procesy a systémy, ktoré podporujú vysokú úroveň dátovej kvality a poskytujú mechanizmy na odhaľovanie a riešenie problémov v reálnom čase.

### 3.3 Odporúčania pre RPO

Štatistický úrad ako OVM zodpovedný za referenčný register (RPO) by mal ďalej pokračovať v pravidelnom meraní kvality dát a čo najskôr uskutočniť ďalšie meranie vrátane zverejnenia výsledkov z merania a oslovenia príslušných OVM poskytujúcich údaje, u ktorých boli identifikované nedostatky v kvalite poskytovaných údajov.

Za týmto účelom by malo MIRRI dôrazne vyžadovať vykonávanie pravidelných meraní dátovej kvality minimálne u všetkých OVM, ktoré majú do svojej zodpovednosti zverený niektorý z referenčných registrov alebo zdrojových registrov.

Meranie dátovej kvality RPO by malo byť vykonávané na základe biznis pravidiel pre jednotlivé parametre dátovej kvality (presnosť, konzistentnosť, správnosť, kompletnosť, unikátnosť, aktuálnosť, strojová spracovateľnosť, interoperabilita a referenčná integrita), aby bolo možné vyhodnocovať dátovú kvalitu RPO ako aj ostatných registrov jednotne. Preto je potrebné, aby pre každý parameter bol navrhnutý dostatočný počet biznis pravidiel, pomocou ktorých bude vykonané meranie. Dostatočné množstvo biznis pravidiel zabezpečí reprezentatívnosť výsledku dátovej kvality pre daný parameter. Dostatočné množstvo biznis pravidiel bude vyplývať z veľkosti registra (množstva atribútov) a existujúcich biznis potrieb na merania dátovej kvality pre zabezpečenie dostatočnej informovanosti o dátovej kvalite registra.

Predpokladom riadenia dátovej kvality v organizácii je vymedzenie kapacít dedikovaných priamo za účelom realizovania s tým spojených aktivít v súlade s nastavenými procesmi. Štatistický úrad by preto mal prostredníctvom svojich interných dokumentov presne stanoviť kto a akým spôsobom riadi dátovú kvalitu RPO. Tieto aktivity musia byť formálne priradené konkrétnemu vlastníkovi (konkrétnej role v organizácii) vrátane štandardizácie postupu pri riadení kvality – cieľom je zabezpečiť, aby agenda bola vykonávaná bez ohľadu na personálne zmeny v organizácii, aby aktivity boli realizované nepretržite a zároveň aby aj proces merania kvality dát RPO bol vykonávaný v súlade s metodikou merania dátovej kvality vo verejnej správe.

Na základe týchto skutočností boli identifikované nasledujúce odporúčania:

- Štatistický úrad by mal v spolupráci s Dátovou kanceláriou (MIRRI) definovať spôsob a postup merania dátovej kvality tak, aby výstup z merania dátovej kvality bol dostatočne informatívny pre konzumentov dát RPO a zadané biznis pravidlá boli zatriedené do parametrov dátovej kvality.
- Je potrebné zabezpečiť kontinuitu vykonávania meraní dátovej kvality a riadenia dátovej kvality RPO bez ohľadu na personálne zmeny na Štatistickom úrade.
- Je potrebné formulovať kampane na zvyšovanie dátovej kvality v zdrojových registroch RPO a realizovať ich v spolupráci so zdrojovými registrami.
- ŠÚ SR by mal v spolupráci s Dátovou kanceláriou vytvoriť metodiku pre zdrojové registre RPO na poskytovanie údajov vo vysokej kvalite.
- ŠÚ SR by mal v spolupráci s Dátovou kanceláriou preveriť možnosti zavedenia prahových limitov na parametre dátovej kvality v RPO a ich zadefinovanie prostredníctvom uznesenia, vyhlášky príp. iného relevantného záväzného aktu.
- ŠÚ SR by mal pravidelne vykonávať merania dátovej kvality RPO a zverejnením výsledkov meraní sprístupniť informácie o dátovej kvalite RPO konzumentom jeho údajov.
- Dátová kancelária (MIRRI) by mala aktívnejšie spolupracovať so ŠÚ SR na zlepšovaní dátovej kvality RPO (ako referenčného registra) najmä vo vzťahu k zdrojovým registrom. V spojení síl vidíme priestor na vytvorenie sústredenejšieho tlaku na zodpovedné OVM. Toto odporúčanie je podobne aplikovateľné aj pre ďalšie referenčné registre (alebo iné prioritné OE).

### 3.4 Odporúčania pre RA

Hoci výsledky merania dátovej kvality RA boli skreslené meraním nad všetkými údajmi (teda aj neplatnými údajmi), posúdenie procesov dátovej kvality RA preukazuje pozitívny dopad dodržiavania Metodiky DK. Vďaka prísnyh kontrolám biznis pravidiel správnosti, presnosti, kompletnosti a unikátnosti je maximalizovaná kvalita pribúdajúcich záznamov. Opakované kampane na čistenie údajov zas postupne zvyšujú kvalitu historických záznamov. Keďže meraním boli zistená pretrvávajúca minoritná nekvalita údajov, najmä z hľadiska kompletnosti, MV SR by malo naďalej pokračovať v aktivitách čistenia údajov, kým RA nedosiahne maximálnu požadovanú úroveň kvality údajov z hľadiska všetkých parametrov dátovej kvality.

Dátová kvalita RA, tak ako každého referenčného registra, by mala byť verejne preukázateľná dostatočne informatívnym spôsobom pre jeho konzumentov. Kvalita RA by mala byť pravidelne meraná nielen pre interné účely čistenia dát v registri, k ich výstupom by mali mať prístup aj konzumenti RA. Na základe toho boli formulované nasledujúce odporúčania:

- MV SR by malo vykonávať pravidelné merania dátovej kvality a v súlade s usmerneniami Dátovej kancelárie zverejňovať výsledky, aby boli informácie o kvalite údajov v RA dostupné pre jeho konzumentov.
- Biznis pravidlá pre meranie a štruktúra zverejňovaných reportov by mali byť založené na požiadavkách konzumentov a schválené Dátovou kanceláriou.
- Zvyšovanie DK by malo byť riadeným procesom založeným na definovaných cieľoch (prahových hodnotách parametrov DK) a plán na dosiahnutie stanovených cieľov by mal zahŕňať úlohy nevyhnutné pre dosiahnutie cieľov.

### 3.5 Odporúčania pre RNZ

Aj napriek tomu, že NIP nemá implementovanú žiadnu metodiku riadenia DK a nevykonáva žiadne merania DK, výsledky merania preukázali vysokú kvalitu údajov z hľadiska väčšiny parametrov DK. Väčšina nekvalitných údajov bola do registra prenesená z neaktuálnej lokálnej kópie Štatistického registra organizácií. Zároveň bolo meraním odhalené malé množstvo duplícít. Tie boli podľa správcu registra spôsobené obmedzením informačného systému, ktorý v špecifických prípadoch vytvára dva záznamy pre jednu evidovanú pokutu (obmedzenie 99 nelegálne zamestnaných osôb na jeden záznam). Časť údajov RNZ (pokuty udelené slovenským právny subjektom) je publikovaná ako tabuľka vo formáte HTML, ale údaje o pokutách udelené zahraničným subjektom sú evidované samostatne, sú neúplné a publikované vo formáte PDF. V rámci stretnutí so správcom registra bolo tiež zaznamenané zistenie, že nové záznamy sú vytvárané a aktualizované v 9 lokálnych verziách databázy a do centrálnej databázy sú prenášané v týždenných intervaloch. Keďže RNZ nebol referenčným registrom, boli nároky na jeho dátovú kvalitu definované len potrebami rezortu MPSVaR SR. V priebehu vypracovania tohto dokumentu však prebiehali prípravy na implementáciu nového informačného systému registra (IS SAWO) a zmenu RNZ na referenčný register. Z toho dôvodu by sa mali primerane zvýšiť aj požiadavky na riadenie jeho dátovej kvality. V svetle týchto skutočností boli formulované nasledujúce odporúčania.

- RNZ by mal byť priamo integrovaný na referenčné dáta RPO, aby mohli byť pri vytváraní záznamu údaje o pokutovanom subjekte stotožňované voči aktuálnym údajom RPO.
- Všetky nové záznamy by mali byť vytvárané v centrálnom systéme, nie v lokálnych kópiách databázy.
- Pri implementácii nového IS je potrebné odstrániť obmedzenia, ktoré prispievajú k vytváraniu duplicitných záznamov.
- Do registra je potrebné zahrnúť aj zoznam pokutovaných fyzických osôb a právnických so sídlom mimo územia Slovenskej republiky.
- Dosiahnutie referenčnej integrity vyžaduje zavedenie jednoznačného referenčného identifikátora.
- Zvýšenie strojovej spracovateľnosti vyžaduje sprístupnenie údajov v otvorenom, strojovo spracovateľnom formáte.

## 4 Zovšeobecnenie odporúčaní - Príručka kvality pre OVM

Na základe zistení v analytickej časti tohto dokumentu (ako i výstupu 1.1.1 Štandardizácia dátovej kvality) predkladáme v tejto kapitole návrh jednotného postupu pri riadení dátovej kvality vo verejnej správe. Sú do nej zahrnuté opatrenia na realizáciu viacerých odporúčaní, uvedených v tomto dokumente, čím dopĺňa predchádzajúcu kapitolu 3. Treba ju samozrejme vnímať nielen ako integrálnu súčasť celého tohto dokumentu ale najmä ako návrh previazaný s kapitolami 7 a 8.

Základom pre úspešné riadenie dátovej kvality na úrovni OVM je pochopenie samotného pojmu dátovej kvality a jej parametrov, ako aj zabezpečenie prioritizácie tejto oblasti v rámci agendy jednotlivých OVM. Preto ešte pred vlastnou realizáciou aktivít spočívajúcich v riadení dátovej kvality, by mal byť na úrovni každého OVM (vrátane MIRRI) deklarovaný záväzok k dodržiavaniu/riadeniu a najmä zlepšovaniu dátovej kvality, ktorý je v súlade s definíciou dátovej kvality. Tento záväzok by mal byť podporovaný na všetkých úrovniach riadenia OVM ako aj v rámci spolupráce medzi OVM a MIRRI.

Inšpiráciou môže byť napríklad potreba zavedenia spoločného rámca dátovej kvality vo verejnej správe viditeľná v podmienkach Spojeného kráľovstva zavedením tzv. Rámca pre riadenie dátovej kvality vo verejnej správe<sup>21</sup>, ktorý okrem iného ako jeden zo svojich princípov uvádza práve nevyhnutnosť zaviazat' sa k dátovej kvalite.

Ďalším príkladom môže byť snaha o zavedenie kvality štatistík v rámci Európskeho štatistického systému (EŠS) prostredníctvom Spoločného rámca kvality EŠS<sup>22</sup> (vzťahuje sa aj na Národný štatistický systém v Slovenskej republike), ktorý obsahuje:

- Vyhlásenie EŠS o kvalite (demonštruje povedomie o kvalite, ktoré v EŠS existuje, ako aj záväzok všetkých jeho členov neustále rozvíjať, vytvárať a šíriť kvalitnú európsku štatistiku),
- Kódex postupov pre európsku štatistiku (zabezpečuje dôveru verejnosti v európsku štatistiku stanovením, ako sa má európska štatistika rozvíjať, tvoriť a šíriť v súlade so štatistickými zásadami a s najlepšimi medzinárodnými štatistickými postupmi),
- Rámec pre zabezpečenie kvality EŠS (predstavuje príručku pre implementáciu Kódexu postupov; obsahuje odporúčané metódy, nástroje a dobrú prax).

Z pohľadu riadenia dátovej kvality vo verejnej správe je možné vnímať podobnosti s riadením kvality štatistík v rámci Národného štatistického systému, kde dochádza k štandardizácii postupov a procesov členov Národného štatistického systému (paralela k jednotlivým OVM v rámci riadenia dátovej kvality) za koordinácie Štatistického úradu (paralela k MIRRI v rámci riadenia dátovej kvality) s cieľom zachovania resp. zvyšovania kvality štatistík. Preto inšpiráciou pre zavedenie riadenia dátovej kvality vo verejnej správe môže byť práve Spoločný rámec kvality EŠS.

Navrhovaná príručka riadenia dátovej kvality pre OVM, ktorá je predmetom tejto kapitoly preto obsahuje:

- vymedzenie dátovej kvality a jej parametrov (pojem dátová kvalita musí byť chápaný a vnímaný jednotne všetkými OVM),
- záväzok k udržiavaniu/riadeniu a zlepšovaniu dátovej kvality,

<sup>21</sup> [The Government Data Quality Framework - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

<sup>22</sup> [Kvalita štatistík \(statistics.sk\)](http://statistics.sk)

- rámec a popis procesu riadenia dátovej kvality na úrovni OVM,
- rámec koordinačnej činnosti a dohľadu vykonávaného MIRRI z pozície centrálného koordinátora riadenia dátovej kvality.

## 4.1 Úvod, kontext a východiská

Potenciál a význam údajov je v súčasnej spoločnosti nespochybniteľný, či už vo verejnom alebo súkromnom sektore. Jednou z priorít Európskej komisie na roky 2019 až 2024 je Európa pripravená na digitálny vek<sup>23</sup> ako súčasť európskej dátovej stratégie Európskej únie<sup>24</sup>. EÚ má za cieľ stať sa nasledovaniahodným príkladom spoločnosti, ktorá za svoju silu vďačí údajom a zároveň cieľom európskej dátovej stratégie je dosiahnuť, aby sa EÚ stala lídrom v spoločnosti založenej na údajoch.

Prístup k údajom a schopnosť ich využívania sú nevyhnutné pre inovácie a rast, pričom inovácie založené na údajoch môžu mať významné a konkrétne prínosy, ako sú napríklad personalizovaná medicína, lepšia mobilita, lepšia tvorba politiky a modernizácia verejných služieb.

Tento trend spočívajúci v digitálnej transformácii spoločnosti ovplyvňuje aj život v podmienkach Slovenskej republiky, kde v kontexte prebiehajúcej transformácie industriálnej spoločnosti na informačnú spoločnosť bola prijatá Stratégia digitálnej transformácie Slovenska 2030<sup>25</sup>.

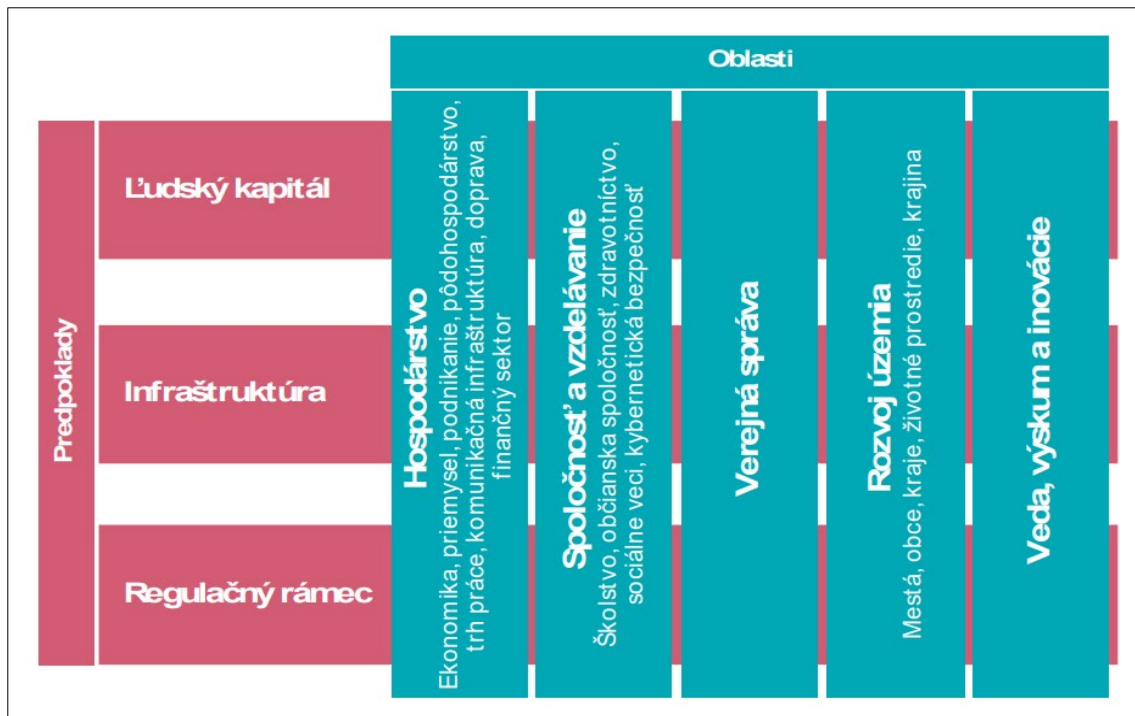
Príručka zároveň predstavuje rámec, na základe ktorého by mali jednotlivé OVM postupovať pri vytváraní riadiacich aktov vo svojej organizácii zameraných na organizačné, procesné a technické zabezpečenie riadenia dátovej kvality.

<sup>23</sup> [Európa pripravená na digitálny vek \(europa.eu\)](https://europa.eu/europa/en/digital-competence)

<sup>24</sup> [Európska dátová stratégia \(europa.eu\)](https://europa.eu/europa/en/digital-competence)

<sup>25</sup> [Strategia-digitalnej-transformacie-Slovenska-2030.pdf \(gov.sk\)](https://www.gov.sk/Assets/Strategia-digitalnej-transformacie-Slovenska-2030.pdf)





**Obrázok 1 Vízia digitálnej transformácie: predpoklady a oblasti<sup>26</sup>**

Ako jedna z oblastí, v ktorej je nevyhnutne potrebné znásobiť svoj potenciál prostredníctvom digitálnej transformácie, bola identifikovaná aj verejná správa. Zlepšenie schopností verejnej správy využívať údaje a inovácie v prospech občanov je jednou zo základných prioritných oblastí Akčného plánu digitálnej transformácie Slovenska<sup>27</sup> (krátkodobý horizont) a zároveň sa prostredníctvom digitálnej transformácie má smerovať k modernej a efektívnej verejnej správe (dlhodobý horizont).

Pre naplnenie potenciálu digitálnej transformácie je potrebná aktivita a súčinnosť jednotlivých OVM ako aj koordinácia príslušných aktivít na úrovni MIRRI. Rozhodovanie na základe dát, poskytovanie digitálnych služieb (možnosť vybaviť všetko online bez potreby osobnej návštevy úradu), aplikovanie princípu jedenkrát a dosť (eliminovanie potreby oznamovať tú istú skutočnosť viacerým úradom) a pod. Toto všetko sú oblasti, v ktorých v podmienkach SR prebiehajú snahy o implementáciu efektívnych riešení v súlade s víziou digitálnej transformácie a ktoré si vyžadujú riadenie dátovej kvality naprieč celou verejnou správou.

Dátová kvalita je ústredným pojmom, ide o nevyhnutný prvok digitálnej transformácie kľúčový pre maximalizáciu potenciálu údajov zbieraných, spracovávaných a poskytovaných verejnou správou. Chápanie dátovej kvality musí byť jednotné naprieč všetkými OVM, aby aj jej riadenie mohlo byť správne koordinované. Cieľom riadenia a zlepšovania dátovej kvality vo verejnej správe nie je dátová kvalita sama o sebe, ale vyššia kvalita života a lepšie kvalitnejšie služby verejnosti (tie sú silno prepojené a závislé práve na primeranej dátovej kvalite).

Cieľom tejto príručky je poskytnúť jednotné vymedzenie pojmu dátová kvalita všeobecne akceptované vo verejnej správe vrátane záväzku OVM zodpovedne

<sup>26</sup> Zdroj: Stratégia digitálnej transformácie Slovenska 2030

<sup>27</sup> <https://www.mirri.gov.sk/wp-content/uploads/2023/01/APDTS-2023-2026.pdf>

a proaktívne pristupovať k riadeniu dátovej kvality. Jednotlivé aspekty riadenia dátovej kvality sú predmetom úpravy viacerých dokumentov, pričom táto príručka sa nesnaží duplikovať už existujúce pokyny/usmernenia/dokumenty – účelom príručky je poskytnúť OVM podklad pre pochopenie významu a parametrov dátovej kvality, ako aj rámec pre implementáciu riadenia kvality údajov vo svojej organizácii vrátane nasmerovania na najrelevantnejšie zdroje informácií z tejto oblasti.

## 4.2 Údaje vo verejnej správe a ich kvalita

Údaje nás sprevádzajú dennodenne, pri rôznych životných situáciách a so zapojením rôznych OVM. Od narodenia (evidencia v matrike, registri fyzických osôb), cez bežné udalosti, ktoré sú súčasťou našich aktivít v podobe štúdia (evidencia v centrálnom registri študentov), sobáša (zmena v evidencii matriky, registri fyzických osôb), kúpy nehnuteľnosti (záznam v katastri nehnuteľností) a pod., až po smrť. Každý deň orgány verejnej moci pri svojej činnosti zbierajú, spracovávajú a poskytujú údaje o rôznych entitách (fyzických osobách, právnických osobách, iných subjektoch).

Zjednodušene možno povedať, že OVM získavajú údaje (buď priamo od občanov alebo od iných OVM, ktoré tieto údaje o občanoch majú), evidujú ich v príslušných evidenciách pomocou ISVS a používajú ich pri svojej činnosti (napr. pri vydávaní rozhodnutí) alebo ich zdieľajú s iným OVM. Zároveň dôležitou skutočnosťou plánovanou do budúcnosti je vytvorenie takého systému zdieľania údajov, v ktorom tieto údaje bude možné zdieľať aj s inými entitami, ako len orgánmi verejnej moci – napríklad zdieľanie údajov z ISVS s bankou pri schvaľovaní hypotéky.

Jeden z najdôležitejších predpokladov fungovania a spoľahlivosti takéhoto systému na zdieľanie údajov je dátová kvalita. Údaje o konkrétnom subjekte (napr. občanovi) evidované v IS jedného OVM by nemali byť v rozpore s rovnakými údajmi o danom subjekte evidovanými iným OVM. OVM, ktorý v konkrétnej životnej situácii zbiera údaje od konkrétneho subjektu by mal byť schopný porovnať tieto údaje s už existujúcimi údajmi, o ktorých vie, že sú „jedinou pravdou“. Malo by byť umožnené využitie už získaných údajov a OVM by nemal od konkrétneho subjektu vyžadovať poskytnutie údajov, ktoré už má k dispozícii iný OVM zo svojej činnosti. Poskytnutie nových údajov od konkrétneho subjektu (napr. zmena v údajoch) by sa malo prejavovať vo všetkých ISVS používajúcich daný údaj. Mali by byť určené také registre, ktoré je možné považovať za zdroj tých správnych údajov, a ktoré je možné ďalej používať inými OVM. Všetky tieto predpoklady a aj mnohé ďalšie priamo súvisia s riadením dátovej kvality a majú významný vplyv na interoperabilitu údajov vo verejnej správe.

Na základe vyššie uvedeného možno konštatovať, že pokiaľ nebude zabezpečená požadovaná úroveň kvality údajov vo verejnej správe, nebude možné poskytovať plnohodnotné digitálne služby a nebude možné zároveň ani hovoriť o efektívnej a modernej verejnej správe.

### 4.2.1 Vymedzenie dátovej kvality vo verejnej správe

Dátová kvalita (kvalita údajov) je merateľná charakteristika údajov v informačných systémoch verejnej správy vyjadrená príslušnými parametrami – presnosť, konzistentnosť, správnosť, kompletnosť, unikátnosť, aktuálnosť, strojová spracovateľnosť a referenčná integrita.



**Presnosť (čistota):** miera, s akou objekt evidencie reprezentuje reálny svet, vyjadrená zhodou s referenčnými údajmi.

**Kompletnosť:** kompletnosť údajov znamená, že všetky údaje z objektu evidencie, považované alebo označené za povinné, sú prítomné v dátovom prvku.

**Aktuálnosť:** údaje sú časovo adekvátne a považované za aktuálne.

**Unikátnosť:** vyhodnotenie duplicity údajov vo vzťahu k jednoznačnému referencovateľnému identifikátoru.

**Referenčná integrita:** 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.

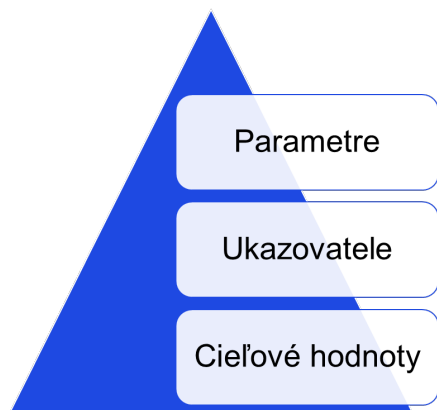
**Strojová spracovateľnosť:** možnosti automatického spracovania údajov plynúce z formátu reprezentácie dát ako napr. spájanie rôznych dát z rôznych zdrojov, či spracovanie dát s ohľadom na ich význam.

**Konzistentnosť:** vzájomné logické vzťahy v rámci objektu evidencie sú správne kompatibilné a v súlade so stanovenými biznis pravidlami.

**Správnosť:** zhoda údajov s kritériami, ktoré stanovujú formát dát.

Úroveň dátovej kvality, resp. jej jednotlivých parametrov sa vyjadruje prostredníctvom hodnôt dosahovaných v rámci konkrétnych ukazovateľov, pričom vo vzťahu k dátovej kvalite rozlišujeme kvalitatívne a kvantitatívne ukazovatele. Kvalitatívne ukazovatele vyjadrujú kvalitu dát z architektonického pohľadu – napríklad strojová spracovateľnosť údajov závisí od ich formátu, t. j. v tomto prípade je v centre záujmu architektúra a nie samotné záznamy údajov. Kvantitatívne ukazovatele predstavujú „tvrdý“ pohľad na dáta – napríklad počet nevyplnených hodnôt – v centre záujmu sú samotné dáta. Z pohľadu vyhodnocovania dátovej kvality je nutné uviesť, že vyhodnocovanie kvantitatívnych ukazovateľov je jednoduchšie, založené na použití konkrétneho vzorca (často automatizovane), zatiaľ čo vyhodnocovanie kvalitatívnych ukazovateľov môže byť zložitejšie – nie vždy existuje vzorec, podľa ktorého sa dá takýto ukazovateľ hodnotiť.

Pre jednotlivé ukazovatele je potrebné definovať cieľové hodnoty – cieľové hodnoty sú ambíciou, ktorú chce organizácia dosiahnuť v rámci daného parametra dátovej kvality. Napríklad pre parameter presnosť je možné určiť ukazovateľ syntaktická presnosť a stanoviť cieľovú hodnotu na 90%, t. j. aspoň 90% záznamov má byť syntakticky bezchybných. Následne na základe merania dátovej kvality a porovnania dosiahnutej hodnoty s cieľovou hodnotou možno posúdiť úroveň dátovej kvality v organizácii a definovať prípadné opatrenia pre zlepšenie dátovej kvality.



**Obrázok 2 Vymedzenie dátovej kvality vo verejnej správe**

Takto vymedzená dátová kvalita sa vzťahuje na všetky údaje, ktorými verejná správa, resp. jednotlivé OVM disponujú. Preto pre príslušných vlastníkov údajov (jednotlivé OVM) vyplýva aj zodpovednosť za kvalitu údajov. Napríklad údaje v referenčných registroch sú údaje, o ktorých platí domnienka správnosti. Vďaka tejto domnienke môžu byť údaje z referenčného registra použité pri činnosti rôznych OVM bez obáv o ich správnosť. Údaje z referenčných registrov sú však získavané z rôznych zdrojových registrov od rôznych OVM. Znamená to, že kvalita údajov referenčných registrov je významne ovplyvnená kvalitou údajov zdrojových registrov.

Stále intenzívnejšia previazanosť údajov vo vlastníctve jednotlivých OVM zvyšuje zároveň nároky na ich participáciu v procese riadenia dátovej kvality bez ohľadu na to, či konkrétne OVM je vlastníkom referenčného registra alebo zdrojového registra niektorého z referenčných registrov a pod. Vo všeobecnosti každý OVM je zodpovedný za zabezpečenie, udržiavanie a rozvíjanie náležitej úrovne dátovej kvality zdieľaných údajov vo vzťahu k ISVS, ktoré spravuje alebo prevádzkuje.

Vymedzenie dátovej kvality prostredníctvom jej parametrov je jednotné naprieč celou verejnou správou, avšak stanovenie konkrétnych ukazovateľov a ich cieľových hodnôt môže byť u jednotlivých OVM odlišné vzhľadom na rôzny charakter rôznych údajov.

#### **4.2.2 Závazok k dátovej kvalite**

Dátová kvalita je jednou z priorít orgánov verejnej moci. Orgány verejnej moci proaktívne pristupujú k riadeniu dátovej kvality prijatím jednotnej definície dátovej kvality a implementáciou procesu riadenia dátovej kvality ako súčasť svojich každodenných činností.

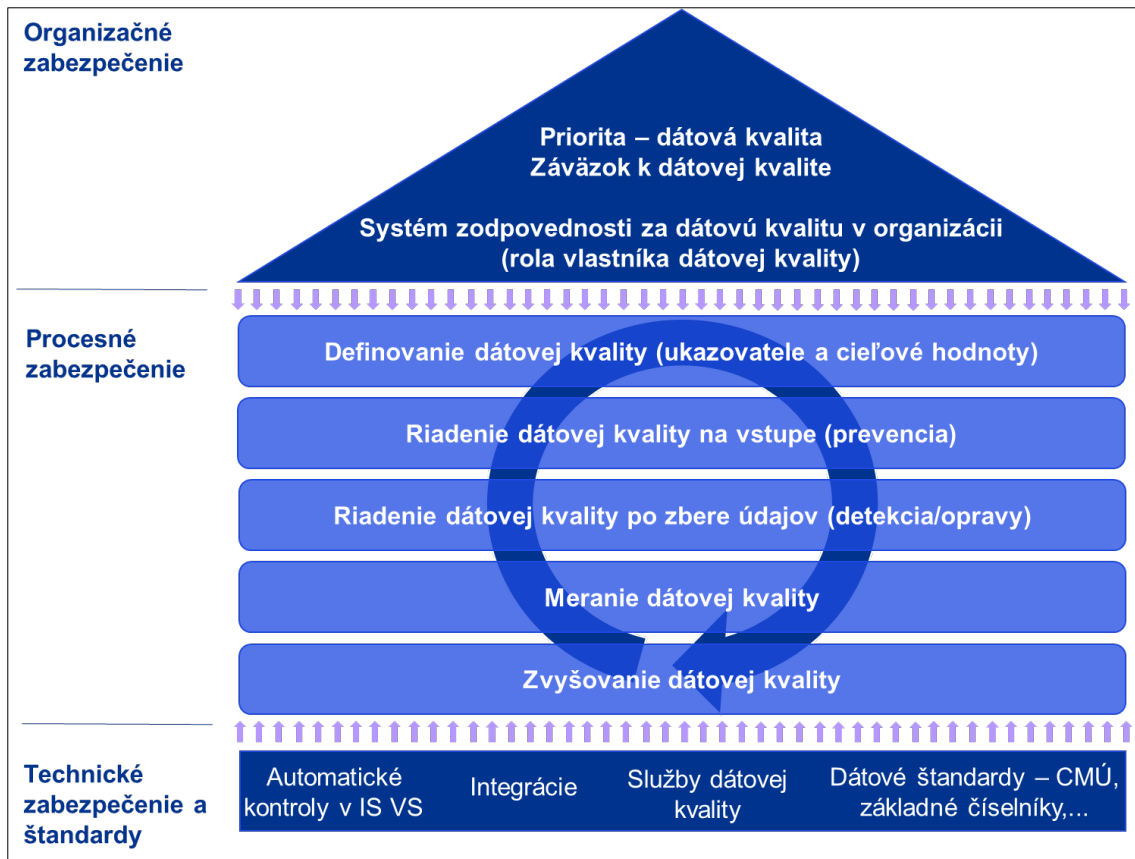
Orgány verejnej moci prispievajú k udržiavaniu náležitej úrovne dátovej kvality a k jej zvyšovaniu:

- definovaním ukazovateľov a cieľových hodnôt ukazovateľov jednotlivých parametrov dátovej kvality, ktoré smerujú k zabezpečeniu presnosti, aktuálnosti, unikátnosti, referenčnej integrity, strojovej spracovateľnosti, konzistentnosti a správnosti údajov evidovaných a zdieľaných v ich informačných systémov,

- organizačným zabezpečením riadenia dátovej kvality v podobe vytvorenia systému zodpovednosti za dátovú kvalitu dovnútra organizácie vrátane zabezpečenia odbornosti a kompetentnosti zamestnancov interagujúcich s dátami,
- zavedením proaktívneho prístupu k riadeniu dátovej kvality v podobe prevencie,
- uplatňovaním procesov merania kvality ako aj vykonávania detekcie a opráv nekvality v údajoch,
- zabezpečením pravidelného merania dátovej kvality v prípade organizácií, ktoré sú správcami referenčných registrov,
- zverejňovaním správ z merania kvality,
- definovaním úloh v podobe nápravných opatrení za účelom zlepšovania kvality a riadením riešenia týchto úloh prostredníctvom ich zaznamenávania v centrálnom registri úloh kvality údajov,
- technickým zabezpečením v podobe implementácie automatických kontrol kvality údajov vo svojich informačných systémoch vrátane zabezpečenia relevantných integrácií,
- zosúladzovaním svojich dátových modelov a štruktúr s príslušnými dátovými štandardmi.

### **4.3 Riadenie dátovej kvality**

Riadenie dátovej kvality je nepretržitý proces vzájomne súvisiacich aktivít smerujúcich k definovaniu, udržaniu a zlepšovaniu dátovej kvality. Definovanie dátovej kvality je nevyhnutným predpokladom jej riadenia – vymedzenie dátovej kvality na účely tohto dokumentu bolo už skôr uvedené. Udržanie dátovej kvality spočíva v nastavení rolí a zodpovedností v rámci organizácie, vrátane realizácie merania dátovej kvality. Zlepšovanie dátovej kvality súvisí predovšetkým s implementáciou nápravných opatrení ako výsledku z merania kvality, ako aj v snahe o proaktívny prístup k riadeniu dátovej kvality.



Obrázok 3 Riadenie dátovej kvality v OVM

#### 4.3.1 Organizačné zabezpečenie riadenia dátovej kvality

Pre OVM je jednou z priorít kvalita údajov v ISVS, ktoré prevádzkuje alebo spravuje – preto záväzok k dátovej kvalite je súčasťou stratégie vedenia a je vytvorený systém zodpovednosti určením vlastníka kvality údajov. Konečnú zodpovednosť za uplatňovanie aktivít súvisiacich s riadením dátovej kvality má vedenie organizácie.

Vedenie musí zabezpečiť:

- odkomunikovanie dátovej kvality ako jednej zo svojich priorít smerom ku všetkým úrovniam riadenia organizácie (dovnútra) ako aj navonok zverejnením záväzku k riadeniu a zvyšovaniu dátovej kvality na svojom webovom sídle,
- vytvorenie roly zodpovednej za proces riadenia dátovej kvality v celej organizácii (v rámci organizácie je určená rola z radov vedenia organizácie zodpovedná za dátovú kvalitu a zároveň táto rola deleguje výkon právomocí a úloh na osobitnú rolu v rámci organizačnej štruktúry napr. prostredníctvom vytvorenia roly dátového kurátora ako delegovaného vlastníka kvality údajov),
- podporu pre všetky aktivity súvisiace s riadením dátovej kvality (vymedzenie personálnych kapacít, finančných prostriedkov, poskytnutie relevantných technológií a pod.).

Vedenie organizácie je centrálnym bodom v rámci zodpovednosti za kvalitu údajov a je dôležité, aby do svojej stratégie jasne definovalo záväzok k riadeniu (udržiavaniu) a zlepšovaniu dátovej kvality, či už smerom dovnútra alebo smerom navonok. Vedenie

organizácie určí delegovaného vlastníka kvality údajov, ktorý zodpovedá za realizáciu procesu riadenia dátovej kvality v organizácii. Hoci delegovaný vlastník sa môže rozhodnúť delegovať niektoré činnosti na ďalšiu úroveň, ale zodpovednosť za riadenie dátovej kvality mu stále zostáva. Delegovaný vlastník kvality údajov zodpovedá najmä za:

- implementáciu a riadenie celého procesu riadenia kvality údajov v organizácii,
- meranie dátovej kvality v organizácii,
- implementáciu nápravných opatrení na zlepšenie dátovej kvality v organizácii,
- súlad údajov s vyhláškou o štandardoch pre ISVS,
- využívanie vyhlásených referenčných údajov v ISVS organizácie,
- vyhlasovanie referenčných údajov, ak je organizácia ich správcom.

Zodpovednosť za kvalitu údajov v organizácii môže byť zverená jednej osobe ale aj celej organizačnej zložke (napr. oddeleniu) v závislosti od charakteru údajov a ISVS, ktoré organizácia spravuje. V oboch prípadoch vedenie organizácie určí sadu výkonnostných ukazovateľov, ktoré budú mať vplyv na hodnotenie osoby/oddelenia zodpovedného za kvalitu údajov, a to konkrétne:

- bolo vykonané meranie dátovej kvality aspoň 1x v priebehu posledného ukončeného kalendárneho roku?
- boli dosiahnuté cieľové hodnoty ukazovateľov dátovej kvality v poslednom ukončenom kalendárnom roku?
- boli implementované nápravné opatrenia určené pre zvýšenie dátovej kvality s termínom v poslednom ukončenom kalendárnom roku?

Hodnotenie takto definovaných výkonnostných ukazovateľov bude vykonávať vedenie organizácie minimálne jedenkrát ročne. Cieľom nastavenia týchto výkonnostných ukazovateľov má byť snaha o motiváciu zamestnancov v oblasti zvyšovania dátovej kvality (nielen formou stanovenia cieľov ale aj motivačných faktorov, medzi ktoré môže patriť napr. finančná motivácia prostredníctvom mimoriadnych odmien v prípade splnenia všetkých výkonnostných ukazovateľov a pod.).

Povinnosťou organizácie v rámci organizačného zabezpečenia je aj zabezpečenie dostatočnej odbornosti osôb, ktorým je pridelená zodpovednosť za kvalitu údajov ako aj osôb, ktoré údaje spracúvajú, resp. participujú v priebehu životného cyklu údajov. Vzdelávanie musí byť zamerané na budovanie kompetencie zamestnancov (odbornú prípravu) ako aj na zvyšovanie povedomia v oblasti kvality údajov. Školenia môžu prebiehať vo forme online vzdelávania, workshopov alebo mentoringu, pričom nemenej dôležitou formou vzdelávania je zdieľanie znalostí a skúseností medzi zamestnancami navzájom alebo aj medzi organizáciami navzájom.

Významným subjektom sú aj konzumenti dát, ktorí môžu pomôcť organizácii pri definovaní požiadaviek na dátovú kvalitu, ako aj pri zvyšovaní povedomia o nej. Je nevyhnutné, aby organizácia proaktívne oslovovala konzumentov dát, spolupracovala s nimi pri riadení dátovej kvality a poskytovala im informácie o kvalite poskytovaných/zverejňovaných dát.

#### 4.3.1.1 *Koordinácia v rámci organizačného zabezpečenia*

Organizačné zabezpečenie riadenia dátovej kvality jednotlivých OVM je centrálné koordinované zo strany MIRRI najmä:

- vymedzením dátovej kvality na legislatívnej úrovni a uložením povinnosti a zodpovednosti jednotlivých OVM za dátovú kvalitu údajov v ISVS, ktoré spravujú alebo prevádzkujú,
- vyčlenením finančných prostriedkov pre OVM za účelom zriadenia kapacít určených na organizačné zabezpečenie riadenia dátovej kvality v jednotlivých OVM,
- pravidelným monitoringom a vyhodnocovaním stavu dátovej kvality, spojeným s podávaním správ a prijímaním nápravných opatrení
- poskytnutím odbornej podpory v podobe konzultácií, školení príp. aj dočasným poskytnutím vlastných kapacít za účelom pomoci pri vytvorení a zavedení systému zvýšenia odbornosti a kompetentnosti zamestnancov OVM zodpovedných za dátovú kvalitu,
- aktívnou komunikáciou na všetkých úrovniach riadenia (nielen napr. vytvorením relevantných pracovných skupín/fór na zdieľanie informácií a skúseností z oblasti dátovej kvality).

#### 4.3.2 **Procesné zabezpečenie riadenia dátovej kvality**

Cieľom procesného zabezpečenia riadenia dátovej kvality je, aby proaktívne riadenie dátovej kvality bolo súčasťou každodenných procesov organizácie. Každý OVM v prvom rade musí poznať údaje, ktoré sú súčasťou jeho ISVS. Za týmto účelom je potrebné určiť aké dáta, z akého zdroja a za akým účelom zbiera. Preto je potrebné, aby organizácia udržiavala a aktualizovala informácie minimálne o:

- používanom dátovom modeli,
- zdroji dát,
- používanom dátovom slovníku,
- používaných referenčných dátach,
- používaných objektoch evidencie.

Takáto dokumentácia musí byť dostupná za všetky ISVS organizácie a vždy musí byť prístupná vlastníčkovi kvality údajov, pretože tieto informácie sú podmienkou pre riadne vykonávanie procesov súvisiacich s celým životným cyklom údajov vrátane riadenia kvality údajov. Prvým krokom je vytvorenie dokumentácie, avšak následne a nemenej dôležité je aktualizovanie tejto dokumentácie v prípade akýchkoľvek zmien. Tieto informácie sú totiž významným vstupom pre riadenie dátovej kvality.

##### 4.3.2.1 *Riadenie dátovej kvality na vstupe (prevencia)*

Pokiaľ má organizácia vytvorenú dokumentáciu a vie aké dáta a za akým účelom získava, je potrebné aby dodržiavala nasledovné pravidlá:

- Organizácia nezbera dáta, ktoré sú referenčnými dátami.

Je nevyhnutné, aby ešte pred zberom údajov organizácia vyhodnotila, či údaj, ktorý potrebuje získať nie je referenčným údajom. Zoznam vyhlásených referenčných registrov je verejne dostupný spolu s uvedením tých údajov registra, ktoré sú vyhlásené za referenčné dáta<sup>28</sup>. Ak je žiadaný údaj referenčným údajom, organizácia údaj nebude zbierať (vytvárať), ale použije hodnotu z referenčného registra (stotožňovanie).

- Organizácia používa definované základné číselníky.

Ešte pred získaním samotných údajov musí organizácia rozhodnúť, aké číselníky sa vzťahujú na dané údaje. Pre zachovanie interoperability údajov je potrebné, aby OVM vo svojich ISVS používali zverejnené základné číselníky<sup>29</sup>. Môže nastať situácia, že potrebný číselník nie je súčasťou zoznamu základných číselníkov (alebo je potrebné upraviť už existujúci číselník), v takom prípade je potrebné číselník definovať a požiadať o jeho zaradenie do zoznamu v súlade s usmernením pre základné číselníky<sup>30</sup>.

- Organizácia používa dátové prvky definované Centrálnym modelom údajov.

Organizácia zosúladí svoj dátový model s centrálnym modelom údajov, t. j. organizácia použije také dátové prvky<sup>31</sup> a definuje ich vo svojom ISVS tak, aby boli v súlade s centrálnym modelom údajov verejnej správy<sup>32</sup>. Pokiaľ centrálny model údajov neposkytuje relevantný dátový prvok pre údaje, ktoré potrebuje OVM zbierať, OVM požiada MIRRI o štandardizáciu a evidenciu nového dátového prvku do centrálného modelu údajov.

- Organizácia určí biznis pravidlá jednotlivých atribútov.

Pre údaje, ktoré organizácia získava, musí na účel riadenia dátovej kvality určiť biznis pravidlá – biznis pravidlá predstavujú formalizované požiadavky pre hodnoty údajov v jednotlivých atribútoch databázy. Sú reprezentáciou toho, čo sú to správne dáta a to na základe existujúcich procesov – preto pri tvorbe biznis pravidiel je veľmi dôležitá znalosť využívania údajov v praxi. Pri definovaní biznis pravidiel je dôležité zapojenie osôb, ktoré reálne dotknutú databázu využívajú v rámci svojej agendy (vecní vlastníci).

Dobrá znalosť biznis pravidiel napomáha pochopiť celý proces využitia konkrétnych údajov, ako aj efektívne merať ukazovatele dátovej kvality – biznis pravidlá sa uplatňujú na jednotlivých atribútoch dátových entít datasetu.

Pri definovaní biznis pravidiel organizácia zohľadní už definované biznis pravidlá v centrálnom zozname biznis pravidiel<sup>33</sup>, pričom pokiaľ organizácia zbiera údaje, ktoré majú byť zdrojovými údajmi niektorého z referenčných registrov, je potrebné, aby organizácia prevzala biznis pravidlá referenčného registra.

- Organizácia uplatňuje automatické kontroly údajov na vstupe.

<sup>28</sup> [RR: Zoznam registrov \(gov.sk\)](#)

<sup>29</sup> [ZČ: Zoznam publikovaných číselníkov \(gov.sk\)](#)

<sup>30</sup> [Číselníky \(gov.sk\)](#)

<sup>31</sup> [Katalóg dátových prvkov \(gov.sk\)](#)

<sup>32</sup> [Štandardy ISVS \(gov.sk\)](#)

<sup>33</sup> [Centrálny zoznam biznis pravidiel – DataLab](#)



Automatické kontroly údajov na vstupe (už pri ich zbere) umožňujú odchytenie prvotných nedostatkov údajov a znižujú tak potrebu následných opráv údajov. ISVS jednotlivých OVM, prostredníctvom ktorých sú údaje získavané, musia mať nastavené automatické kontroly údajov voči nastaveným biznis pravidlám, ako aj voči jednotlivým ukazovateľom dátovej kvality, pri ktorých je možné definovať vzorec pre systémovú kontrolu.

#### 4.3.2.2 Riadenie dátovej kvality po zbere údajov (detekcia/opravy)

Nie všetky „nesprávnosti“ v údajoch je možné eliminovať už na vstupe, teda v rámci prevencie. Pre zavedenie proaktívneho prístupu k riadeniu dátovej kvality v organizácii je nevyhnutné, aby jednotlivé OVM vykonávali aktivity spojené s identifikáciou chýb v údajoch už existujúcich v ISVS, t. j. aktivity v rámci detekcie. Krokom, ktorý bezprostredne nasleduje po detekcii je vykonanie opráv v identifikovaných chybných údajoch. Nástrojom na realizáciu týchto aktivít sú napr. centrálné poskytnuté služby dátovej kvality rozvíjané v rámci projektov Dátovej integrácie (DI) a Centrálnej integračnej platformy (CIP) – detailne popísané v kapitole 3.1.

Základom detekcie je okrem existencie definovaných biznis pravidiel aj vykonanie profilácie údajov. Profilácia údajov predstavuje analýzu údajov pre parametre dátovej kvality správnosť, úplnosť, jedinečnosť, konzistentnosť a primeranosť. Na profiláciu údajov slúžia služby dátovej kvality DQ1 – Základná analýza dátovej kvality a DQ2 – Validácia voči objektu evidencie. Po identifikácii nekvality údajov na základe profilácie dát organizácia využije službu dátovej kvality DQ3 – Asistované čistenie. Prostredníctvom tejto služby je možné zvýšiť kvalitu údajov datasetu prostredníctvom vykonávania nápravných opatrení navrhnutých službou. Spôsob využitia služieb dátovej kvality, ako aj ich funkcie, sú bližšie špecifikované v dokumente č. 1.2.2 Služby dátovej kvality.

#### **Príklady nekvality údajov a spôsoby ich riešenia**

##### **Chýbajúce hodnoty**

Chýbajúce hodnoty nemusia byť problém v prípade, ak sú to hodnoty, ktoré nie sú nevyhnutné pre kompletnosť záznamu. V takomto prípade je však otázne, či tieto hodnoty je potrebné zbierať a v databázach uchovávať. V prípade, ak sú chýbajúce hodnoty podstatné, existujú rôzne techniky, ako ich doplniť:

- identifikovanie referenčného registra, v ktorom sa údaj nachádza a príprava procedúry na doplnenie chýbajúceho údaju,
- využitie imputačných techník, ako napríklad:
  - imputácia využívajúca medián alebo priemer;
  - imputácia s použitím najčastejšej hodnoty alebo nulových / konštantných hodnôt;
  - využitie algoritmu „najbližšieho suseda“;
  - imputácia použitím multivariačnej imputácie pomocou reťazenej rovnice;
  - imputácia využívajúca „Deep Learning“.

##### **Duplicitné záznamy**



Duplicitné záznamy predstavujú problém najmä pre nastavenie jednoznačnej komunikácie, ale aj pre vyhodnocovanie a analýzy, ktoré sa nad údajmi realizujú. Množstvo duplicitných záznamov, s ktorými sa stretávame, patrí do dvoch odlišných typov:

Neunikátne kľúče – jedná sa napr. o dva záznamy v tej istej tabuľke, ktoré majú rovnaký kód alebo kľúč, ale môžu mať rôzne hodnoty a významy. Toto sa môže stať pri pomiešaní údajov alebo v prípade, ak údaje pochádzajú z nedatabázových zdrojov, ako sú textové súbory, (csv súbory z importu csv) alebo súbory programu Excel a pod.

Duplicitný význam – ide o častejší problém a často aj náročnejší na vyriešenie. Tieto druhy duplikátov sú často najškodlivejšie pre dobrú analýzu. Aj keď existujú nástroje na detekciu duplicit (v oblasti mien a adries), často je potrebný práve zásah človeka (ideálne skúseného odborníka), ktorý dokáže správne identifikovať duplicitu. Pre tieto potreby je nevyhnutné aktualizovať mapu duplicit a zaznamenávať ich vyriešenie.

V kontexte vyššie uvedeného je v rámci odstraňovania daného problému možné:

- zlúčenie záznamov – predstavuje komparáciu záznamov a následne zlúčenie údajov z oboch záznamov do jedného, ak sa jedná o dopĺňujúce informácie, ktoré jednotlivé záznamy obsahujú,
- odstránenie duplicitného záznamu – v prípade, ak sa jedná o totožný duplicitný záznam, je potrebné jeden z nich odstrániť.

### **Chybné hodnoty**

Chybné hodnoty v údajoch sú často dôsledkom nezavedeného manažmentu dát a neexistujúceho koncepčného riešenia na sledovanie dátovej kvality. Práve chybné hodnoty sú jedným z adeptov na odstraňovanie preventívnymi prístupmi, ako je kontrola a vynucovanie dodržiavania biznis pravidiel alebo referencovanie na zdrojové registre. Napriek tomu nie je ľahké zabezpečiť kvalitu údajov a správnych hodnôt v databázach v plnej miere, a preto je potrebné vysporiadať sa aj s týmito problémami vo fáze opráv.

Použitie referenčných dát: Ak došlo k problémom v dátach napr. na základe preklepov a chybné údaje boli identifikované, vhodným riešením je využitie referenčných údajov (ak tieto existujú) na „prepísanie“ chybných dát.

Referenčné dáta sú horšie ako dáta v systéme: Môže nastať situácia, kde kvalita údajov v referenčnom registri je horšia ako kvalita dát v zdrojovom systéme. V tomto prípade je potrebné pri identifikácii nezrovnalostí kontaktovať vlastníka referenčného registra a definovať spôsob, ako sa daný údaj v referenčnom registri opraví.

Automatizované opravy nepovoľuje legislatíva: V súvislosti s vyššie uvedeným vzniká niekedy problém, kde automatické opravy v referenčných registroch nie je možné realizovať z titulu legislatívnych obmedzení a teda úprava údajov je často procesne aj časovo zdĺhavá. Opravu musíme vykonať v zdrojovom registri, ktorý údaj pre referenčný register poskytuje.

### **Neaktuálne hodnoty**

Neaktuálne hodnoty v databázach sú často spôsobené buď:

- pasivitou majiteľov údajov (napr. v prípade presťahovania nenahlási obyvateľ zmenu trvalého bydliska alebo neposkytne poisťovniam zmenené údaje), alebo
- zlým nastavením procesu zmeny a zápisu údajov do databázy (zmena statusu konania alebo zmena sudcu a pod.).

Základným nástrojom na eliminovanie týchto problémov je definovanie jasných postupov, ako majú byť konkrétne údaje aktualizované:

- ak ide o údaje, ktoré sú automaticky aktualizované v zdrojových a referenčných systémoch, tieto dáta by sa mali automaticky preniesť do všetkých databáz, ktoré s týmito údajmi pracujú,
- ak ide o zmenu údajov v procese, ktorý dataset popisuje alebo dokumentuje, musí byť jasne určené, kto a akým spôsobom aktualizuje údaj. Tento postup zároveň zabezpečuje prístup do príslušnej databázy.

Zavedením týchto opatrení sa však neodstránia všetky chybné a neaktuálne hodnoty. V tomto prípade by bola riešením len manuálna kontrola a stotožňovanie údajov s realitou v momente, keď sa daný údaj má použiť.

### **Nekonzistentné formáty**

Nekonzistentné formáty často spôsobujú problémy pri spájaní údajov alebo pri vyhodnocovaní rôznych aspektov, ktoré údaje prinášajú. Rovnako vznikajú problémy aj pri využívaní údajov v systémoch, kedy je tvar formátu aj jeho technické prevedenie v poriadku, ale obsahovo je formát rôzny. Napr. v prípade dátumu môže nastať problém – MM/DD/RR vs. DD/MM/RR. Prehodenie dňa a mesiaca môže v systémoch generovať zásadné chyby – napr. ak systém validuje počet mesiacov a zrazu je v hodnote číslo 15. V prípade konaní, pre ktoré sú dátumy podstatné, to môže mať značné negatívne dopady.

Nekonzistentnosť formátov je rovnako kandidátom na posilnenie prevencie, a to najmä nastavením exaktného formátu pre danú položku. Ide o vstupnú validačnú kontrolu jednotlivých položiek.

V prípade, ak sa identifikuje nesprávny formát vo fáze detekcie, je možné ho opraviť nasledovnými spôsobmi:

- nastavením automatizovaného algoritmu zmeny formátov (najčastejšie sa vyskytujúcich) na správny formát a následnú validáciu formátov prostredníctvom validačných nástrojov resp. algoritmov,
- „ručným“ prepísaním formátu do správneho tvaru a nastavením parametrov pre daný formát.

#### **4.3.2.3 Meranie dátovej kvality**

Významným zdrojom pre identifikáciu nekvality v údajoch a potrebu opráv resp. zmien sú výsledky meraní dátovej kvality. Vykonávanie meraní dátovej kvality je veľmi dôležitým nástrojom riadenia dátovej kvality, pričom meranie dátovej kvality je povinné pre všetky OVM, ktoré sú správcami referenčných údajov alebo vlastníkov zdrojových údajov pre niektorý z referenčných registrov.

OVM, ktorých povinnosťou je vykonávať meranie dátovej kvality, realizujú merania pravidelne, minimálne jedenkrát ročne, a to podľa Metodiky merania dátovej kvality vo verejnej správe<sup>34</sup>. Táto metodika predstavuje návod, ako chápať a uskutočňovať meranie dátovej kvality v rámci procesu riadenia dátovej kvality, ako definovať ukazovatele dátovej kvality vrátane ich výpočtu a stanovenia cieľových hodnôt.

Výsledkom každého merania dátovej kvality musí byť report v podobe správy z merania dátovej kvality, ktorý je OVM povinný zverejňovať na svojom webovom sídle. Súčasťou správy z merania dátovej kvality sú:

- informácie o uskutočnenom meraní v podobe karty merania (identifikácia OVM, identifikácia merania, identifikácia ISVS a objektu evidencie/datasetu, časový údaj merania, osoby vykonávajúce meranie a prípadné poznámky),
- metadáta v podobe štruktúrovanej informácie o meranom objekte evidencie/datasetu,
- dátový model znázorňujúci vzťahy jednotlivých atribútov,
- zoznam použitých biznis pravidiel priradených k jednotlivým parametrom dátovej kvality,
- výsledky merania (dosiahnuté hodnoty) v rámci jednotlivých parametrov dátovej kvality,
- návrh nápravných opatrení.

Okrem existujúcej metodiky merania dátovej kvality vo verejnej správe slúži ako pomoc pre OVM v rámci realizácie aktivít spojených s riadením dátovej kvality predovšetkým s meraním dátovej kvality aj videoškolenie zverejnené na oficiálnom kanáli Dátovej kancelárie MIRRI<sup>35</sup>.

#### 4.3.2.4 Zvyšovanie dátovej kvality

Zvyšovanie kvality údajov možno chápať ako kvalitatívne a kvantitatívne zlepšenia parametrov kvality údajov zaznamenané dvomi alebo viacerými meraniami kvality toho istého údaje v priebehu určitého obdobia, v rámci ktorého boli vykonané nápravné opatrenia alebo opatrenia na zamedzenie vzniku nekvalitných údajov.

OVM, ktoré realizujú merania dátovej kvality sú povinné definovať nápravné opatrenia pre zistené nekvality. V rámci tejto činnosti je potrebné rozlišovať, či:

- A: ide o nekvalitu, ktorej potenciálne riešenie je v kompetencii OVM, ktorý meranie vykonával

Príklad: Meranie dátovej kvality vykonané nad datasetom OVM ukázalo nekvalitu v údajoch, ktorá je spôsobená absenciou automatickej kontroly vyplnenosti povinných polí pri zadávaní údajov do ISVS, ktorého je OVM správcom.

- B: ide o nekvalitu, ktorej potenciálne riešenie si vyžaduje súčinnosť iného OVM.

<sup>34</sup> [KPMG KIDS Report v5.2a \(datalab.digital\)](#)

<sup>35</sup> [Školenie - Dátová kvalita - 1. Úvod - YouTube](#)

Príklad: Meranie dátovej kvality vykonané nad údajmi referenčného registra ukázalo nekvalitu v údajoch, ktorá je však spôsobená nekvalitou údajov v zdrojovom registri.

V obidvoch vyššie uvedených príkladoch je potrebné mať zavedené procesy zlepšovania kvality, prostredníctvom ktorých budú takéto prípady riešené.

A: Pokiaľ OVM identifikuje nekvalitu v údajoch, za ktoré je primárne zodpovedný (a oprava je plne v jej kompetencii), je nevyhnutné, aby v rámci svojho interného prostredia:

- zaevidoval identifikovanú nekvalitu a nápravné opatrenie do centrálného registra identifikovaných úloh v dátovej kvalite (vytvorenie a udržiavanie registra úloh v dátovej kvalite je v zodpovednosti vlastníka dátovej kvality organizácie),
- definoval nápravné opatrenie (spôsob riešenia identifikovanej nekvality),
- pridelil zodpovednosť za implementáciu nápravného opatrenia (úlohy) konkrétnej osobe v organizácii,
- stanovil dátum pre implementáciu nápravného opatrenia.

Osoba zodpovedná za implementáciu nápravného opatrenia (väčšinou vlastník dátovej kvality v organizácii, avšak nie nevyhnutne) je povinná zaevidovať vykonanie nápravného opatrenia, príp. dôvody nesplnenia stanoveného termínu a návrh riešenia. Úloha sa považuje za splnenú až preukázaním vyriešenia dátovej nekvality ďalším meraním dátovej kvality. Za uzavretie úlohy z nápravného opatrenia v registri úloh dátovej kvality je vždy zodpovedný vlastník dátovej kvality v organizácii.

B: Pokiaľ OVM identifikuje nekvalitu v údajoch, ktoré spravuje iný OVM, je povinný o tejto skutočnosti informovať zodpovedný príslušný OVM. Informovaný OVM následne postupuje podľa procesu uvedeného vyššie (proces A).

Za účelom zvyšovania dátovej kvality sú všetky OVM povinné aktívne získavať informácie, ktoré môžu byť podnetom pre zvýšenie dátovej kvality, a to najmä:

- prostredníctvom zvyšovania angažovanosti OVM v rámci riadenia dátovej kvality (participácia na relevantných diskusiách odbornej verejnosti napríklad účasťou v príslušných pracovných skupinách, zdieľanie skúseností medzi jednotlivými OVM, oboznamovanie sa s informáciami v oblasti dátovej kvality publikovanými dátovou kanceláriou MIRIRI a pod.),
- pravidelným monitoringom legislatívnych požiadaviek a pripravovaných legislatívnych zmien s dopadom na kvalitu údajov,
- získavaním spätnej väzby od používateľov údajov.

#### 4.3.2.5 *Koordinácia v rámci procesného zabezpečenia*

Procesné zabezpečenie riadenia dátovej kvality jednotlivých OVM je centrálné koordinované zo strany MIRRI najmä:

- vytvorením centrálnych zoznamov základných číselníkov a biznis pravidiel vrátane zverejnenia informácie o postupoch pridávania/úprav v týchto zoznamoch iniciovaných jednotlivými OVM,
- zabezpečenie udržiavania vyššie uvedených zoznamov a zabezpečenie promptnej reakcie na podnety zo strany OVM vrátane vyčlenenia vlastných kapacít za účelom správy príslušných centrálnych zoznamov,
- vytvorením centrálného registra úloh dátovej kvality a monitorovaním plnenia otvorených úloh,
- riadením a podporou vzájomnej súčinnosti jednotlivých OVM pri zvyšovaní dátovej kvality naprieč ISVS.

### 4.3.3 Technické zabezpečenie riadenia dátovej kvality a dátové štandardy

Kvalita údajov a proces riadenia kvality údajov je v OVM cielený na ISVS, ktoré OVM spravuje. Správu ISVS má vykonávať tak, aby OVM najmä:

- Zabezpečil, že prevencia nekvality údajov je v maximálnej možnej miere automatizovaná už pri zbere/vkladaní/vytváraní údajov do jeho ISVS (súčasťou informačných systémov sú automatické kontroly konzistentnosti dát ako aj automatické kontroly dát voči definovaným biznis pravidlám).

Údaje, ktoré pri zbere/zadávaní do informačného systému neprejdú automatickými kontrolami, môžu byť podľa definovaných kontrol automaticky odmietnuté spolu s vygenerovaním informácie o tom, ktorá kontrola bola príčinou nekvality dát (táto informácia je dôležitá pre zabezpečenie čo najrýchlejšej opravy zadávaných údajov, či už ide o zobrazenie hlásenia napríklad pri zadávaní údajov prostredníctvom formulárov alebo ide napríklad o vygenerovanie notifikácie pre OVM, ktoré poskytuje údaje do ISVS prostredníctvom automatizovaných dávok).

Vzhľadom na charakter údajov, ako aj konkrétne kontroly si môže OVM stanoviť, že pre niektoré automatické kontroly (alebo niektoré výsledky automatických kontrol) bude vyžadované individuálne posúdenie dát na vstupe. V takomto prípade nebudú údaje hneď zamietnuté, ale informačný systém bude údaje smerovať na posúdenie na to určenou rolou, ktorá vyhodnotí kvalitu údajov a spôsob riešenia v danom prípade.

- V prípade zavádzania nového informačného systému alebo nového dátového objektu, príp. inej relevantnej zmeny, boli tieto zavádzané už v súlade s aktuálnymi dátovými štandardmi s použitím základných číselníkov a štruktúr centrálného modelu údajov.  
Mnohé ISVS, ktoré sú v súčasnosti používané, nemajú vytvorené dátové modely a štruktúry, ktoré by boli v súlade s centrálnym modelom údajov, základnými číselníkmi, dátovými prvkami a pod. Je nevyhnutné, aby OVM v prípade zavádzania nových informačných systémov alebo v prípade zavádzania zmien, definovali a implementovali požiadavky na zosúladenie svojich dátových štruktúr s centrálnym modelom údajov a základnými číselníkmi.
- V prípade používania referenčných údajov bola vytvorená integrácia na informačné systémy referenčných registrov.  
Informačný systém OVM, ktorý používa referenčné údaje bude mať vytvorenú integráciu na referenčný údaje tak, aby OVM tieto referenčné údaje získaval online. Cieľom je minimalizovať také prípady používania referenčných údajov, ako

napríklad vytváranie ich lokálnych kópií u jednotlivých OVM, čím sa vytvára potenciálna nekvalita minimálne v rámci aktuálnosti takýchto údajov.

- Boli dodržané štandardy pre ISVS definované prostredníctvom aktuálnych usmernení (v súčasnosti Vyhlášky č. 78/2020 o štandardoch pre informačné systémy verejnej správy<sup>36</sup>).
- Na riadenie dátovej kvality, predovšetkým na analýzu dát prostredníctvom ich profilácie, na čistenie dát, stotožnenie s referenčnými údajmi, ako aj na transformáciu dát do strojovo spracovateľnej formy (vždy, keď je to možné) boli využité centrálné poskytované služby dátovej kvality.  
Služby dátovej kvality sú bližšie popísané v dokumente č. 1.2.2 Služby dátovej kvality.

Pri zabezpečovaní dátovej kvality jednotlivé OVM dodržiavajú aktuálne dátové štandardy schválené Komisiou pre štandardizáciu ISVS, ktoré sú verejne dostupné<sup>37</sup>. Pokiaľ je potrebné zavedenie nového štandardu, OVM postupuje podľa publikovaného usmernenia v súlade so štandardizačným procesom<sup>38</sup>.

#### 4.3.3.1 *Koordinácia v rámci technického zabezpečenia*

Technické zabezpečenie riadenia dátovej kvality vrátane štandardov je centrálné koordinované zo strany MIRRI najmä:

- vyčlenením balíka finančných prostriedkov na nástroje zvyšovania dátovej kvality v ISVS pre jednotlivé OVM,
- zabezpečením aktuálnosti a dostupnosti štandardov pre ISVS (technických aj dátových štandardov),
- poskytovaním prístupu k službám dátovej kvality a zabezpečením fungovania a ďalšieho rozvoja týchto služieb.

## 4.4 **Koordinácia a dohľad nad riadením dátovej kvality OVM**

Jednotlivé OVM realizujú riadenie dátovej kvality vo svojich organizáciách na základe aktivít popísaných vyššie v tejto príručke a v súlade s jednotným vymedzením dátovej kvality a záväzku k dátovej kvalite. Vzhľadom na komplexnosť dátovej kvality ISVS a previazanosť resp. zvyšovanie interoperability údajov je nevyhnutné, aby riadenie dátovej kvality bolo koordinované na vyššej, centrálnej úrovni prostredníctvom dátovej kancelárie MIRRI.

OVM v tejto súvislosti vo vzťahu k dátovej kancelárii MIRRI pristupujú ako ku koordinátorovi riadenia dátovej kvality, najmä:

- získavajú centrálné usmernenia (v podobe prevzatia jednotného vymedzenia dátovej kvality, zosúladovania interných aspektov riadenia dátovej kvality so štandardmi a zoznamami publikovanými dátovou kanceláriou MIRRI – za týmto účelom je zo strany koordinátora vytvorený jeden verejne dostupný priestor, ktorý predstavuje znalostnú databázu pre OVM v oblasti riadenia dátovej kvality),

<sup>36</sup> [78/2020 Z.z. - Vyhláška Úradu podpredsedu vlády SR... - SLOV-LEX](#)

<sup>37</sup> [Štandardy ISVS \(gov.sk\)](#)

<sup>38</sup> [Štandardizačný proces \(gov.sk\)](#)



- predkladajú Dátovej kancelárii na posúdenie vlastné stratégie riadenia kvality údajov a plány na zvyšovanie dátovej kvality a ich pravidelné aktualizácie (napr. ale nielen vytvorením a aktualizáciou Príručky kvality)
- aktívne oslovujú dátovú kanceláriu MIRRI v prípade potreby konzultácií a výkladu centrálnych usmernení pri ich implementácii do interného prostredia OVM (za týmto účelom sú zo strany koordinátora dedikované príslušné kapacity),
- reportujú stav dátovej kvality vo svojej organizácii prostredníctvom správ z meraní dátovej kvality a centrálnemu zoznamu úloh dátovej kvality (zo strany koordinátora je vykonávaný pravidelný monitoring stavu dátovej kvality v jednotlivých OVM),
- podrobujú sa pravidelným auditom dátovej kvality vykonávaným zo strany koordinátora.

#### **4.4.1 Koordinačné nástroje zabezpečujúce dohľad nad riadením dátovej kvality OVM**

Okrem poskytovanej podpory zo strany MIRRI je významným aspektom koordinácie riadenia dátovej kvality aj dohľad nad implementáciou procesov riadenia kvality a ich vplyvu na úroveň dátovej kvality u OVM. Za týmto účelom využíva koordinátor dva základné nástroje dohľadu:

- hodnotenie stavu úloh dátovej kvality,
- audity.

##### **4.4.1.1 Hodnotenie stavu úloh dátovej kvality**

Dátová kancelária MIRRI v pozícii koordinátora vedie centrálny register úloh dátovej kvality, ktorý je verejne dostupný. V centrálnom registri úloh dátovej kvality jednotlivé OVM evidujú úlohy, ktoré boli v ich organizácii identifikované v súvislosti s:

- riadením dátovej kvality,
- výsledkami uskutočnených meraní a s nimi spojenými nápravnými opatreniami,
- výsledkami auditov vykonaných zo strany koordinátora.

Úloha v centrálnom registri úloh dátovej kvality je evidovaná minimálne v nasledovnej štruktúre:

- ID úlohy,
- popis úlohy,
- popis riešenia,
- OVM zodpovedný za riešenie úlohy,
- dotknutý parameter kvality,
- termín na implementáciu riešenia úlohy,
- stav úlohy (otvorená/uzatvorená),
- výsledok riešenia.

Dátová kancelária MIRRI pravidelne monitoruje stav a výsledky riešenia úloh zaevidovaných v centrálnom zozname úloh dátovej kvality. Vstupom do monitoringu sú

pravidelné správy o stave dátovej kvality, predkladané OVM (v štruktúre výstupnej súhrnnej správy). Výstupom monitoringu je súhrnná správa, ktorá obsahuje najmä:

- zhodnotenie stavu úloh a ich zmien oproti poslednej správe z monitoringu,
- identifikovanie úloh, ktoré neboli vyriešené v termíne stanovenom pre riešenie úloh,
- informácia o dôvodoch nesplnenia úloh a o ďalšom postupe (po konzultácii so zodpovedným OVM),
- úspešné príklady riadenia dátovej kvality na základe informácií získaných z monitoringu (ak také sú) ako podklad pre zdieľanie skúseností medzi jednotlivými OVM a podpora pri riešení podobných úloh v zodpovednosti rôznych OVM.

Správu z monitoringu zverejňuje dátová kancelária MIRRI na svojom webovom sídle, resp. na dohodnutom verejne dostupnom priestore dedikovanom pre oblasť dátovej kvality vo verejnej správe.

#### 4.4.1.2 *Audity*

Súčasťou koordinácie a riadenia dátovej kvality vo verejnej správe je vykonávanie auditov zameraných na dátovú kvalitu a s tým súvisiacich procesov. Výkon auditov bude zastrešený a realizovaný zo strany dátovej kancelárie MIRRI prostredníctvom tímu audítorov. Cieľom takýchto auditov je preverenie zabezpečenia interného prostredia OVM (organizačného, procesného a technického) vo vzťahu k riadeniu dátovej kvality.

Dátová kancelária bude vykonávať nasledovné typy auditov:

- pravidelné audity,
- ad hoc audity.

Pravidelné audity sa týkajú tých OVM, ktoré sú správcami referenčných údajov (v prípade týchto OVM budú audity realizované minimálne jedenkrát za dva roky). Ad hoc audity sa môžu týkať ktoréhokoľvek OVM, ktorý je zaradený do plánu auditov. Podnetom na zaradenie ad hoc auditu do plánu auditov sú najmä informácie z vyhodnotenia stavu úloh dátovej kvality, informácie získané z už vykonaných auditov, ako aj informácie získané z koordinačnej činnosti dátovej kancelárie MIRRI.

#### **Zásady auditu**

Pri výkone auditu je potrebné dodržiavať nasledujúce zásady.

- Výkon auditu sa uskutoční minimálne 1x u každého OVM v rámci dvojročného cyklu.
- Koordinátor spracováva plán auditov na 2 ročné obdobie. Plán auditov je následne rozpracovaný minimálne na obdobie 1 kalendárneho roku. Plán identifikuje počet vykonávaných auditov v danom roku, vrátane časového vymedzenia jednotlivých auditov, identifikácie auditovaných subjektov a predmetu auditu.
- Výkon individuálneho auditu bude zahájený oznámením o začiatku výkonu auditu.
- Audit vykonáva tím audítorov, tvorený minimálne dvoma osobami za koordinátora, pričom tento tím je možné podľa uváženia rozšíriť o odborníka, ktorý nie je súčasťou auditovaného OVM.



- Audit sa vykonáva podľa plánu auditu, ktorý obsahuje minimálne:
  - cieľ auditu,
  - predmet auditu,
  - kritériá auditu,
  - zloženie audítorského tímu,
  - dátum a miesto, kde bude audit vykonaný,
  - názov a očakávaný čas trvania každej hlavnej preverovanej činnosti,
  - zodpovedných za preverovanú činnosť (funkcia / označenie subjektu/útvary),
  - označenie dokumentov preverovaných v rámci auditu.
- Jedným z podkladov pre realizáciu auditu môžu byť aj existujúce správy z meraní dátovej kvality ako aj správa z hodnotenia stavu dátovej kvality.
- Výstupom auditu bude správa z auditu spracovaná minimálne v rozsahu:
  - informácie o auditovanom subjekte,
  - definovanie zamerania auditu,
  - zhodnotenie predchádzajúcich auditov/zistení (ak relevantné),
  - popis vykonanej audítorskej práce,
  - názor audítora,
  - zistenia,
  - odporúčania pre systémové zlepšenia (nápravné/preventívne opatrenia a námety pre zlepšenie).
- Závery z auditu sú prezentované na záverečnom pohovore audítorov so zodpovednými osobami auditovaného subjektu.
- Auditovaný subjekt má právo sa k záverom z auditu vyjadriť a poskytnúť dodatočné argumenty/dôkazy/dokumenty v prípade, že nesúhlasí so zisteniami uvedenými v správe.
- Audítorský tím dodatočnú dokumentáciu posúdi a v prípade, že je to relevantné upraví závery z auditu.
- Auditovaný OVM je na základe zistení povinný prijať primerané opatrenia a určiť lehoty na realizáciu prijatých opatrení a identifikovať osobu/osoby zodpovedné za ich splnenie.
- Audítorský tím posúdi primeranosť opatrení a po ich vzájomnom odsúhlasení s auditovaným subjektom sú uvedené v správe z auditu.
- Opatrenia formulované v správe z auditu sa prenášajú do centrálného zoznamu úloh dátovej kvality prostredníctvom definovania s tým súvisiacich úloh.

### **Hlavné fázy auditu**

- 1 Zostavenie plánu auditov (koordinátor - tím audítorov)
- 2 Realizácia sebahodnotenia (auditovaný OVM)
- 3 Príprava, poskytnutie a vyhodnotenie podkladov (auditovaný OVM a koordinátor - tím audítorov)
- 4 Realizácia auditu u OVM (koordinátor – tím audítorov)
- 5 Vypracovanie a schválenie správy z auditu (koordinátor - tím audítorov)

## 6 Realizácia odporúčaní na zlepšenie (auditovaný OVM)

## 5 Návrh postupov a metód pre čistenie údajov

Čistenie údajov je dôležitý proces na zabezpečenie ich presnosti a spoľahlivosti. Chyby, nezrovnalosti, duplicity, odľahlé hodnoty a chýbajúce hodnoty môžu ohroziť kvalitu údajov a následné analýzy, čo vedie k chybným poznatkom a neefektívnemu rozhodovaniu. Implementácia efektívnych praktík čistenia dát je nevyhnutná pre organizácie, aby zvýšili hodnotu svojich dátových aktív a podporili informované rozhodovanie.

Samotné čistenie však nie je izolovaný proces, naopak je integrálnou časťou procesov správy dát. Pokiaľ tieto nie sú nastavené správne a konzistentne, nepovedie čistenie dát k želaným výsledkom, čím sa myslí najmä ich dlhodobá udržateľnosť a efektivita. V ďalšom texte teda uvádzame niektoré súvislosti a overené postupy, ktoré je dobré vziať do úvahy pri nastavovaní procesov čistenia dát.

### 5.1 Osvedčené postupy v širšom kontexte správy dát

Aby sa zabezpečilo efektívne čistenie údajov a udržateľnosť výsledkov, organizácie by sa mali riadiť osvedčenými postupmi, ktoré idú nad rámec špecifických techník a zaviesť efektívne procesy správy dát. Posúdenie nastavenia procesov čistenia dát v celkovom rámci správy dát môže vychádzať z odpovedí na otázky z nasledujúcej tabuľky a podľa výsledku posúdenia je potom možné prípadne upraviť nastavenie procesov.

**Tabuľka 79 Posúdenie nastavenia postupov správy dát**

Príklady otázok	Postupy
<p>1. Máme zavedený formálny rámec správy údajov?</p> <p>Sú úlohy a zodpovednosti jasne definované v rámci riadenia údajov?</p> <p>Sú zdokumentované a dodržiavané procesy správy údajov vrátane čistenia údajov?</p> <p>Existuje pravidelný proces hodnotenia alebo auditu na vyhodnotenie účinnosti nášho rámca správy údajov?</p>	<p><b>Rámec správy údajov:</b> Implementácia rámca správy údajov, ktorý zabezpečuje zodpovednosť, vlastníctvo a zodpovednosť za kvalitu údajov v celej organizácii.</p>
<p>2. Ako sú snahy o čistenie údajov koordinované a zosúladené s inými aktivitami správy údajov (ako je napríklad data governance, data stewardship a zabezpečenie kvality údajov)?</p>	<p><b>Integrácia do procesov správy údajov:</b> Integrácia čistenia údajov ako neoddeliteľnej súčasti bežných procesov správy údajov vrátane ich získavania, transformácie a analýzy.</p>
<p>3. Existujú mechanizmy na meranie a sledovanie vplyvu čistenia údajov na metriky kvality údajov?</p> <p>Ako často sa vykonávajú hodnotenia kvality údajov a aké kľúčové ukazovatele výkonnosti sa používajú na hodnotenie úspešnosti činností čistenia údajov?</p> <p>Ako sa výsledky čistenia dát oznamujú (reportujú) a využívajú na neustále zlepšovanie?</p>	<p><b>Monitorovanie a hodnotenie:</b> Pravidelné monitorovanie a vyhodnocovanie účinnosti úsilia o čistenie údajov.</p>

Príklady otázok	Postupy
<p>4. Rozumieme našim údajom, ich štruktúre, vzájomným závislostiam a ich celkovej kvalite?</p> <p>(Príklad typických otázok kladených počas profilovania údajov:</p> <p><i>Spíňa konkrétny stĺpec normy alebo vzory? Je stĺpec údajov zaznamenaný ako reťazec alebo číslo?</i></p> <p><i>Koľko je tam vynechaných hodnôt?</i></p> <p><i>Koľko jedinečných hodnôt je v stĺpci a ako sú rozdelené?</i></p> <p><i>Je tento súbor údajov prepojený s iným alebo má s ním vzťah?)</i></p> <p>5. Existujú zdokumentované pravidlá a usmernenia týkajúce sa kvality údajov, ktorými sa riadia činnosti spojené s čistením údajov?</p> <p>Existujú preddefinované pravidlá alebo protokoly na riešenie chýbajúcich údajov?</p> <p>Existujú pokyny na správu jedinečných hodnôt v stĺpci, ako je identifikácia a riešenie duplikátov alebo nezrovnalostí?</p>	<p><b>Profilovanie údajov:</b> Vykonávanie komplexného profilovania údajov na pochopenie kvality, štruktúry a obsahu súboru údajov pred začatím procesu čistenia.</p> <p><b>Stanovenie pravidiel kvality údajov:</b> Definovanie pravidiel kvality údajov, ktoré sú v súlade s organizačnými cieľmi a požiadavkami a umožňujú konzistentné a systematické čistenie údajov.</p>

## 5.2 Najčastejšie používané postupy a metódy čistenia údajov

Na riešenie takých problémov, akými sú chyby, duplicity, odľahlé hodnoty pri snahe o dodržanie parametrov strojovej spracovateľnosti, konzistencie, aktuálnosti a jedinečnosti, správnosti, presnosti a komplexnosti, by sa mal dodržať pracovný postup čistenia dát, ktorý zahŕňa niekoľko krokov:

**1. Strojová spracovateľnosť:** Údaje musia byť vo formáte, ktorý môžu stroje ľahko spracovať a analyzovať. Na zlepšenie spracovateľnosti strojov je možné použiť nasledujúce metódy:

- Štandardizácia formátov dátumu a číselných vyjadrení.  
*Ak napríklad množina údajov obsahuje dátumy v rôznych formátoch ako „MM/DD/RRRR“, „DD-MM-RRRR“ alebo „RRRR/MM/DD“, všetky dátumy sa skonvertujú do jednotného formátu (napr. „RRRR-MM-DD“), čo zabezpečí jednotnosť, zjednodušuje výpočty založené na dátume a umožňuje lepšiu automatizáciu analýz súvisiacich s časom.*
- Konverzia kategorických údajov na numerické reprezentácie  
*Napríklad stroje často vyžadujú numerické vstupy na analýzu. Konverzia kategorických premenných ako „Vysoká“, „Stredná“ a „Nízka“ na číselné hodnoty, ako sú 3, 2 a 1 (alebo jednorazové kódovanie), uľahčuje matematické operácie a modelovacie úlohy.*
- Riešenie špeciálnych znakov, symbolov alebo problémov s kódovaním.

*Ide napríklad o odstránenie medzier na začiatku/na konci, konverziu nekonzistentných formátov a opravu typografických chýb. Štrukturálne čistenie zvyšuje integritu údajov a uľahčuje bezproblémovú analýzu a integráciu.*

**2. Konzistencia:** zaisťuje, že dátové prvky majú rovnaký význam a formát v celom súbore údajov. Na dosiahnutie konzistentnosti je možné použiť nasledujúce metódy čistenia údajov:

- Štandardizácia jednotiek (napr. regionálne názvy a kódy).  
*Napríklad v množine údajov, ktorá obsahuje adresy zákazníkov z rôznych regiónov Slovenska môžu existovať rozdiely v reprezentácii regiónov, ako napríklad „Banská Bystrica“ vs. „Banskobystrický kraj“. Štandardizáciou regionálnych názvov a kódov do spoločného formátu, ako je použitie oficiálnych ISO kódov (napr. „SK-BC“ pre Banskobystrický kraj), môžu byť údaje konzistentné a ľahšie analyzovateľné.*
- Zabezpečenie konzistentných konvencií pomenovania premenných a kategórií.  
*Veľmeme si napríklad súbor údajov obsahujúci informácie o vládnych výdavkoch v rôznych sektoroch. Súbor údajov môže mať nekonzistentné konvencie pomenovania pre sektory, ako napríklad „Vzdelávanie“, „Vzdelávanie“ a „Vzdelávacie služby“. Na zabezpečenie konzistentnosti je možné implementovať proces čistenia údajov, aby sa štandardizovali konvencie pomenovania do spoločného formátu, ako je napríklad „Vzdelávanie“. Táto štandardizácia uľahčí agregáciu a presnú analýzu údajov v rôznych sektoroch.*

**3. Aktuálnosť a jedinečnosť:** Aktuálnosť sa vzťahuje na aktuálnosť údajov a jedinečnosť zaisťuje, že každý záznam je odlišný. Na zachovanie aktuálnosti a jedinečnosti možno použiť nasledujúce metódy:

- Identifikácia a aktualizácia zastaraných záznamov o aktuálne informácie.  
*V databáze môžu mať niektoré záznamy zastarané kontaktné informácie. Krížovým odkazom na údaje s najnovšími záznamami alebo použitím externých zdrojov údajov je možné aktualizovať podrobnosti, aby sa zabezpečili presné a aktuálne informácie.*
- Odstránenie duplicitných záznamov.  
*Identifikácia a odstránenie duplicitných záznamov je rozhodujúce pre presnú analýzu. Odstránenie duplikátov zlepšuje integritu údajov, eliminuje redundanciu a umožňuje efektívne ukladanie a analýzu.*
- Používanie verzií údajov alebo časovej pečiatky na sledovanie zmien v súbore údajov.  
*Napríklad, keď sa 15. júla 2023 prijímú nové údaje o miere nezamestnanosti pre konkrétny región, budú označené ako „15-07-2023 10:00 AM“. Časová pečiatka zachytáva presný okamih, kedy boli údaje zaznamenané alebo aktualizované, čím sa zabezpečuje pôvod údajov a pomáha pri sledovaní línie údajov.*
- Identifikácia chýbajúcich hodnôt  
*Chýbajúce hodnoty sa môžu vyskytnúť v dôsledku neúplného zadania údajov, straty údajov alebo úmyselného vynechania. Techniky ako imputačný odhad alebo vyplnenie chýbajúcich hodnôt na základe existujúcich vzorov. Riešenie chýbajúcich hodnôt minimalizuje odchýlky a medzery v údajoch, čo umožňuje spoľahlivejšiu analýzu.*

**4. Správnosť:** zahŕňa zabezpečenie toho, aby údaje presne reprezentovali entity alebo udalosti skutočného sveta. Na zvýšenie správnosti obvykle slúžia nasledujúce metódy čistenia údajov:

- Krížové odkazy na údaje s dôveryhodnými (externými) zdrojmi na overenie ich presnosti.

*Napríklad prepojenie údajov so spoľahlivými referenčnými registrami, ako je Register právnických osôb (RPO) a Register adries (RA), zvyšuje presnosť a úplnosť údajov. Prepojením údajov s týmito externými zdrojmi organizácia zaistí, že jej údaje sú v súlade s dôveryhodnými a aktuálnymi informáciami.*

- Aplikovanie pravidiel overovania údajov na kontrolu nezrovnalostí alebo nelogických hodnôt.

*Napríklad v prípade množiny údajov, ktorá obsahuje informácie o veku, použitie pravidiel overenia na kontrolu, či vek spadá do primeraného rozsahu (napr. 1 až 120), pomáha identifikovať a opravovať chybné záznamy. Napríklad označovanie záznamov s vekom mimo očakávaného rozsahu na ďalšie preskúmanie a overenie.*

**5. Presnosť a komplexnosť:** úzko súvisí so správnosťou a vzťahuje sa na to, ako dobre údaje predstavujú skutočné hodnoty alebo stavy. Metódy na zlepšenie presnosti zahŕňajú:

- Profilovanie údajov

*Analýzou a vizualizáciou údajov pomocou štatistických metód, ako sú priemer, štandardná odchýlka, rozsah alebo kvantily, je možné nájsť neočakávané a chybné hodnoty. V tejto fáze môže byť užitočná aj vizualizácia. Napríklad vizualizáciou priemerného príjmu v regiónoch možno zistiť, či existujú nejaké odľahlé hodnoty.*

- Odstránenie odľahlých hodnôt

*Tento krok eliminuje dátové body, ktoré sa výrazne odchyľujú od očakávaných vzorov. Odľahlé hodnoty môžu vzniknúť v dôsledku chýb merania alebo anomálií. Odstránenie odľahlých hodnôt zvyšuje presnosť údajov a zosúladuje súbor údajov s väčšinou údajových bodov.*

- Zjednodušenie dátových štruktúr

*Napríklad súbor údajov obsahuje širokú škálu premenných vrátane meteorologických údajov, úrovni znečisťujúcich látok a geografických súradníc. Na zníženie zložitosti a zlepšenie ovládateľnosti môžu byť denné údaje agregované do mesačných priemerov (samozrejme v prípadoch, keď to je vecne prípustné).*

Dátová kvalita a s tým spojená potreba čistenia údajov je relevantnou témou počas celého životného cyklu údajov (od ich získania, evidovania v systéme až po ich výmaz). Možno teda hovoriť o istej postupnosti, ktorá sa vzťahuje k čisteniu údajov:

#### 1 Prevencia

*Kedykoľvek je to možné, je treba uprednostniť čistenie údajov už na vstupe, t. j. už pri zbieraní údajov, resp. pri zadávaní údajov do systému. V rámci prevencie sa snažíme zabrániť vstupu zlých údajov do systému, pričom čím efektívnejšia prevencia je, tým je nižšia pravdepodobnosť potreby náročných a nákladných opráv v systémoch.*

#### 2 Detekcia

*Detekcia zahŕňa identifikovanie nekvalitných, chybných či chýbajúcich údajov, ktoré už v systéme sú – predpokladom účinnej detekcie je uplatňovanie proaktívneho prístupu k vyhľadávaniu zlých údajov, ktorý je realizovaný pravidelne.*

### 3 Opravy

*V prípade nekvalitných, chybných či chýbajúcich údajov, ktoré neboli „odchytené“ na vstupe, je potrebné vykonať opravy.*

Z pohľadu zabezpečenia dátovej kvality sú významné všetky tri vyššie uvedené oblasti, avšak organizácia by sa mala snažiť znižovať potrebu vykonávania opráv využitím relevantných metód už v rámci prevencie i detekcie tak, aby postupne minimalizovala zásahy do systému súvisiace s opravou údajov.

## 5.3 Návrh postupov čistenia údajov hodnotených registrov

Každý z hodnotených registrov už má implementované niektoré mechanizmy čistenia údajov a zvyšovania dátovej kvality. Všetky vymenované metódy čistenia údajov sú popísané v predchádzajúcej kapitole.

### 5.3.1 RPO

Kvalita dát RPO je významne ovplyvnená kvalitou dát jeho zdrojových registrov. Z toho dôvodu je vhodnejšie, aby riadenie dátovej kvality RPO do väčšej časti zahrnulo aj riadenie dátovej kvality jeho zdrojových registrov. Ako je uvedené v kapitole 3.3 Odporúčania pre RPO, ŠÚ SR by mal v spolupráci s Dátovou kanceláriou vytvoriť metodiku pre zdrojové registre RPO na poskytovanie údajov vo vysokej kvalite a v rámci nej sa zamerať na metódy a postupy čistenia údajov vhodné pre jednotlivé registre.

### 5.3.2 RA

Riadenie DK v RA už má implementované všetky metódy čistenia dátovej kvality, ktoré boli potrebné. Z hľadiska postupov sa správa registra zameriava na prevenciu vzniku nových nekvalitných záznamov, detekciu existujúcich nekvalít v historických záznamoch a čistenie záznamov, u ktorých boli zistené nezhody. V rámci prevencie sa implementovali tieto postupy:

- Štandardizácia formátov dátumu a číselných vyjadrení – pri vytváraní nových záznamov sú vynútené kontroly numerických údajov a dátumov.
- Riešenie špeciálnych znakov, symbolov alebo problémov s kódovaním – pri vytváraní nových záznamov sú vynútené kontroly neplatných alfanumerických znakov a nadbytočných medzier.
- Štandardizácia jednotiek (napr. regionálne názvy a kódy) – register využíva centrálné číselníky pre registre verejnej správy a má definované aj vlastné odvetvové číselníky.
- Odstránenie duplicitných záznamov – kontroly pri vytváraní záznamu zamedzujú vytvoreniu duplicitných záznamov.
- Používanie verzií údajov alebo časovej pečiatky na sledovanie zmien v súbore údajov – register uchováva časové rezy záznamov nielen s rozdelením na platné a neplatné záznamy, ale aj so schopnosťou identifikovať platné hodnoty v konkrétnom momente v minulosti.
- Identifikácia chýbajúcich hodnôt – pri vytváraní záznamu kontroly vynucujú kompletnosť povinných atribútov.



V rámci detekcie nekvality v historických záznamoch a čistenia údajov využívajú všetky vyššie uvedené postupy, vrátane aplikovania pravidiel overovania údajov na kontrolu nezrovnalostí alebo nelogických hodnôt.

### 5.3.3 RNZ

Riadenie dátovej kvality RNZ sa v súčasnosti zameriava len na prevenciu vzniku chýb pri nových záznamoch. Správca registra nevykonáva žiadne aktivity v oblasti detekcie nekvalitných záznamov a čistenia údajov. Pri vytváraní nových záznamov boli doteraz implementované nasledujúce postupy:

- štandardizácia formátov dátumu a číselných vyjadrení,
- štandardizácia jednotiek,
- identifikácia chýbajúcich hodnôt.

V súvislosti s implementáciou nového informačného systému registra a so zmenou RNZ na referenčný register by sa mal správca registra zamerať aj na detekciu už existujúcich nekvalitných záznamov a čistenie údajov. Hoci meranie preukázalo vysokú dátovú kvalitu meraného datasetu, postupy detekcie slúžia nielen na zistenie chýb, ale aj na preukázanie dátovej kvality. Správca registra by mal s použitím týchto postupov vykonávať pravidelné kontroly kvality registra a sprístupniť ich výsledky konzumentom. Na zväzovanie je, aby boli v riadení dátovej kvality RNZ zakomponované aj nasledujúce postupy:

- Riešenie špeciálnych znakov, symbolov alebo problémov s kódovaním.
- Štandardizácia jednotiek (najmä využívanie centrálnych číselníkov).
- Identifikácia a aktualizácia zastaraných záznamov o aktuálne informácie.
- Odstránenie duplicitných záznamov.
- Používanie verzií údajov alebo časovej pečiatky na sledovanie zmien v súbore údajov.
- Aplikovanie pravidiel overovania údajov na kontrolu nezrovnalostí alebo nelogických hodnôt.

Vzhľadom na vyššie percento zistených nezhôd v údajoch preberaných zo ŠRO je nevyhnutné, aby boli údaje pokutovaných subjektov pri vytváraní nových záznamov stotožňované voči aktuálnym údajom RPO. Pri následných kontrolách dátovej kvality by mal byť využitý postup krížových odkazov na údaje RPO na overenie ich presnosti.



## 6 Návrh algoritmov pre zlepšenie dátovej kvality

V tejto kapitole sa pokúsime uviesť niekoľko inovatívnych prístupov k zlepšovaniu kvality dát, ktoré uvádzame ako inšpiráciu pre ďalší rozvoj správy dát v dimenzii dátovej kvality. Nakoľko výstup 1.2.2 Služby dátovej kvality dáva väčší priestor detailom existujúcich služieb dátovej kvality, ako i súvisiacemu návrhu automatizácie, záujemcu o ďalšie informácie teda odkazujeme na tento výstup.

Avšak, ako každá metóda, aj tieto inovatívne metódy majú svoje limity a sú situácie, kedy je lepšie použiť metódy tradičnejšie. Rozhodnutie, kedy použiť pokročilé metódy čistenia údajov (ako je napr. AI/ML), v porovnaní s tradičnými metódami, závisí od niekoľkých faktorov. Každý prístup má svoje silné a slabé stránky a výber by mal byť založený na špecifických charakteristikách spracúvaných údajov a cieľoch procesu čistenia údajov. Tu je niekoľko úvah, ktoré pomôžu usmerniť rozhodnutie, kedy použiť AI/ML na čistenie údajov:

*Zložitosť a objem údajov:* Metódy AI/ML vynikajú pri práci s veľkými a zložitými súbormi údajov. Tradičné algoritmicke metódy môžu mať problém zvládnuť veľké množstvo údajov alebo keď majú údaje zložité vzory a vzťahy, prípadne sa vyžaduje spracúvanie v (semi) reálnom čase. Modely strojového učenia sa môžu učiť zo vzorov a prispôbovať sa zmenám v údajoch, vďaka čomu sú pre takéto prípady vhodnejšie.

*Neštruktúrované alebo čiastočne štruktúrované údaje:* Ak sú údaje neštruktúrované alebo čiastočne štruktúrované (napr. text, obrázky, zvuk), metódy AI/ML (ako napr. spracovanie prirodzeného jazyka - NLP) alebo počítačové videnie môžu byť vysoko efektívne. Tradičné algoritmy sú zvyčajne navrhnuté pre štruktúrované údaje a ich použitie pre neštruktúrované údaje nemusí priniesť presné výsledky.

*Automatizácia procesu:* Techniky AI/ML môžu často automatizovať úlohy čistenia údajov, čím sa znižuje potreba manuálneho zásahu. Tradičné metódy môžu vyžadovať viac manuálneho kódovania založeného na pravidlách, čím sú menej efektívne pre väčšie súbory údajov alebo časté aktualizácie.

*Nejednoznačnosť a chybovosť údajov:* Modely AI/ML dokážu zvládnuť neistotu a chyby v údajoch elegantnejšie ako tradičné algoritmy. Dokážu odvodiť chýbajúce hodnoty, opraviť chyby a ľahšie identifikovať odľahlé hodnoty, vďaka čomu sú vhodné pre súbory údajov s inherentným šumom alebo chýbajúcimi informáciami.

*Škálovateľnosť a prispôbivosť:* Keď sa proces čistenia údajov musí škálovať s rastom údajov alebo sa musí prispôbiť meniacim sa vzorom údajov v priebehu času, modely AI/ML môžu byť flexibilnejšie a robustnejšie. Tradičné metódy môžu vyžadovať časté úpravy, aby sa vyrovnali s novými výzvami.

*Skúmanie komplexných vzťahov:* Techniky AI/ML môžu identifikovať zložité vzťahy medzi premennými, ktoré nemusia byť evidentné prostredníctvom tradičných prístupov. Môžu odhaliť skryté vzorce, korelácie a závislosti v údajoch, čo vedie k presnejšiemu čisteniu a lepšiemu vhľadu do charakteru údajov.

*Odborné znalosti a zdroje domény:* Metódy AI/ML môžu vyžadovať viac odborných znalostí v oblasti strojového učenia a prístupu k výpočtovým zdrojom, ako sú napríklad výkonné hardvérové platformy a platformy na spracovanie údajov. Tradičné metódy môžu byť dostupnejšie pre tých, ktorí majú obmedzené technické schopnosti alebo zdroje.

Stručne povedané, pokročilé metódy čistenia údajov, ako je AI/ML, sú najprínosnejšie pri práci s veľkými, komplexnými, neštruktúrovanými súbormi údajov alebo keď sú kritické automatizácia, škálovateľnosť a adaptabilita. Avšak pre menšie, dobre štruktúrované súbory údajov s jednoduchými potrebami čistenia údajov môžu byť tradičné algoritmické metódy stále efektívne a ľahšie implementovateľné. Výber by mal byť založený na špecifických charakteristikách údajov, projektových požiadavkách a dostupných zdrojoch.

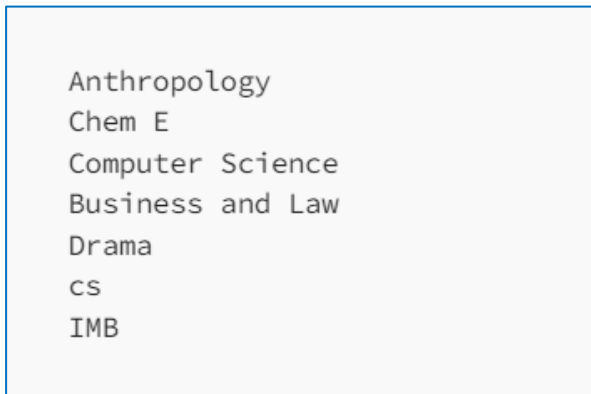
Na ilustráciu postupov uvádzame viacero prípadov použitia strojového učenia (ML) a rozsiahlych jazykových modelov (LLM) na zlepšenie kvality dát. Tieto sa dopĺňajú s informáciami, uvedenými vo výstupe 1.2.2 Služby dátovej kvality, preto odporúčame vnímať ich v spoločnom kontexte oboch výstupov.

## 6.1 Modelové prípady použitia pokročilých algoritmických metód

### 6.1.1 Study Fetch – Zefektívnenie čistenia údajov pomocou veľkých jazykových modelov (LLM)

Tento prípad použitia sa zameriava na súbor údajov obsahujúci neštruktúrované textové údaje, ktoré predstavovali výzvy pre tradičné prístupy čistenia údajov<sup>39</sup>.

Spoločnosť Study Fetch uskutočnila prieskum medzi vysokoškólakmi, pričom získala viac ako 10 000 odpovedí. Pri analýze údajov však narazili na problém. „Hlavným“ poľom v prieskume bolo pole s voľným textom, ktoré umožnilo respondentom zadať ľubovoľnú hodnotu. Výsledkom boli neštruktúrované a nekonzistentné údaje, čo predstavovalo ťažkosti pri zmysluplnej analýze.



```
Anthropology
Chem E
Computer Science
Business and Law
Drama
cs
IMB
```

#### Obrázok 4 Surové výsledky prieskumu

Aby dali týmto údajom zmysel, potrebovali riešenie, ktoré by mohlo tieto odpovede kategorizovať do známych hlavných oblastí. V snahe očistiť tieto údaje od herézy a ušetriť hodiny triedenia v Exceli sa núkal alternatívny prístup: použitie výkonného veľkého jazykového modelu (LLM).

Rozhodli sa využiť GPT (Generative Pre-trained Transformers spoločnosti OpenAI, výkonné LLM s viac ako 175 miliardami parametrov. Tento model bol trénovaný na veľkom množstve webových údajov a mohol využiť posilňujúce učenie s ľudskou

<sup>39</sup> [From Chaos to Clarity: Streamlining Data Cleansing Using Large Language Models | by Naresh Ram | Jun, 2023 | Towards Data Science \(medium.com\)](#)

spätnou väzbou na poskytnutie presných odpovedí. Vyzbrojení týmto LLM boli pripravení riešiť úlohu čistenia dát.

Tím vytvoril zadanie, ktoré obsahovalo preddefinované kategórie pre vysokoškolské študijné programy:

- 1 Umenie a humanitné vedy: Literatúra a umenie.
- 2 Spoločenské vedy: psychológia, sociológia a politológia.
- 3 Obchod a manažment: Marketing, financie, ekonómia a MBA.
- 4 Biologické vedy: Biológia, environmentálne vedy a oblasti súvisiace s poľnohospodárstvom.
- 5 Inžinierstvo a technológia: Elektrotechnika, strojárstvo atď.
- 6 Medicína, zdravie a súvisiace vedy: Ošetrovateľstvo, lekárska škola atď.
- 7 Všetky ostatné programy: Všetko, čo nespadá do vyššie uvedených kategórií.

Pomocou tohto zadania požiadali LLM, aby odpovede z prieskumu podľa toho kategorizoval. Výstup LLM poskytol mapovanie medzi poradovými číslami odpovedí prieskumu a zodpovedajúcimi kategóriami programov. Napríklad odpoveď „Dráma“, bola priradená k „Umenie a humanitné vedy“ a odpoveď „IMB“, bola priradená k „Všetky ostatné programy“.

S pomocou LLM úspešne kategorizovali odpovede prieskumu a získali cenné poznatky. Ručné zadávanie každého bloku odpovedí prieskumu do prehliadača a výpočet mapovania by však bolo časovo náročné a náchylné na chyby. Na prekonanie tejto výzvy využili API OpenAI, ktoré im umožnilo automatizovať proces čistenia dát. Tím využil REST rozhranie na interakciu s LLM. Uskutočnili volania API pomocou Node.js a knižnice Node OpenAI, pričom odovzdali zadanie a ďalšie relevantné parametre. Každá interakcia vygenerovala odpoveď obsahujúcu požadované mapovania. Implementovali slučku na spracovanie údajov v dávkach, čím sa zabezpečilo efektívne vykonávanie. Využitím API a automatizáciou procesu výrazne znížili manuálne úsilie a dosiahli úsporu nákladov. Celá úloha čistenia dát, vrátane viacerých spustení a testovania, stála menej ako 1 dolár a trvala približne tri hodiny kódovania a samotného behu programu.

Okrem čistenia údajov je možné LLM použiť aj na rôzne iné úlohy súvisiace s údajmi. Niektoré potenciálne prípady použitia zahŕňajú:

- analýza a štandardizácia údajov,
- deduplikácia údajov,
- sumarizácia údajov a
- analýza sentimentu.

Tieto modely majú potenciál zefektívniť analýzu údajov a poskytnúť cenné poznatky v rôznych oblastiach. Je však dôležité poznamenať, že LLM by sa mali používať v spojení s inými technikami a ľudským dohľadom. Sú to mocné nástroje, ale môžu so zdanlivou istotou generovať nesprávne informácie, známe ako halucinácie. Správna doménová odbornosť, porozumenie kontextu a manuálna kontrola sú nevyhnutné na zabezpečenie integrity údajov.

Na záver možno povedať, že využitím veľkých jazykových modelov, ako je GPT, môžu podniky zefektívniť procesy čistenia údajov a efektívnejšie získať cenné poznatky. Tieto modely v kombinácii s automatizáciou prostredníctvom rozhraní API poskytujú

nákladovo efektívne riešenia problémov súvisiacich s údajmi. Či už ide o normalizáciu odpovedí na prieskum, analýzu zákazníckych recenzií alebo zlepšenie kvality údajov, LLM ponúkajú obrovský potenciál pre transformáciu úloh zameraných na údaje.

## 6.1.2 Uber – Zabezpečenie kvality údajov pomocou strojového učenia

Uber, revolučná platforma pre zdieľanie jász (u nás sa jedná o taxislužbu), funguje v masívnom rozsahu približne 14 miliónov jász denne. Riadenie kvality údajov v takom obrovskom rozsahu predstavuje skľučujúcu zodpovednosť. Na riešenie tohto problému Uber využíva mechanizmus známy ako Data Quality Monitor (DQM), ktorý využíva silu strojového učenia na zabezpečenie integrity údajov<sup>40</sup>.

**Detekcia anomálií pomocou modelu Holt-Wintersovho časového radu<sup>41</sup>:** DQM využíva pokročilé štatistické techniky, ako je analýza hlavných komponentov<sup>42</sup>, na analýzu historických vzorcov a trendov predpovedaných údajov pomocou modelu Holt-Wintersovho časového radu. Porovnaním projektovaných hodnôt s historickými trendmi DQM efektívne zisťuje anomálie v reálnom čase. Tento prístup pomáha Uberu identifikovať akúkoľvek odchýlku od očakávaných vzorcov údajov a okamžite ich upozorní na potenciálne problémy s kvalitou údajov.

**Bodovanie a stanovovanie priorít:** Aby sa predišlo zahlcovaniu vlastníkov údajov množstvom nepretržitých upozornení, DQM používa premyslený prístup priradovaním skóre k anomáliám na základe ich závažnosti. Uber uprednostnením upozornení podľa ich vplyvu zaisťuje, že vlastníci údajov sa môžu najskôr zamerať na najkritickejšie problémy, čím zefektívnia svoje úsilie o zlepšenie kvality údajov.

## 6.1.3 Netflix – Detekcia anomálií a ďalšie so strojovým učeníím

Netflix, globálna populárna streamovacia služba nastavuje štandard pre prispôsobené odporúčania obsahu a bezproblémové používateľské skúsenosti a rozširuje svoje schopnosti založené na ML nad rámec odporúčaní obsahu, aby riešil problémy s kvalitou údajov<sup>43</sup>.

**Monitorovanie stavu systému pomocou algoritmov klastrovania založených na hustote (Density-Based Clustering Algorithm - DBSCAN):** Hoci je primárne navrhnutý na monitorovanie stavu systému, Netflix uplatňuje princípy algoritmu DBSCAN na identifikáciu anomálií v údajoch. Využitím algoritmu DBSCAN dokáže Netflix efektívne identifikovať systémy, ktoré sa odchyľujú od normy, čím pomáha udržiavať kvalitu údajov s pôsobivou mierou presnosti 93 %.

Tu je podrobnejšie vysvetlenie toho, ako DBSCAN funguje:

- 1 Zhlukovanie na základe hustoty: DBSCAN zoskupuje dátové body na základe ich hustoty vo vysokorozmernom priestore. Definuje dva dôležité parametre: „epsilon“

<sup>40</sup> [How Are the Tech Giants Ensuring Data Quality with ML? | by Jyoti Dhiman | Geek Culture | Medium](#)

<sup>41</sup> [7.3 Holt-Winters' seasonal method | Forecasting: Principles and Practice \(2nd ed\) \(otexts.com\)](#)

<sup>42</sup> Principal component analysis (PCA) is a popular technique for analyzing large datasets containing a high number of dimensions/features per observation, increasing the interpretability of data while preserving the maximum amount of information, and enabling the visualization of multidimensional data. Formally, PCA is a statistical technique for reducing the dimensionality of a dataset.

<sup>43</sup> [How Are the Tech Giants Ensuring Data Quality with ML? | by Jyoti Dhiman | Geek Culture | Medium](#)

- ( $\epsilon$ ), ktorý predstavuje maximálnu vzdialenosť medzi dvoma dátovými bodmi, aby sa mohli považovať za súčasť toho istého klastra, a „minPts“, ktorý nastavuje minimálny počet bodov potrebných na vytvorenie hustého región (klastera).
- 2 **Reprezentácia údajov:** Netflix pravdepodobne predstavuje správanie svojich systémov ako údajové body vo viacrozmernom priestore. Každý údajový bod môže obsahovať rôzne atribúty alebo metriky, ktoré popisujú stav a výkon systému v danom čase. Tieto atribúty môžu zahŕňať veci ako využitie centrálnej procesorovej jednotky, spotreba pamäte, sieťová prevádzka, časy odozvy a ďalšie relevantné systémové metriky.
  - 3 **Identifikácia normálneho správania:** Aby mohol Netflix monitorovať stav svojich systémov, musí najprv stanoviť základnú líniu normálneho správania. Zvyčajne sa to robí zhromažďovaním historických údajov zo systémov počas pravidelných prevádzkových období, keď všetko funguje správne. Algoritmus DBSCAN sa potom použije na tieto historické údaje, aby sa identifikovali zhluky normálneho správania. Dátové body, ktoré spadajú do týchto zhlukov, predstavujú typické správanie systému za normálnych okolností.
  - 4 **Detekcia anomálií:** Akonáhle je stanovená základná línia normálneho správania, algoritmus DBSCAN sa aplikuje v reálnom čase na nové prichádzajúce dátové body zo živých systémov. Ak nový údajový bod spadá do jedného zo zavedených zhlukov normálneho správania, považuje sa za súčasť normy a nepodniknú sa žiadne ďalšie kroky.
  - 5 **Identifikácia anomálií:** Ak nový údajový bod nespadá do žiadneho z vytvorených klastrov, považuje sa za anomáliu alebo odľahlú hodnotu. Anomálie môžu naznačovať potenciálne problémy, abnormality alebo poruchy v systéme. Monitorovací systém Netflixu pravdepodobne spustí výstrahu alebo prijme vhodné opatrenia, keď sa zistí anomália. Tieto akcie môžu zahŕňať upovedomenie systémových administrátorov, zaprotokolovanie udalosti pre ďalšiu analýzu alebo spustenie automatických procedúr obnovy, aby sa zabezpečilo, že si systém zachová svoj celkový výkon a kvalitu.
  - 6 **Ladenie parametrov:** Presnosť procesu detekcie anomálií je ovplyvnená výberom parametrov epsilon ( $\epsilon$ ) a minPts. Nájdenie optimálnych hodnôt pre tieto parametre môže zahŕňať experimentovanie a doladovanie tak, aby vyhovovali špecifickým charakteristikám monitorovaných systémov.

**Detekcia anomálií pomocou robustnej analýzy hlavných komponentov (Robust Principal Component Analysis - RPCA):** Netflix využíva výkonný mechanizmus Netflix RAD na zisťovanie anomálií v rôznych prípadoch použitia, ako sú anomálie platieb a anomálie pri registrácii. Využitím robustnej analýzy hlavných komponentov (RPCA) dokáže Netflix v rámci svojich údajov identifikovať reprezentácie nízkej úrovne, náhodný šum a odľahlé hodnoty. Tento prístup zabezpečuje efektívne zisťovanie anomálií údajov, čo prispieva k celkovému zlepšeniu kvality údajov. V prípade záujmu o detail sa dá viac informácií nájsť priamo v kóde implementácie RPCA<sup>44</sup>.

## 6.2 Algoritmy pre pokročilé čistenie dát - Talend a OpenMetadata

Tradičné metódy spracovania dát sa často spoliehali na manuálne kódovanie, zložité SQL dotazy a štandardné ETL (Extract, Transform, Load) procesy. Hoci tieto metódy plnili svoj účel, čelili obmedzeniam pri práci s veľkými objemami údajov, zložitými štruktúrami údajov a rýchlo sa meniacimi požiadavkami organizácií. Okrem toho,

<sup>44</sup> [Surus/resources/examples/pig/rad.pig at master · Netflix/Surus · GitHub](https://github.com/Netflix/Surus/tree/master/resources/examples/pig/rad)



nedostatok automatizácie spôsobil, že to bolo časovo náročné a náchylné na chyby, čo bránilo celkovej integrácii údajov a procesu analýzy.

Na druhej strane pokročilé metódy AI/ML ponúkajú zmenu paradigmy v spracovaní údajov. Umožňujú organizáciám využiť silu automatizácie, prediktívnej analýzy a rozpoznávania vzorov na získanie zmysluplných poznatkov z rozsiahlych a rôznorodých súborov údajov. Tieto techniky dokážu zvládnuť zložité dátové vzťahy, odhaliť anomálie a optimalizovať procesy transformácie dát, čo vedie k efektívnejšiemu a presnejšiemu rozhodovaniu.

Nasledovný text sumarizuje tradičné a pokročilé metódy v nástroji Talend, ktorý sa využíva v projekte DI. Popis nástroja OpenMetadata, plánovaného na nasadenie v projekte CIP/CSRU sa popis sa nachádza vo výstupe 1.2.2. Niektoré z opísaných nástrojov, postupov alebo metód však zatiaľ nie sú k dispozícii pre používateľov v rámci projektov DI a CIP/CSRU. Zároveň poznamenávame, že tento text v žiadnom prípade nenahrádza relevantnú používateľskú alebo technickú dokumentáciu, na ktorú týmto čitateľa odkazujeme.

## 6.2.1 Tradičné algoritmické metódy v Talende

Algoritmické metódy v Talende sa používajú na čistenie, štandardizáciu a obohatenie údajov, čím sa zabezpečuje, že spĺňajú požadované štandardy kvality. Niektoré z bežných metód v Talende zahŕňajú:

- **Profilovanie údajov:** Talend poskytuje funkciu profilovania údajov, ktorá zahŕňa analýzu obsahu a štruktúry údajov s cieľom identifikovať potenciálne problémy s kvalitou údajov. Automaticky zisťuje vzory, distribúcie údajov a typy údajov, čím pomáha používateľom pochopiť kvalitu ich údajov.
- **Detekcia duplicitných záznamov:** Talend má algoritmy, ktoré dokážu identifikovať a spravovať duplicitné záznamy v súboroch údajov. Tieto algoritmy pomáhajú zabezpečiť konzistentnosť údajov a vyhnúť sa nadbytočným informáciám.
- **Čistenie údajov:** Talend ponúka rôzne funkcie čistenia údajov, ktoré môžu opraviť, odstrániť alebo nahradiť nepresné alebo neúplné údaje. To zahŕňa techniky, ako je overenie adresy, obohatenie údajov a štandardizácia údajov.
- **Data Matching:** Algoritmy zhody dát v Talende umožňujú používateľom identifikovať a prepojiť záznamy, ktoré odkazujú na rovnakú entitu v rôznych súboroch údajov. Je to užitočné pri vytváraní jednotného zobrazenia údajov z viacerých zdrojov.
- **Pravidlá kvality dát:** Talend umožňuje používateľom definovať a aplikovať vlastné pravidlá kvality dát. Tieto pravidlá pomáhajú pri overovaní údajov podľa špecifických kritérií a zabezpečujú, aby údaje vyhovovali požadovaným štandardom.
- **Profilovanie kvality údajov:** Talend umožňuje používateľom vytvárať úlohy profilovania kvality údajov na automatizáciu procesu analýzy kvality údajov. Tieto úlohy možno naplánovať tak, aby sa pravidelne spúšťali, aby sa v priebehu času monitorovala kvalita údajov.
- **Overenie údajov:** Talend poskytuje celý rad overovacích funkcií na overenie integrity a presnosti údajov, ako je overenie typu údajov, kontroly rozsahu a kontroly konzistencie.

- **Štandardizácia dát:** Talend podporuje algoritmy štandardizácie dát, ktoré zaisťujú, že dáta sú formátované jednotne a konzistentne v rámci dátových množín.
- **Obohacovanie údajov:** Talend môže obohatiť údaje integráciou externých zdrojov údajov a poskytnúť tak dodatočný kontext alebo podrobnosti o existujúcich údajoch.
- **Data Monitoring and Reporting:** Talend ponúka možnosti monitorovania kvality dát, ktoré používateľom umožňujú sledovať metriky kvality dát a generovať správy na vyhodnotenie celkového stavu ich dát.

## 6.2.2 Pokročilé metódy a algoritmy strojového učenia v Talende

Talend ako robustná platforma ponúka širokú škálu nástrojov, ktoré používateľom umožňujú využívať možnosti AI/ML vo svojich pracovných postupoch integrácie údajov. Algoritmy strojového učenia sú zoskupené do štyroch oblastí podľa toho, ako fungujú, pričom každá obsahuje rôzne komponenty ML pripravené na použitie<sup>45</sup>:

### 1. Klasifikačné algoritmy (binárna a viaczložková klasifikácia)

Klasifikácia v strojovom učení je technika dolovania údajov, ktorá sa používa na nájdenie vzorov vo veľkých súboroch údajov. Na identifikáciu, do ktorej množiny kategórií (subpopulácií) pozorovanie patrí sa používa množinu trénovacích údajov obsahujúcich pozorovania (inštancie), ktorých príslušnosť ku kategórii je známa.

Príklady použitia klasifikačných algoritmov zahŕňajú detekciu spamu, kategorizáciu obrázkov a ťažbu textu na zisťovanie nálady zákazníkov. Cieľom je predpovedať podskupinu triedy (alebo označenie – label) podľa známeho príkladu.

Komponenty klasifikácie strojového učenia v Talende zahŕňajú funkcie tClassify, tClassifySVM, tDecisionTreeModel, tGradientBoostedTreeModel, tLogicRegressionModel, tNaiveBayesModel, tPredict, tRandomForestModel a tSVMModel.

### 2. Algoritmy zhlukovania (klastrovania)

Klastrová analýza (zhlukovanie) je primárnou úlohou prieskumného získavania údajov a bežnou technikou používanou pri štatistickej analýze údajov. Napríklad zoskupovanie typu K-means je typom učenia bez dozoru. Je to jeden z najjednoduchších algoritmov učenia bez dozoru, ktorý sa používa na riešenie problému klasifikácie danej množiny údajov prostredníctvom určitého počtu zhlukov. Príklady použitia pre K-means zahŕňajú cenovú segmentáciu, určenie lojality zákazníkov a odhalenie podvodov.

Komponenty klastrovania strojového učenia Talend zahŕňajú tKMeansModel, tPredict a tPredictCluster.

### 3. Odporúčacie algoritmy

Nazýva sa to aj systém odporúčaní a je to podtrieda filtrovania informácií, ktorá sa snaží predpovedať hodnotenie alebo preferencie, ktoré by používateľ dal položke. Kolaboratívne filtrovanie je jedným typom algoritmu odporúčaní. Kolaboratívne filtrovanie môže byť založené na používateľoch alebo na položkách. Cieľom oboch

<sup>45</sup> [Machine Learning Tools: Talend's Comprehensive Platform | Talend](#)

prístupov je automaticky predpovedať používateľov alebo položky (t. j. filtrovať) na základe preferencií mnohých používateľov alebo položiek (t. j. spolupráca).

Dva typy komponentov odporúčaní strojového učenia Talend sú tALSModel a tRecommend. Algoritmy odporúčacieho systému je možné skombinovať s technikami hlbokého učenia, aby ste mohli predpovedať obrovské objemy veľkých dát.

Komponenty odporúčaní strojového učenia Talend zahŕňajú tALSModel a tRecommend.

#### **4. Regresné algoritmy**

Regresné testovanie je štatistický proces na odhadovanie vzťahu medzi premennými. Zameriava sa na vzťah medzi závislou premennou a jednou alebo viacerými nezávislými premennými alebo „prediktormi“.

Komponenty regresie strojového učenia Talend zahŕňajú tModelEncoder, tLinearRegressionModel a tPredict.



## 7 Naplánovanie zlepšenia dátovej kvality

V tejto kapitole sa nachádza časový harmonogram pre implementáciu jednotlivých odporúčaní na zvýšenie dátovej kvality. Je rozdelený po jednotlivých odporúčaníach, pri ktorých je indikovaná organizácia zodpovedná za implementáciu a odporúčaný termín, do ktorého je potrebné implementáciu vykonať. Každé odporúčanie je odvodené od zistenia, na ktoré je odkazované cez jednotlivé kapitoly priamo pri odporúčaní.

**Tabuľka 80 Odporúčania na zlepšenie dátovej kvality**

ID	Odporúčanie	Zodpovednosť za implementáciu	Odporúčaný termín
1.	Zvýšiť počet OVM integrovaných na CMÚ, najmä všetky tie, ktoré spravujú referenčné registre (kapitola 3.2)	Dátová kancelária	Q4 2024
2.	Zamerať sa pri navrhovaní nových projektov na osvedčenú prax v ostatných štátoch EÚ (kapitola 3.2)	Dátová kancelária	Q4 2023
3.	Zaviesť vynútiteľnú povinnosť pravidelne merať DK a jednotným spôsobom zverejňovať výsledky meraní (kapitola 3.2)	Dátová kancelária	Q2 2024
4.	Navrhnuť štruktúru a obsah jednotného reportu dátovej kvality pre zverejňovanie	Dátová kancelária	Q4 2023
5.	Zverejňovať pravidelné reporty dátovej kvality jednotným spôsobom (kapitola 3.2)*	OVM	Q2 2024
6.	Jasne dodefinovať roly a zodpovednosti v rámci riadenia DK na úrovni MIRRI a na úrovni OVM (kapitola 3.2)	Dátová kancelária	Q1 2024
7.	Zjednotiť pravidlá riadenia dátovej kvality naprieč OVM (kapitola 3.2)*	OVM a Dátová kancelária	Q1 2024
8.	Riadenie zmien zaradiť medzi prioritné úlohy a zodpovednosti Dátovej kancelárie (kapitola 3.2)	Dátová kancelária	Q4 2023
9.	Štatistický úrad by mal v spolupráci s Dátovou kanceláriou (MIRRI) definovať spôsob a postup merania dátovej kvality tak, aby výstup z merania dátovej kvality bol dostatočne informatívny pre konzumentov dát RPO a	ŠÚ a Dátová kancelária	Q4 2023

ID	Odporúčanie	Zodpovednosť za implementáciu	Odporúčany termín
	zadefinované biznis pravidlá boli zatriedené do parametrov dátovej kvality. (kapitola 3.3)		
10.	Je potrebné zabezpečiť kontinuitu vykonávania meraní dátovej kvality a riadenia dátovej kvality RPO bez ohľadu na personálne zmeny Štatistického úradu. (kapitola 3.3)*	ŠÚ	Q4 2023
11.	Je potrebné formulovať kampane na zvyšovanie dátovej kvality v zdrojových registroch RPO a realizovať ich v spolupráci so zdrojovými registrami. (kapitola 3.3)*	ŠÚ a Zdrojové registre RPO, Dátová kancelária	Q2 2024
12.	ŠÚ SR by mal v spolupráci s Dátovou kanceláriou vytvoriť metodiku pre zdrojové registre RPO na poskytovanie údajov vo vysokej kvalite. (kapitola 3.3)*	ŠÚ a Dátová kancelária	Q1 2024
11.	ŠÚ SR by mal v spolupráci s Dátovou kanceláriou preveriť možnosti zavedenia prahových limitov na parametre dátovej kvality v RPO a ich zadefinovanie prostredníctvom uznesenia, vyhlášky príp. iného relevantného záväzného aktu. (kapitola 3.3)	ŠÚ a Dátová kancelária	Q3 2024
12.	ŠÚ SR by mal pravidelne vykonávať merania dátovej kvality RPO a zverejnením výsledkov meraní sprístupniť informácie o dátovej kvalite RPO konzumentom jeho údajov. (kapitola 3.3)	ŠÚ a Dátová kancelária	Q4 2023
13.	MV SR by malo vykonávať pravidelné merania dátovej kvality a v súlade s usmerneniami Dátovej kancelárie zverejňovať výsledky, aby boli informácie o kvalite údajov v RA dostupné pre jeho konzumentov. (kapitola 3.4)*	MV SR a Dátová kancelária	Q4 2023
14.	Biznis pravidlá pre meranie a štruktúra zverejňovaných reportov by mali byť založené na požiadavkách konzumentov a schválené Dátovou kanceláriou. (kapitola 3.4, RA)*	MV SR a Dátová kancelária	Q4 2023
15.	Zvyšovanie DK by malo byť riadeným procesom založeným na definovaných cieľoch (prahových hodnotách parametrov DK) a plán na dosiahnutie stanovených cieľov by mal zahŕňať úlohy nevyhnutné pre dosiahnutie cieľov. Tento proces by mal byť nepretržitý. (kapitola 3.4, RA)*	MV SR	Q1 2024

ID	Odporúčanie	Zodpovednosť za implementáciu	Odporúčaná termín
16.	RNZ by mal byť priamo integrovaný na referenčné dáta RPO, aby mohli byť pri vytváraní záznamu údaje o pokutovanom subjekte stotožňované voči aktuálnym údajom RPO. (kapitoly 3.5)	MPSVR SR, Dátová kancelária	Q2 2024
17.	Všetky referenčné registre by mali byť priamo integrované na referenčné dáta RPO*	OVM a Dátová kancelária	Q4 2024
18.	Všetky nové záznamy by mali byť vytvárané v centrálnom systéme, nie v lokálnych kópiách databázy. (kapitoly 3.5, RNZ)	MPSVR SR	Q2 2024
19.	Pri implementácii nového IS je potrebné odstrániť obmedzenia, ktoré prispievajú k vytváraniu duplicitných záznamov. (kapitola 3.5, RNZ)	MPSVR SR, Dátová kancelária	Q2 2024
20.	Do registra je potrebné zahrnúť aj zoznam pokutovaných fyzických osôb a právnických so sídlom mimo územia Slovenskej republiky. (kapitoly 3.5, RNZ)	MPSVR SR, Dátová kancelária	Q2 2024
21.	Dosiahnutie referenčnej integrity vyžaduje zavedenie jednoznačného referenčného identifikátora. (kapitoly 3.5, RNZ)*	MPSVR SR, Dátová kancelária	Q2 2024
22.	Zvýšenie strojovej spracovateľnosti vyžaduje sprístupnenie údajov v otvorenom formáte. (kapitoly 3.5, RNZ)*	MPSVR SR, Dátová kancelária	Q2 2024
23.	OVM musia zabezpečiť Organizačné riadenie dátovej kvality v rámci ich organizačného prostredia (kapitola 4.3.1)*	OVM	Q4 2024
24.	Koordinácia v rámci organizačného zabezpečenia riadenia dátovej kvality v rámci OVM (kapitola 4.3.1.1 a kapitola 4.3.2.5)*	OVM a Dátová kancelária	Q4 2024
25.	Zabezpečenie procesného riadenia dátovej kvality v prostredí OVM (kapitola 4.3.2)*	OVM	Q4 2025
26.	Technické zabezpečenie riadenia dátovej kvality a dátové štandardy v prostredí OVM. (kapitola 4.3.3)*	OVM	Q4 2025

ID	Odporúčanie	Zodpovednosť za implementáciu	Odporúčany termín
27.	Adoptovanie postupov a požiadaviek príručky kvality pre OVM (kapitola 4)*	OVM a Dátová kancelária	Q4 2025

\*Platí takisto pre všetky referenčné registre

## 8 Dohľad nad realizáciou odporúčaní pre zlepšenie dátovej kvality

Dohľad nad realizáciou odporúčaní pre zlepšenie dátovej kvality adresovaných OVM a naformulovaných už skôr v tomto dokumente príp. v iných relevantných výstupoch zameraných na dátovú kvalitu publikovaných dátovou kanceláriou MIRRI (napr. odporúčania v dokumente č. 1.1.1) je v kompetencii koordinátora riadenia dátovej kvality, t. j. v kompetencii MIRRI. Účelom takéhoto dohľadu je predovšetkým získanie informácií o plnení/nepĺnení úloh súvisiacich s odporúčaniami a poskytnutie podpory OVM smerujúcej k naplneniu podstaty odporúčaní a implementácii vhodných riešení.

Predpokladom pre zlepšenie dátovej kvality a implementáciu s ňou súvisiacich opatrení, aby OVM zodpovedné za realizáciu opatrení boli informované o týchto opatreniach, o dôvode a kontexte týchto opatrení ako aj o možnostiach ich naplnenia v praxi a aby sa s nimi stotožnili a podporovali ich realizáciu. Preto nevyhnutnou súčasťou dohľadu realizovaného zo strany MIRRI je v prvom kroku odkomunikovanie odporúčaní s dotknutými OVM. Za týmto účelom MIRRI ako súčasť svojej koordinačnej činnosti zabezpečuje informovanosť OVM, ktorých sa odporúčania týkajú, minimálne:

- Zverejnením odporúčaní a s tým súvisiacej dokumentácie prostredníctvom verejne dostupného priestoru dedikovaného pre dátovú kvalitu (napríklad dohodnutá stránka na webovom sídle dátovej kancelárie MIRRI).
- Prezentovaním a vysvetlením odporúčaní na viacerých úrovniach hierarchie riadenia prostredníctvom vhodného komunikačného kanála (stretnutie MIRRI a vlastníkov dátovej kvality u OVM, stretnutia dátových kurátorov, stretnutia relevantných pracovných skupín a pod.).  
Cieľom takejto prezentácie nie je iba informovať OVM o aktuálnych odporúčaníach dátovej kvality, ale aj vysvetlenie dôvodov pre formulované odporúčania vrátane odkomunikovania očakávaní koordinátora vo vzťahu k napĺňaniu odporúčaní jednotlivými OVM a získanie súhlasu a podpory pre navrhnuté odporúčania a komunikované očakávania. Na to, aby OVM mohli efektívne a úspešne pristupovať k napĺňaniu týchto očakávaní, musia pochopiť podstatu odporúčaní a pozitívne dôsledky v prípade ich implementácie do prostredia svojej organizácie. Preto povinnou súčasťou prezentácie odporúčaní je aj uvedenie predpokladaného budúceho stavu a jeho výhod oproti aktuálnemu stavu, najmä za účelom spomenutého získania súhlasu a podpory zo strany vedenia OVM.
- Vytvorením centrálného registra úloh dátovej kvality (nástroj koordinácie zabezpečujúci dohľad nad riadením dátovej kvality OVM, bližšie pozri kapitolu 4.4.1.1 *Hodnotenie stavu úloh dátovej kvality*), do ktorého sú prenesené odprezentované odporúčania, resp. úlohy vyplývajúce z týchto odporúčaní pre OVM.

Za plnenie úloh dátovej kvality v centrálnom registri úloh dátovej kvality sú zodpovední vlastníci dátovej kvality príslušného OVM. Pri vytváraní úlohy dátovej kvality postupuje OVM tak, aby riešenie úlohy smerovalo k napĺňaniu odkomunikovaných (prezentovaných) očakávaní koordinátora v súvislosti s príslušnými odporúčaniami. Vytvorenie a pridelenie úlohy v centrálnom registri úloh je preto výsledkom vzájomnej komunikácie medzi koordinátorom dátovej kvality (MIRRI) a príslušným OVM.

Samotný dohľad nad realizáciou odporúčaní pre zlepšenie dátovej kvality vykonáva MIRRI prostredníctvom:

- hodnotenia stavu úloh dátovej kvality (pozri kapitolu 4.4.1.1 Hodnotenie stavu úloh dátovej kvality) vykonávaného pravidelne, minimálne jedenkrát ročne,
- výkonom auditov u OVM podľa plánu auditov (pozri 4.4.1.2 Audity),
- aktívnou participáciou na plnení úloh spolu s OVM minimálne prostredníctvom poskytovania konzultácií a usmernení OVM pri plnení ich úloh vrátane dohľadu pri zavádzaní procesu riadenia kvality do prostredia OVM.

Na účel dohľadu je zo strany MIRRI potrebné zaviesť povinnosť OVM vytvoriť a udržiavať vo svojej organizácii riadiaci akt zameraný na implementáciu procesov riadenia dátovej kvality (v súlade s navrhovaným rámcom v podobe príručky popísanej v kapitole 4 *Zovšeobecnenie odporúčaní - Príručka kvality pre OVM* tohto dokumentu), ako integrálnej súčasť správy a riadenia dát. Dohľad MIRRI v tejto oblasti je potom vykonávaný prostredníctvom posudzovania, pripomienkovania a prípadne i schvaľovania takýchto riadiacich aktov ešte pred ich samotným zavedením v príslušnom OVM. Prostredníctvom schvaľovania môže MIRRI zabezpečiť, aby dátová kvalita v OVM bola riadená v súlade s očakávaniami koordinátora (MIRRI) a jednotne naprieč všetkými OVM.

Primárnym cieľom takto nastaveného dohľadu nie je „zistiť chyby a trestať“, ale podporovať a motivovať OVM k zvyšovaniu dátovej kvality. Okrem vlastného výkonu dohľadu (prostredníctvom hodnotenia úloh, výkonu auditov a schvaľovania interných riadiacich aktov riadenia dátovej kvality) a poskytovania podpory pri riešení úloh, je preto kľúčovým aj motivovanie OVM prostredníctvom podpory pri realizácii dopytových projektov súvisiacich s dátovou kvalitou.

V súvislosti s dohľadom treba zdôrazniť jeho potenciál vo vzťahu k motivácii OVM plniť odporúčania - informácie získané koordinátorom pri výkone dohľadu je žiaduce ďalej využiť spôsobom, ktorý bude zvyšovať ochotu OVM aktívne a zodpovedne pristupovať k zlepšovaniu dátovej kvality. Za týmto účelom na základe zistení z dohľadu MIRRI:

- Aktívne zbiera a verejne prezentuje úspešné prípady implementácie odporúčaní z reálnej praxe jednotlivých OVM (prostredníctvom verejne dostupného priestoru dedikovaného pre dátovú kvalitu, napríklad v rámci datalab.digital),
- Vytvorí platformu aj na osobné zdieľanie skúseností a aplikáciu overených postupov z úspešných prípadov implementácie u niektorého OVM. Zvyšuje to jednak pozitívne vnímanie riešenia konkrétnych situácií v praxi, keď takéto riešenia sa na základe existujúcich úspešných príkladov javia ako dosiahnuteľnejšie a tiež zefektívňuje realizáciu v podobných situáciách u iných OVM
- Pri plnení úloh dátovej kvality by OVM mali postupovať podľa overenej praxe a tam, kde to nie je nevyhnutné, by mali využiť riešenia už aplikované inými OVM (prípadne sa nimi inšpirovať alebo si ich upraviť vzhľadom na špecifiká svojej organizácie). V tejto súvislosti by bolo užitočné uvažovať aj o vytvorení mechanizmu krátkodobých stáží, ktorých cieľom je získanie konkrétnych praktických skúseností pracovníkov (napr. dátových kurátorov) u takého OVM, ktoré je napred v oblasti riadenia a zvyšovania dátovej kvality. Podobne by takéto stáže mohli byť organizované v relevantných prípadoch aj na MIRRI.

Úlohou MIRRI ako koordinátora dátovej kvality a realizátora s tým súvisiaceho dohľadu je preto nielen výkon samotného dohľadu, ale takisto vytvorenie prostredia

a predpokladov na efektívnu implementáciu opatrení využitím zistení z dohľadu tak, aby ďalej podporovali a motivovali OVM k zlepšovaniu dátovej kvality. Táto úloha však prekračuje rámec samotnej dátovej kancelárie a musí byť vlastnená na vyššej až najvyššej úrovni riadenia MIRRI, nakoľko vyžaduje pravidelnú komunikáciu a získavanie súhlasu a podpory na zodpovedajúcich hierarchických úrovniach jednotlivých OVM.

## 9 Register tabuliek a obrázkov

Tabuľka 1 Biznis pravidlá pre meranie presnosti RPO .....	4
Tabuľka 2 Biznis pravidlá pre meranie správnosti RPO .....	5
Tabuľka 3 Biznis pravidlá pre meranie kompletnosti RPO .....	7
Tabuľka 4 Biznis pravidlá pre meranie konzistentnosti RPO .....	12
Tabuľka 5 Biznis pravidlá pre meranie referenčnej integrity RPO .....	15
Tabuľka 6 Zoznam meraných biznis pravidiel nad RA .....	18
Tabuľka 7 Biznis pravidlá s overenou presnosťou 100 % .....	27
Tabuľka 8 Výsledok merania pravidla BP_MVSR_RA_004 .....	29
Tabuľka 9 Výsledok merania pravidla BP_MVSR_RA_020 .....	29
Tabuľka 10 Výsledok merania pravidla BP_MVSR_RA_038 .....	29
Tabuľka 11 Výsledok merania pravidla BP_MVSR_RA_057 .....	30
Tabuľka 12 Výsledok merania pravidla BP_MVSR_RA_061 .....	30
Tabuľka 13 Výsledok merania pravidla BP_MVSR_RA_075 .....	30
Tabuľka 14 Výsledok merania pravidla BP_MVSR_RA_090 .....	31
Tabuľka 15 Výsledok merania pravidla BP_MVSR_RA_092 .....	31
Tabuľka 16 Výsledok merania pravidla BP_MVSR_RA_094 .....	31
Tabuľka 17 Výsledok merania pravidla BP_MVSR_RA_095 .....	32
Tabuľka 18 Výsledok merania pravidla BP_MVSR_RA_098 .....	32
Tabuľka 19 Výsledok merania pravidla BP_MVSR_RA_111 .....	32
Tabuľka 20 Výsledok merania pravidla BP_MVSR_RA_116 .....	33
Tabuľka 21 Výsledok merania pravidla BP_MVSR_RA_121 .....	33
Tabuľka 22 Výsledok merania pravidla BP_MVSR_RA_125 .....	33
Tabuľka 23 Výsledok merania pravidla BP_MVSR_RA_126 .....	34
Tabuľka 24 Výsledok merania pravidla BP_MVSR_RA_132 .....	34
Tabuľka 25 Výsledok merania pravidla BP_MVSR_RA_134 .....	34
Tabuľka 26 Výsledok merania pravidla BP_MVSR_RA_138 .....	35
Tabuľka 27 Výsledok merania pravidla BP_MVSR_RA_139 .....	35
Tabuľka 28 Výsledok merania pravidla BP_MVSR_RA_140 .....	35
Tabuľka 29 Výsledok merania pravidla BP_MVSR_RA_141 .....	36
Tabuľka 30 Výsledok merania pravidla BP_MVSR_RA_142 .....	36
Tabuľka 31 Výsledok merania pravidla BP_MVSR_RA_145 .....	37
Tabuľka 32 Biznis pravidlá s overenou správnosťou 100 % .....	37
Tabuľka 33 Výsledok merania pravidla BP_MVSR_RA_012 .....	39
Tabuľka 34 Výsledok merania pravidla BP_MVSR_RA_028 .....	39
Tabuľka 35 Výsledok merania pravidla BP_MVSR_RA_065 .....	39
Tabuľka 36 Výsledok merania pravidla BP_MVSR_RA_077 .....	40
Tabuľka 37 Výsledok merania pravidla BP_MVSR_RA_078 .....	40
Tabuľka 38 Biznis pravidlá s overenou kompletnosťou 100 % .....	40
Tabuľka 39 Výsledok merania pravidla BP_MVSR_RA_033 .....	43
Tabuľka 40 Výsledok merania pravidla BP_MVSR_RA_093 .....	44
Tabuľka 41 Výsledok merania pravidla BP_MVSR_RA_100 .....	44
Tabuľka 42 Výsledok merania pravidla BP_MVSR_RA_101 .....	45
Tabuľka 43 Výsledok merania pravidla BP_MVSR_RA_103 .....	45
Tabuľka 44 Výsledok merania pravidla BP_MVSR_RA_104 .....	45
Tabuľka 45 Výsledok merania pravidla BP_MVSR_RA_113 .....	46
Tabuľka 46 Výsledok merania pravidla BP_MVSR_RA_120 .....	46
Tabuľka 47 Výsledok merania pravidla BP_MVSR_RA_127 .....	47
Tabuľka 48 Výsledok merania pravidla BP_MVSR_RA_128 .....	47
Tabuľka 49 Výsledok merania pravidla BP_MVSR_RA_119 .....	47
Tabuľka 50 Výsledok merania pravidla BP_MVSR_RA_052 .....	48
Tabuľka 51 Výsledok merania pravidla BP_MVSR_RA_071 .....	48
Tabuľka 52 Výsledok merania pravidla BP_MVSR_RA_085 .....	49



Tabuľka 53 Výsledok merania pravidla BP_MVSR_RA_086.....	49
Tabuľka 54 Výsledok merania pravidla BP_MVSR_RA_106.....	50
Tabuľka 55 Výsledok merania pravidla BP_MVSR_RA_107.....	50
Tabuľka 56 Výsledok merania pravidla BP_MVSR_RA_123.....	50
Tabuľka 57 Výsledok merania pravidla BP_MVSR_RA_124.....	51
Tabuľka 58 Výsledok merania pravidla BP_MVSR_RA_118.....	51
Tabuľka 59 Biznis pravidlá so zistenou presnosťou nižšou než 100 % .....	52
Tabuľka 60 Biznis pravidlá so zistenou správnosťou nižšou než 100 % .....	54
Tabuľka 61 Biznis pravidlá so zistenou kompletnosťou nižšou než 100 %.....	55
Tabuľka 62 Biznis pravidlá so zistenou konzistentnosťou nižšou než 100 % .....	56
Tabuľka 63 Zoznam meraných biznis pravidiel nad RNZ .....	59
Tabuľka 64 Výsledok merania pravidla BP_NIP_RNZ_002.....	61
Tabuľka 65 Výsledok merania pravidla BP_NIP_RNZ_005.....	61
Tabuľka 66 Výsledok merania pravidla BP_NIP_RNZ_010.....	62
Tabuľka 67 Výsledok merania pravidla BP_NIP_RNZ_013.....	62
Tabuľka 68 Výsledok merania pravidla BP_NIP_RNZ_014.....	62
Tabuľka 69 Výsledok merania pravidla BP_NIP_RNZ_004.....	63
Tabuľka 70 Výsledok merania pravidla BP_NIP_RNZ_009.....	63
Tabuľka 71 Výsledok merania pravidla BP_NIP_RNZ_016.....	64
Tabuľka 72 Výsledky merania kompletnosti povinných atribútov .....	64
Tabuľka 73 Výsledok merania pravidla BP_NIP_RNZ_015.....	65
Tabuľka 74 Výsledok merania dátovej kvality v RNZ.....	65
Tabuľka 75 Celkový stav iniciatív DK podľa oblastí .....	69
Tabuľka 76 Centrálna úroveň: Vyhodnotenie realizácie opatrení .....	70
Tabuľka 77 Rezortná úroveň: Vyhodnotenie realizácie opatrení .....	71
Tabuľka 78 Všeobecné odporúčania pre zlepšenie dátovej kvality .....	73
Tabuľka 79 Posúdenie nastavenia postupov správy dát .....	100
Tabuľka 80 Odporúčania na zlepšenie dátovej kvality.....	114
Obrázok 1 Vízia digitálnej transformácie: predpoklady a oblasti .....	80
Obrázok 2 Vymedzenie dátovej kvality vo verejnej správe .....	83
Obrázok 3 Riadenie dátovej kvality v OVM.....	85
Obrázok 4 Surové výsledky prieskumu.....	107

# Error! Reference

## Contact us

**Rudolf Sedmina**  
partner  
[rsedmina@kpmg.sk](mailto:rsedmina@kpmg.sk)

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

[www.kpmg.com](http://www.kpmg.com)

© 2023 Copyright owned by one or more of the KPMG International entities. KPMG International entities provide no services to clients. All rights reserved.

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavour to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation.

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.