



Výstup č. 1.1.1

Štandardizácia 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	25.05.2023	
2.0	Mário Finka	15.06.2023	Zpracovanie pripomienok

Revízia

	Recenzent	Dátum revízie
1	Andrej Fukas	09.06.2023
2	Viktória Šunderlíková	09.06.2023
3		
4		
5		

Schválenie

	Schvaľovateľ	Podpis	Dátum schválenia
1			
2			
3			
4			
5			

Slovník

CES	Centrálny ekonomický systém
CMÚ	Centrálny model údajov
dataset	Množina údajov, množina dát, niekedy aj register
Dátový kurátor	Odborný zamestnanec zodpovedný za správu a udržiavanie kvality, presnosti a užitočnosti dát počas ich životného cyklu
DK	Dátová kvalita
ID-SK	Metodické usmernenie a jednotný dizajn manuál elektronických služieb a webových sídiel SR
ISVS	Informačný/é systém/y verejnej správy
KPI	Kľúčové ukazovatele výkonnosti (Key Performance Indicators)
Metodika merania dátovej kvality (Metodika merania DK, Metodika)	Metodika merania dátovej kvality vo verejnej správe, ktorá bola vydaná v roku 2019 v rámci národného projektu „Zlepšenie využívania údajov vo verejnej správe“ (Zmluva o dielo č. 321/2018).
MF SR	Ministerstvo financií SR
MIRRI	Ministerstvo investícií, regionálneho rozvoja a informatizácie SR
MŠVVaŠ SR	Ministerstvo školstva, vedy, výskumu a športu SR
MV SR	Ministerstvo vnútra SR
NSZÚ	Národný systém zdieľania údajov
OVM	Orgány verejnej moci
RA	Register adries
RIS	Rezortný informačný systém
RPO	Register právnických osôb
ŠÚ SR	Štatistický úrad SR
Zákon o údajoch	Pripravovaný Zákon o správe, zabezpečení prístupovania a používania vybraných kategórií údajov evidovaných v informačných systémoch verejnej správy (zákon o údajoch).

Obsah

1	Úvod a zhrnutie	1
2	Posúdenie súčasného stavu	2
2.1	Súčasná a chystaná legislatíva	2
2.1.1	Dátová legislatíva EÚ	2
2.1.2	Súčasná dátová legislatíva SR	2
2.1.3	Pripravovaný zákon o údajoch	3
2.2	Stav v riadení dátovej kvality	6
2.3	Podrobnejšie vyhodnotenie odporúčaní	11
2.3.1	Dátové štandardy	11
2.3.2	Organizácia	1415
2.3.3	Proces	17
2.3.4	Technológie	1920
2.4	Stav v oblasti merania kvality dát	22
2.4.1	Doterajšie aktivity merania DK	22
2.4.2	Požiadavky a politiky riadenia dátovej kvality	23
2.4.3	Predchádzajúce odporúčania a ich relevantnosť	24
2.5	Centrálne projekty riadenia dát s väzbou na dátovú kvalitu	27
2.5.1	Dátová integrácia (DI)	27
2.5.2	Centrálna integračná platforma (CIP)	31
2.5.3	Konsolidovaná analytická vrstva (KAV)	34
2.5.4	Životné situácie (ŽS)	3435
3	Návrh úprav a aktualizácia štandardu pre meranie dátovej kvality	36
3.1	Návrh na úpravu odporúčaní	36
3.1.1	Dátové štandardy	36
3.1.2	Organizácia	38
3.1.3	Proces	40
3.1.4	Technológie	42
3.2	Nové odporúčania	44
3.3	Návrhy na zlepšenie merania kvality dát	46
4	Definícia a spôsob merania pre kľúčové výkonnostné indikátory	48
4.1	Príprava na meranie kvality	49
4.1.1	Definovanie objektu	49
4.1.2	Získanie dát	50
4.1.3	Výber parametrov	50
4.1.4	Výber ukazovateľov (kľúčové výkonnostné ukazovatele, KPI)	52
4.2	Vykonanie merania	56
4.3	Výstup z merania	57
5	Popis alternatívnych nástrojov pre analýzu kvality dát	58
5.1	O prístupe k výberu nástroja pre dátovú kvalitu	58

5.2	Rozšírenie výberu odporúčaných nástrojov	63
6	Definícia systému pravidelnej aktualizácie a reportovania	64
6.1	Zavedené postupy hlásenia a aktualizácií dátovej kvality a podmienok hodnotenia	66
6.1.1	Štatistický úrad - RPO	66
6.1.2	Otvorené údaje	68
6.1.3	Ministerstvo vnútra – Register adries	68
6.1.4	Ministerstvo školstva, vedy, výskumu a športu – Rezortný informačný systém	69
6.2	Model reportovania	70
6.2.1	Návrh obsahu reportov dátovej kvality	72
6.2.2	Návrh prístupu k aktualizácii a reportovaniu	73
7	Definícia požiadaviek na ISVS pre automatizované meranie kvality	74
7.1	Všeobecné požiadavky automatizácie merania DK	74
7.2	Požiadavky na automatizáciu merania DK v kontexte ISVS	75
7.3	Odporúčania pre automatizáciu merania DK v ISVS	77
8	Registre	78
8.1	Register tabuliek	78
8.2	Register obrázkov	78

1 Úvod a zhrnutie

Účelom tohto dokumentu je vytvorenie prehľadu vývoja merania dátovej kvality vo verejnej správe od roku 2019, kedy bola vydaná Metodika merania dátovej kvality vo verejnej správe (Metodika merania DK alebo aj Metodika) a predloženie návrhov na jej aktualizáciu.

V rámci revízie Metodiky bolo realizované preskúmanie súčasného stavu merania DK u orgánov verejnej moci, stav implementácie odporúčaní pre zlepšenie merania DK a stav jednotlivých projektov a iniciatív v oblasti merania DK, ktorým sa zaoberá kapitola 2.

Na tomto základe je v kapitolách 3 a 4 ponúknutý návrh úprav štandardu a spôsobu merania výkonnostných indikátorov (KPI). Ďalej sú popísané alternatívne nástroje pre analýzu kvality dát (kap. 5) a návrh pre systém reportovania (podávanie správ o kvalite) v kapitole 6. Záverom v kapitole 7 sumarizujeme požiadavky na informačné systémy verejnej správy súvisiace s meraním dátovej kvality.

Meranie dátovej kvality je iba jedným z komponentov komplexu riadenia kvality dát, preto sa v niektorých prípadoch odkazujeme na nadväzujúce výstupy 1.2.1 Opatrenia pre zvýšenie dátovej kvality a 1.2.2 Služby dátovej kvality. Preto zároveň predmetom tohto dokumentu nie je alebo je len okrajovo všeobecný stav správy dát a ich kvality a vzhľadom na časové súvislosti tvorby tohto dokumentu ani vyhodnotenie strojového merania kvality vzoriek dát. Vo viacerých prípadoch bolo potrebné rozhodnúť, kam zaradiť niektoré aspekty merania kvality. V prípade, že sme sa rozhodli (najmä kvôli konzistentnosti) ich zaradiť do vyššie uvedených výstupov v kontexte s ostatnými aspektami, je to v texte spomenuté.

2 Posúdenie súčasného stavu

Táto časť popisuje aktivity a stav merania dátovej kvality na Slovensku od roku 2019, pričom treba poznamenať, že vzhľadom na tesnú previazanosť s ostatnými procesmi riadenia DK často opisujeme aj tie. Ďalej kapitola pojednáva aj o aktuálnom stave dátovej legislatívy a prebiehajúcich projektoch.

2.1 Súčasná a chystaná legislatíva

2.1.1 Dátová legislatíva EÚ

V roku 2020 Európska komisia predstavila Európsku dátovú stratégiu. Jej pilierom je nariadenie o harmonizovaných pravidlách o spravodlivom prístupe k údajom a ich využívaní, známe aj ako Zákon o údajoch (Data Act), ktorý Európska komisia prijala 23. februára 2022. Data Act dopĺňa ďalšie nariadenie o správe a riadení údajov (European Data Governance Act), ktoré nadobudlo účinnosť 23. júna 2022 a po 15-mesačnom období odkladu bude platiť od septembra 2023. Zatiaľ čo nariadenie o správe a riadení údajov vytvára procesy a štruktúry na uľahčenie zdieľania údajov, zákon o údajoch objasňuje, kto môže z údajov vytvárať hodnotu a za akých podmienok.

2.1.2 Súčasná dátová legislatíva SR

Na Slovensku bola dátová legislatíva čiastočne pokrytá viacerými zákonmi a vyhláškami, ktoré čiastkovo usmerňujú narábanie s dátami vo svojich oblastiach. Medzi najdôležitejšie z nich patria:

- Zákon č. 211/2000 Z. z. o slobodnom prístupe k informáciám a o zmene a doplnení niektorých zákonov (zákon o slobode informácií), ktorý ustanovuje aké údaje musia povinné osoby zverejniť alebo sprístupniť a akým spôsobom.
- Zákon č. 305/2013 Z. z. o elektronickej podobe výkonu pôsobnosti orgánov verejnej moci a o zmene a doplnení niektorých zákonov (zákon o e-Governmente), ktorý niektoré informačné systémy pre výkon pôsobnosti orgánov verejnej moci v elektronickej podobe.
- Zákon č. 177/2018 Z. z. o niektorých opatreniach na znižovanie administratívnej záťaže využívaním informačných systémov verejnej správy (zákon proti byrokracii), ktorý ustanovuje, že orgány verejnej moci sú pri svojej úradnej činnosti povinné a oprávnené získavať a používať údaje evidované v informačných systémoch verejnej správy.
- Zákon č. 95/2019 Z. z. o informačných technológiách vo verejnej správe a o zmene a doplnení niektorých zákonov (zákon o ITVS), ktorý ustanovuje organizáciu správy informačných technológií verejnej správy a základné požiadavky kladené na informačné technológie verejnej správy.
- Vyhláška úradu podpredsedu vlády SR pre investície a informatizáciu č. 78/2020 Z. z. o štandardoch pre informačné technológie verejnej správy a Vyhláška úradu podpredsedu vlády SR pre investície a informatizáciu č. 85/2020 Z. z. o riadení projektov.

Z hľadiska dátovej kvality je najrelevantnejším usmernením Vyhláška úradu podpredsedu vlády SR pre investície a informatizáciu č. 78/2020, ktorá ustanovuje

štandardy pre opisný jazyk dátových prvkov (§ 12), prenos dátových prvkov (§ 13), výmenu údajov medzi informačnými systémami verejnej správy (§ 30), referencovateľný identifikátor (§ 32), dereferenciáciu (§ 33), kvalitu datasetu poskytovaného orgánom riadenia (§ 38), otvorené údaje (§ 39), ich poskytovanie (§ 40) a základné číselníky (§ 51).

Vyhláška 78/2020 identifikuje 6 úrovní kvality datasetu, pričom najvyššia úroveň kvality (5) vyžaduje splnenie všetkých predošlých podmienok:

- dataset musí byť dostupný vo webovom prostredí (úroveň 1),
- jeho obsah musí byť štruktúrovaný tak, aby bolo možné jeho automatizované spracovanie (úroveň 2),
- jeho obsah musí byť poskytovaný v otvorenom formáte, nezávislom na konkrétnom proprietárnom softvéri (úroveň 3),
- na identifikáciu jeho údajov a ich vzťahov sa používajú referencovateľné identifikátory (úroveň 4) a
- dataset a jeho interné a externé vzťahy sú prepájané prostredníctvom referencovateľných identifikátorov a sú opísané prostredníctvom Centrálného modelu údajov (úroveň 5).

Hoci Vyhláška určuje aspoň podmienky kvality samotného datasetu, žiadna doterajšia legislatíva neustanovuje povinnosti z hľadiska kvality samotných údajov, jej zlepšovania alebo merania.

2.1.3 **Pripravovaný zákon o údajoch**

V súlade s Európskou dátovou stratégiou pristúpilo aj Slovensku ku konsolidácii dátovej legislatívy v pripravovanom Zákone o správe, zabezpečení sprístupňovania a používania vybraných kategórií údajov evidovaných v informačných systémoch verejnej správy (Zákon o údajoch), ktorý pokryje medzery v doterajšej dátovej legislatíve. Návrh zákona o údajoch sa týka najmä:

- postavenia analytických jednotiek a spracúvania tzv. analytických údajov,
- postavenia a práv dátového kurátora,
- definovania MIRRI ako garanta dátovej kvality,
- povinností v oblasti nového konceptu tzv. „mojich údajov“,
- základných číselníkov, referenčných údajov a povinností súvisiacich s referencovaním, otvorených údajov, ako aj súvisiacich povinností dotknutých orgánov verejnej moci (tzv. povinných subjektov), ktoré spravujú údaje vo svojich informačných systémoch.

Prínos nového zákona o údajoch v porovnaní s doterajšou legislatívou

Zákon berie do úvahy a legislatívne ustanovuje aj odporúčania z Metodiky. Zavádza pojem dátovej kvality a jej riadenia a určuje orgánom verejnej moci a právnickým osobám povinnosť plniť si povinnosti týkajúce sa dátovej kvality. Definuje pôsobnosť MIRRI ako správcu národného systému zdieľania údajov v oblasti správy, zabezpečenia sprístupňovania a používania vybraných kategórií údajov evidovaných v informačných systémoch verejnej správy. Poskytuje MIRRI možnosť (aj keď neprikazuje) vykonať minimálne raz ročne kontrolu plnenia povinností a uložiť OVM alebo právnickej osobe nápravné opatrenie. Zákon o údajoch tiež definuje rolu, určuje

postavenie a povinnosti dátového kurátora z hľadiska definovania, zabezpečenia a merania dátovej kvality a ustanovuje MIRRI ako centrálného dátového kurátora, ktorý usmerňuje, metodicky riadi a pravidelne školí všetkých dátových kurátorov.

Pozitívnym prínosom navrhovaného zákona o údajoch je najmä konsolidácia správy a poskytovania údajov, zavedenia pojmu dátovej kvality, určenie zodpovedností zapojených inštitúcií a kompetencií MIRRI z hľadiska metodického riadenia dátovej kvality, kontroly jej dodržiavania a možnosti uloženia nápravných opatrení.

Nedodržanie povinností vyplývajúcich zo zákona nie je však spojené so žiadnymi negatívnymi dopadmi, čo v praxi znamená, že dodržiavanie zákona je v podstate dobrovoľné. Garant zákona (MIRRI) deklaroval, že nepovažuje za efektívne sankcionovať OVM za nedodržiavanie odporúčaní a usmernení Centrálného dátového kurátora, ale považuje za vhodnejšie motivovať ich prostredníctvom podpory pri realizácii dopytových projektov súvisiacich s dátovou kvalitou. Napriek tomu existuje možnosť pozastavenia financovania týchto projektov v prípade, ak OVM nedodržiava svoje povinnosti v rámci dátovej kvality, ktoré vyplývajú zo Zákona o údajoch a ďalšej legislatívy.

Definovanie roly dátového kurátora

Vo všeobecnosti je dátový kurátor odborný zamestnanec, zodpovedný za správu a udržiavanie kvality, presnosti a užitočnosti dát počas ich životného cyklu. V kontexte kvality údajov zohráva kurátor údajov kľúčovú úlohu pri zabezpečovaní toho, aby boli údaje správne usporiadané, zdokumentované a prístupné a aby spĺňali potreby zamýšľaných používateľov. Implementáciou osvedčených postupov pre správu, spracovanie a dokumentáciu údajov môže kurátor údajov pomôcť zabezpečiť, aby boli údaje presné, spoľahlivé a užitočné na účely rozhodovania. Niektoré špecifické úlohy a povinnosti kurátora údajov v oblasti kvality údajov môžu zahŕňať definovanie dátových štandardov (vývoj a presadzovanie dátových štandardov, ako sú konvencie formátovania údajov alebo konvencie pomenovania, aby sa zabezpečila konzistentnosť a presnosť údajov), zhromažďovanie a vyhodnocovanie údajov (identifikácia relevantných zdrojov údajov, zhromažďovanie a usporiadanie údajov a hodnotenie kvality údajov pomocou techník, ako je profilovanie údajov alebo čistenie údajov), udržiavanie integrity údajov (zabezpečenie, že údaje sú presné, aktuálne a úplné, a že sa identifikujú a riešia všetky chyby alebo nezrovnalosti), správu metadát (vývoj a údržba metadát, ako sú dátové slovníky alebo dátové katalógy, s cieľom poskytnúť informácie o význame, štruktúre a kontexte dát) a poskytovanie podpory používateľom pri navigácii a interpretácii údajov a pri otázkach alebo problémoch súvisiacich s údajmi.

Povinnostiam dátového kurátora je určená štvrtá časť zákona o údajoch, ktorá definuje, že dátový kurátor musí spĺňať odborné a osobnostné predpoklady podľa tohto zákona a má dohliadať na dodržiavanie povinností pri správe, zabezpečení prístupnosti a používania vybraných kategórií údajov evidovaných v registroch, **najmä povinností súvisiacich s dátovou kvalitou**. Dátový kurátor je podľa navrhovaného zákona povinný:

- koordinovať činnosti súvisiace s riadením dátovej kvality, meraním dátovej kvality a strategickým rozvojom v oblasti údajov,
- dohliadať na výkon merania dátovej kvality, vyjadrovať sa k výsledkom uskutočneného merania dátovej kvality a navrhovať opatrenia na nápravu nedostatkov zistených pri meraní dátovej kvality,
- zverejniť výsledky merania dátovej kvality vo forme otvorených údajov,

- prijímať a vyhodnocovať návrhy a podnety týkajúce sa dátovej kvality,
- vzdelávať sa na úseku dátovej kvality a v spôsoboch jej riadenia a merania.

Definovanie roly MIRRI ako gestora DK na Slovensku

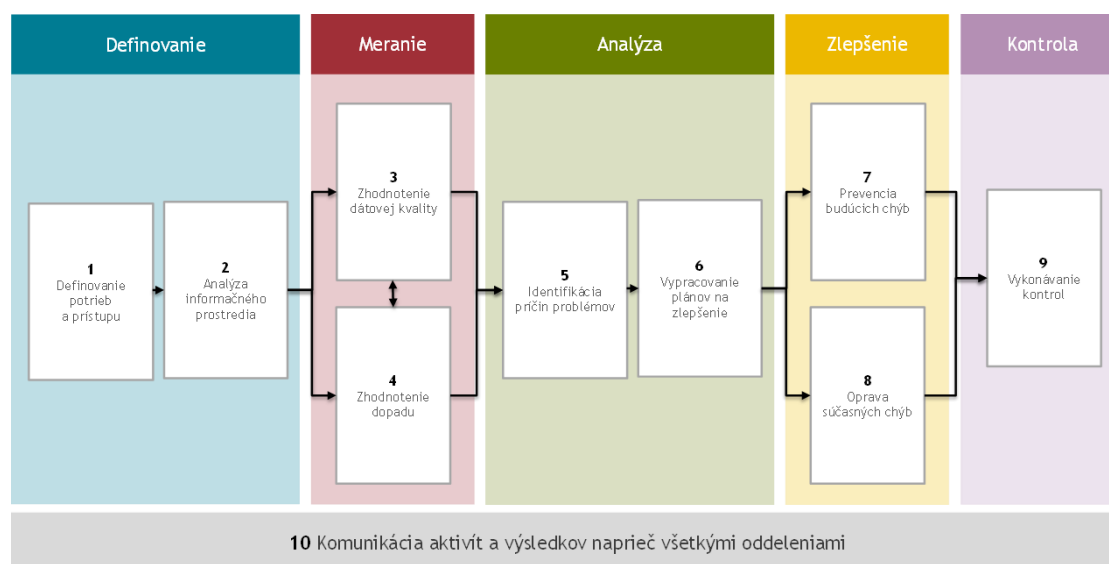
Okrem ustanovení týkajúcich sa dátovej kvality zákon o údajoch určuje MIRRI ako správcu Národného systému zdieľania údajov (NSZÚ), ktorý zabezpečuje vzájomnú komunikáciu informačných systémov na účely zdieľania údajov evidovaných v registroch, a to najmä na účely výkonu verejnej moci. Je určený na integráciu údajov a ich synchronizáciu pri referencovaní. Zabezpečuje jednotný spôsob ich poskytovania a zdieľania údajov evidovaných v registroch, ich pseudonymizáciu a anonymizáciu a sprístupnenie pre analytickú činnosť, ako aj evidenciu právnych základov pre ich získavanie.

Ďalšie pravidlá, ktoré zákon prináša

Zákon popisuje aj pravidlá správy, evidencie a používania referenčných údajov a registrov a venuje samostatnú časť aj stotožneniu údajov. Stotožnenie údajov je potvrdenie vzťahu medzi objektom evidencie a konkrétnym hmotným objektom alebo konkrétnym nehmotným objektom skutočného sveta v registri a identifikovanie konkrétneho hmotného objektu alebo konkrétneho nehmotného objektu v registri a v referenčnom registri, a to najmä na účely následného bezchybného referencovania. Zákon nešpecifikuje spôsoby stotožnenia, ale vymedzuje jeho účel, spoluprácu správcu registra a správcu referenčného registra, postup v prípade neúspešného stotožnenia a jeho dopad.

2.2 Stav v riadení 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 správe¹ (ďalej Metodika merania DK alebo len Metodika). Uvedený národný projekt medzitým pokračuje a trvá do konca roka 2023. Jeho cieľom je transformovať fungovanie inštitúcií verejnej správy tak, aby dokázali maximálne efektívne spravovať a zdieľať údaje, využívať údaje pre lepšie rozhodovanie na základe faktov a dôkazov, pre zlepšenie efektivity a adresnosti služieb na základe lepšieho využívania dát (zavedenie princípu „jeden krát a dost“) a v neposlednom rade pre publikovanie otvorených údajov, aby bola podporená inovatívna komunita, ktorá dokáže z údajov vyťažiť ďalšie hodnoty. Projekt by mal priniesť reálne zlepšenie manažmentu údajov vo verejnej správe a vytvoriť podmienky pre transformáciu štátu do podoby organizácie fungujúcej na základe využívania znalostí („data-driven state“). Po skončení by mali mať zapojené inštitúcie verejnej správy dostatočnú kapacitu pokračovať v tomto snažení a spravovať svoje dáta moderným spôsobom.



Obrázok 1 Proces zlepšovania dátovej kvality

Metodika merania DK navrhla 10 krokový proces merania a zlepšovania dátovej kvality, ktorý je zachytený na diagrame vyššie. 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. Nasledovná tabuľka ponúka prehľad ich napĺňania určený tak, ako sa nám to podarilo zistiť na základe stretnutí s Dátovou kanceláriou a s dátovými kurátormi. Nižšie je potom možné nájsť podrobné informácie za centrálnu a lokálnu úroveň, vyhodnotenú na úrovni jednotlivého navrhovaného opatrenia, opäť podľa informácií získaných od MIRRI v písomnej podobe alebo osobnými pohovormi.

¹ <https://datalab.digital/wp-content/uploads/Metodika-merania-d%C3%A1tovej-kvality-vo-verejnej-spr%C3%A1ve.pdf>

Tabuľka 1 Celkový stav iniciatív podľa oblastí

Dátové štandardy	Väčšina iniciatív prebieha, ale mnohé nie sú na správnej ceste
Organizačné zabezpečenie	U väčšiny iniciatív nie je zaznamenaný žiadúci pokrok alebo zaostávajú
Procesné zabezpečenie	Stav väčšiny iniciatív – žiadny pokrok alebo zaostáva
Technológie pre DK	Čiastočný pokrok, je potrebné zlepšenie a ďalší rozvoj

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

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

Tabuľka 2 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 možnosť 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 3 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 organizáciách 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 a 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 údaja 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

Na základe prehľadu plnenia by mali byť v jednotlivých oblastiach prioritne riešené tieto opatrenia:

Za oblasť Štandardy:

- Implementácia biznis pravidiel a pravidiel pre meranie kvality dát v organizáciách a nastavenie ich riadenia, vrátane stanovenia jasných KPI a ich prahových hodnôt ako aj určovania, ktoré atribúty / záznamy potrebujú merať kvalitu dát.
- Vytvorenie lokálnych (rezortných) dátových modelov, ich napojenie na centrálny model údajov (CMÚ) a stanovenie osobnej zodpovednosti za túto oblasť centrálnu a lokálnu, ako aj zjednotenie pravidiel tvorby a údržby dátových modelov a dátovej architektúry.

Za oblasť Organizácia:

- Budovanie kapacít pre organizačné zabezpečenie riadenia kvality dát formou dátových kurátorov.
- Zabezpečenie školení pre riadenie a správu kvality dát na rezortoch a vybudovanie „call centra“ pre poskytovanie poradenstva v oblasti kvality dát.

Za oblasť Procesy:

- Definovanie stratégie pre kvalitu dát (vrátane definovania jasných KPI, ktoré si organizácie následne osvoja) a neustále sledovanie jej implementácie v praxi, jej vyhodnocovanie a aktualizácia v pravidelných intervaloch.
- Jasné definovanie povinných a nepovinných údajov pre jednotlivé záznamy tak, aby pri vyhodnocovaní kvality bolo možné vidieť štatistiky pre povinné aj nepovinné údaje.
- Zabezpečenie jednoduchšieho prístupu k súborom údajov potrebným na meranie a riadenie kvality údajov

Za oblasť Technológia:

- Nastavenie technologických pravidiel pre prácu s databázami (jasne definované pravidlá správy dát, ktoré môžu byť podporené existujúcimi technológiami).
- Nastavenie pravidiel tak, aby záznamy nebolo možné doplniť (vytvoriť), ak neobsahujú všetky potrebné náležitosti (povinné polia) a ak nie sú dodržané definované pravidlá pre zadávanie údajov.

2.3 Podrobnejšie vyhodnotenie odporúčaní

Odporúčania na zlepšenie dátovej kvality verejnej správy na Slovensku, ktoré boli navrhnuté v rámci Metodiky merania DK, sa zameriavali na opatrenia, ktoré je potrebné realizovať na centrálnej alebo na lokálnej úrovni. Tieto odporúčania boli síce čiastočne implementované v praxi, ale ich aplikácia je vo väčšine prípadov nedostatočná. Od roku 2019 došlo aj k zmene priorít, ktoré v nich boli zahrnuté. Táto kapitola podrobnejšie popisuje a vyhodnocuje stav jednotlivých odporúčaní na základe informácií získaných počas stretnutí s Dátovou kanceláriou.

2.3.1 Dátové štandardy

Odporúčanie 1 (lokálna úroveň) – Zavedenie jasných biznis pravidiel pre riadenie dátovej kvality vrátane väzby biznis pravidiel na legislatívu upravujúcu dotknuté údaje.

Toto odporúčanie je určené pre lokálnu úroveň, ale jeho implementácia je ovplyvnená aktivitami odporúčaní 4 a 5 na centrálnej úrovni. OVM by mali preberať a lokálne využívať centralizované dátové pravidlá. Pokiaľ vznikne na lokálnej úrovni potreba definovať nové dátové pravidlá, mali by byť najprv zaradené do CMÚ a až následne prevzaté na lokálnej úrovni. Vďaka tomu centrálné úložisko rieši potrebu jasného definovania dátových pravidiel aj na lokálnej úrovni, aj keď súčasťou Dátového slovníka a CMÚ nie je definovanie väzby na legislatívny kontext.

Stav: Prebieha

Odporúčanie 2 (lokálna úroveň) – Implementácia biznis pravidiel a 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.

Rovnako ako v prípade predošlého odporúčania, na centrálnej úrovni budú dátové pravidlá nastavené v CMÚ. Pravidlá z centrálného modelu už začali využívať Ministerstvo vnútra SR (Register adries) a Ministerstvo financií SR (Centrálny ekonomický systém). Predpokladá sa, že postupne ich začnú využívať aj ostatné OVM. Vďaka existencii centrálného úložiska dátových pravidiel považujeme opatrenie za čiastočne implementované. Organizácie využívajú dátové pravidlá na vstupe údajov do registrov, ale nemajú definované súvisiace biznis pravidlá a KPI a nevykonávajú pravidelné merania dátovej kvality. Nie sú definované žiadne aktivity zamerané na to, aby došlo k zlepšeniu v tejto oblasti.

Stav: Zaostáva

Odporúčanie 3 (lokálna úroveň) – Definovanie presného formátu pre každý zadávaný údaj, aby bola zabezpečená konzistencia všetkých zdrojov, ktoré sú v organizáciách využívané.

Na centrálnej úrovni už je viditeľná snaha o zabezpečenie konzistentnosti dát prostredníctvom ontológií definovaných v CMÚ. Hoci nie je známa úroveň zabezpečenia konzistentnosti dát na lokálnej úrovni, v rámci komunikácie so správcami

lokálnych registrov všetci oslovení potvrdili, že lokálne toto odporúčanie priebežne aplikujú v záujme skvalitnenia údajov vstupujúcich do registrov.

Stav: Prebieha

Odporúčanie 4 (centrálna úroveň) – Vypracovanie jednotného dátového slovníka pre verejnú správu ako komponent centrálného dátového modelu.

V súčasnosti je vytváraný dátový slovník v rámci Talend riešenia v projekte Dátovej integrácie (DI), tento dátový slovník je limitovaný týmto špecifickým nástrojom pre procedúry dátovej kvality a nie je možné ho v súčasnej verzii využiť v rámci iných riešení. Preto sa do budúca odporúča vytvorenie dátového slovníku, ktorý bude postavený v open source riešení (napr. Apache Atlas), ktoré bude slúžiť ako úložisko pre dátový slovník, ktorý bude možné integrovať do rôznych nástrojov na spravovanie dátovej kvality.

Stav: Prebieha

Odporúčanie 5 (centrálna úroveň) – Vytvorenie centrálného dátového modelu verejnej správy a definovanie kompetencií a organizačného zabezpečenia dátového modelu.

Z hľadiska pravidiel pre riadenie dátovej kvality a merania dátovej kvality je dôležité vytvorenie jednotného dátového slovníka a zároveň nepretržité rozširovanie ontológií potrebných pre zavádzanie interoperability údajov v štátnej správe. V rámci tejto aktivity vznikol Centrálny model údajov (CMÚ) a Dátový slovník. Dátový slovník je zatiaľ nedokončený centrálny repozitár jednoduchých dátových pravidiel, v ktorých dátové pravidlo definuje jednu podmienku naviazanú len na jeden parameter (napr. podmienka vyplnenosti alebo podmienka konkrétneho formátu). V CMÚ sa zatiaľ nachádzajú ontológie a niektoré pravidlá vyjadrené pomocou shapes (napríklad ontológia fyzickej osoby²), ale postupne Dátová kancelária plánuje mapovať a schvaľovať nové ontológie, ktoré budú slúžiť na postupné zlepšovanie interoperability medzi jednotlivými OVM. CMÚ zatiaľ využívajú Ministerstvo vnútra SR v Registri adres a Ministerstvo financií v Centrálnom ekonomickom systéme. Postupne budú k CMÚ pripájané ďalšie OVM. Plánovaná je aj možnosť exportu ontológií a dátových pravidiel vo viacerých formátoch, aby ich mohli prevziať aj registre, ktoré nebudú mať prístup k centrálnym úložiskám. CMÚ a Dátový slovník spoločne určujú pravidlá pre každý zadávaný údaj z lokálnej úrovne. S CMÚ je spojené riziko nedostatočných personálnych kapacít na jeho ďalšie rozširovanie a údržbu v požadovaných časových horizontoch.

Stav: Prebieha

Odporúčanie 6 (lokálna úroveň) – Vytvorenie lokálnych (rezortných) dátových modelov a zavedenie osobnej zodpovednosti pre túto oblasť centrálnu aj na každom rezorte, plus jednotné pravidlá pre tvorbu a udržiavanie dátových modelov a dátovej architektúry.

OVM si môžu vytvárať vlastné lokálne modely, musia však používať dátové prvky z CMÚ. Ak identifikujú dátový prvok, ktorý zatiaľ nie je definovaný v CMÚ a slúži na

² <https://metais.vicempremier.gov.sk/confluence/download/attachments/59836112/physical-person-shapes.ttl?version=4&modificationDate=1634679873643&api=v2>

výmenu dát medzi OVM, mali by konzultovať s MIRRI, či je ho potrebné zaradiť do CMÚ, alebo bude využitá štandardizovaná zahraničná ontológia. OVM by sa mali integrovať na CMÚ, ktorý by mal byť udržiavaný v súlade s ich požiadavkami. Proces integrácie OVM na CMÚ je definovaný, ale zatiaľ sú naň integrované len Register adries a Centrálny ekonomický systém. V budúcnosti je plánovaná integrácia ďalších registrov.

Stav: Prebieha

Odporúčanie 7 (centrálna úroveň) – 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.

Metodika je definovaná a podporená aktuálne platnou legislatívou. V súčasnosti ju nevyužívajú všetky OVM, aj keď všetky OVM k nej majú prístup a podporu zo strany Dátovej kancelárie. MIRRI nemá v súčasnosti legislatívne postavenie na presadzovanie jej dodržiavania, ale v príprave je novelizácia, ktorá spresní metodické usmernenia aj postavenie MIRRI (Zákon o údajoch popísaný v kapitole **Chyba!** **Nenašiel sa žiaden zdroj odkazov.2-3-3**). Na podporu zlepšenia riadenia dátovej kvality môže MIRRI využívať pozitívnu motiváciu prostredníctvom dotačných výziev.

Stav: Zaostáva

Odporúčanie 8 (lokálna úroveň) – Zabezpečenie referencovania získavaných údajov organizáciou alebo údajov vedených v databázach organizácií na existujúce referenčné registre.

Pri vyhodnocovaní opatrenia neboli k dispozícii podrobné informácie za všetky OVM, ale v určitej forme je stotožňovanie a referencovanie zabezpečené vo viacerých rezortoch.

Stav: Prebieha

Odporúčanie 9 (lokálna úroveň) – Zabezpečenie súladu súčasného stavu s existujúcimi štandardami dátovej kvality.

Podľa vyjadrenia Dátovej kancelárie je metodika štandardov definovaná a podporená legislatívou, pripravuje sa jej novelizácia (pripravovaný Zákon o údajoch). Nie všetky OVM ju však využívajú, aj napriek tomu, že je k dispozícii. MIRRI nemá legislatívne postavenie na vynútenie.

Stav: Zaostáva

Odporúčanie 10 (centrálna úroveň) – Definovanie oblastí dátovej kvality pre zabezpečenie líderstva v tejto oblasti na úrovni EU.

Toto odporúčanie určuje ambíciu, ktorá je súčasnej situácii v dátovej kvalite na Slovensku príliš vzdialená. V súčasnosti je dôležité zamerať sa na zlepšenie riadenia dátovej kvality, napríklad pripravovanou dátovou legislatívou, vytvorením jednotného bodu pre distribúciu biznis pravidiel a ontológií a najmä zvýšením počtu OVM, ktoré pri riadení dátovej kvality postupujú v súlade s metodickými usmerneniami Dátovej kancelárie.

Stav: Žiadny pokrok

Odporúčanie 11 (lokálna úroveň) – 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.

Podľa vyjadrenia Dátovej kancelárie je metodika štandardov definovaná a podporená legislatívou, pripravuje sa jej novelizácia (pripravovaný Zákon o údajoch). Nie všetky OVM ju však využívajú, aj napriek tomu, že je k dispozícii. MIRRI nemá legislatívne postavenie na vynútenie.

Stav: Zaostáva

Odporúčanie 12 (centrálna úroveň) – 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.

Rovnako ako v prípade odporúčania 7, aj v tomto prípade je metodika definovaná a podporená aktuálne platnou legislatívou. V súčasnosti ju nevyžívajú všetky OVM, aj keď všetky OVM k nej majú prístup a podporu zo strany Dátovej kancelárie. MIRRI nemá v súčasnosti legislatívne postavenie na presadzovanie jej dodržiavania, ale v príprave je novelizácia, ktorá spresní metodické usmernenia aj postavenie MIRRI (Zákon o údajoch popísaný v kapitole **[Chyba! Nenašiel sa žiaden zdroj odkazov.2-3-3](#)**).

Stav: Prebieha

2.3.2 Organizácia

Opatrenie 13 (centrálna úroveň) – Definovanie jasných kompetencií, právomocí a zodpovednosti pre pozície dátových kurátorov.

Kompetencie, právomoci a zodpovednosti dátových kurátorov sú jasne zadefinované v pripravovanom Zákone o údajoch, ktorý však s najväčšou pravdepodobnosťou nebude prijatý v roku 2023. Zatiaľ bola rola dátového kurátora zriadená uznesením Vlády SR a jeho právomoci a kompetencie boli definované metodickým usmernením.

Na zvážení je, či je rola dátového kurátora jediná, ktorá by mala byť podporená legislatívou. Riadenie a zvyšovanie dátovej kvality vyžaduje aj ďalšie roly, ako

napríklad dátový architekt alebo dátový analytik. Navyše, samotné definovanie rol negarantuje, že všetky OVM tieto pozície obsadia a že sa aktívne zapoja do procesov riadenia dátovej kvality. V prípade tohto opatrenia by bolo vhodné, aby Dátová kancelária zjednotila a metodicky nastavila procesy dátovej kvality na všetkých OVM.

Stav: Prebieha

Odporúčanie 14 (lokálna úroveň) – Posilnenie kapacít pre organizačné zabezpečenie riadenia dátovej kvality vo forme dátových kurátorov.

V mnohých organizáciách existuje rola dátového kurátora len formálne. Rola je síce pridelená zamestnancom, ktorých úlohou je však vykonávať aj iné aktivity a majú aj iné zodpovednosti, navyše ich povinnosti a kompetencie ako kurátorov nie sú jasne stanovené a nie sú na túto prácu vyškolení.

Pozn. Počas kompletizácie tohto výstupu prebiehali na MIRRI školenia pre dátových kurátorov (pozri nižšie).

Stav: Zaostáva

Opatrenie 15 (centrálna úroveň) – Zabezpečenie školení pre oblasť riadenia a správy dátovej kvality v rezorte a vybudovanie „call centra“ pre zabezpečenie poradenstva v oblasti dátovej kvality.

Toto odporúčanie zahŕňa časť školení a časť vybudovania call-centra pre zabezpečenie poradenstva. Pokiaľ ide o školenia, Dátová kancelária publikovala viaceré metodické usmernenia pre zvýšenie dátovej kvality a pred dvoma rokmi boli realizované školenia na tvorbu biznis procesov. V priebehu prvého polroka 2023 prebiehajú školenia pre dátových stewardov na používanie konzoly Talend Data Stewardship (TDS) a v príprave sú školenia na tému interoperability.

V súčasnosti nie je plánované vytvorenie call-centra. V rámci aktivít Dátovej kancelárie vznikla iniciatíva delegovať komunikačných manažérov pre jednotlivé OVM, ktorí budú mať prehľad o tom, aké projekty na OVM prebiehajú.

Stav: Zaostáva

Odporúčanie 16 (centrálna úroveň) – Posilnenie kompetencie Dátovej kancelárie v oblasti dátovej kvality tak, aby v súčasnosti existujúce pravidlá vedela kontrolovať a podmieňovať realizáciu projektov napr. z dopytových výziev a vyvodzovať prípadné konsekvencie.

Dátová kancelária má už dnes kompetencie kontroly v oblasti dátovej kvality, vrátane dodržiavania publikačného minima, ale nemá žiadne kompetencie z hľadiska vyvodzovania následkov. To sa nezmení ani prijatím nového Zákona o dátach. Dátová kancelária potvrdila, že udeľovanie sankcií za nedodržiavanie povinností nepovažuje

za efektívne. Vhodnejšie je prihliadať na dodržiavanie pravidiel pri prideľovaní finančných prostriedkov prostredníctvom dotačných výziev.

Stav: Zaostáva

Odporúčanie 17 (lokálna úroveň) – 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.

V čase prípravy tohto dokumentu neboli k dispozícii konkrétne informácie o tom, ako jednotlivé OVM aplikujú kontrolu a overovanie údajov pri vstupe do svojich registrov. Je však možné konštatovať, že hoci väčšina OVM implementovala vo svojich vstupných formulároch a importoch určitú úroveň automatizovanej kontroly (napríklad upozornenie na nesprávny formát zadávaných údajov vo formulároch alebo zamietnutie importu v nesprávnej štruktúre) OVM zväčša nestotožňujú údaje pri zápise do registra, aby eliminovali vznik nekvalitných záznamov a nutnosť ich následnej prácnej kontroly a opravy.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 18 (centrálna úroveň) – Nastavenie pravidiel pre opravy identifikovaných nezrovnalostí v záznamoch tak, aby bol proces čo najkratší a aby bol realizovateľný v čase vzniku alebo odhalenia nekonzistentnosti v údajoch.

Pravidlá opravy údajov sú definované v § 53 zákona o e-Governmente³, ale jednotlivé OVM môžu k oprave údajov definovať vlastný prístup, častokrát na základe súvisiacej legislatívy. V súčasnosti ponúka Dátová kancelária pre OVM na opravu údajov len konzolu Talend Data Stewardship na platforme Talend, ale zabezpečenie licencií na prístup k nástroju je zodpovednosťou OVM. OVM však môžu na opravu údajov využívať vlastné nástroje.

Pri definovaní pravidiel opravy údajov je potrebné vziať do úvahy aj to, či ich môže vykonať správca registra, alebo ich musia vykonať zdrojové registre, ako je tomu napríklad v prípade RPO, čo má dopad na trvanie procesu opravy záznamov.

Stav: Prebieha

Odporúčanie 19 (centrálna úroveň) – Jasné nastavenie pravidiel hodnotenia a odmeňovania pre zodpovedné osoby za dátovú kvalitu reflektujúc 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“.

Toto odporúčanie nepovažuje Dátová kancelária za zmysluplné. Dôležité je, aby bol v každom OVM určený niekto, kto zodpovedá za dátovú kvalitu. Chýba však

³ [305/2013 Z.z. - Zákon o elektronickej podobe výkonu... - SLOV-LEX](#)

vykonávanie pravidelných meraní DK a publikovanie jednotných výsledkov krížom cez všetky rezorty, na základe ktorých by bolo možné zostavovať povinné rebríčky dátovej kvality. To sa deje len v prípade otvorených údajov, u ktorých je na európskej úrovni automatizovane posudzovaná kvalita metadát (nie samotných dát) na základe centrálnych pravidiel. Výsledkom jednotného reportovania naprieč rezortami by bola tzv. tabuľa hanby a nízke umiestnenie v rebríčku by mohlo mať dopad na schvaľovanie financovania ďalších projektov.

Stav: Žiadny pokrok

Odporúčanie 20 (lokálna úroveň) – Implementovanie zmeny v oblasti dátovej kultúry a interpretácie údajov, ktoré má organizácia vo „vlastníctve“ vrátane zavedenia biznis metadát.

Počas diskusie s Dátovou kanceláriou nebolo možné posúdiť, či bolo toto opatrenie implementované a na akej úrovni. Na lokálnej úrovni boli pravdepodobne nejaké zmeny implementované, ale k dispozícii nie sú žiadne informácie. Podobne ako napríklad pri odporúčaní 10, v porovnaní s mnohými členskými štátmi EÚ má Slovensko v zlepšovaní riadenia a merania dátovej kvality väčšie medzery, ktorých zaplnenie má vyššiu prioritu.

Stav: Žiadny pokrok nie je publikovaný

2.3.3 Proces

Odporúčanie 21 (centrálna úroveň) – Definovanie stratégie pre dátovú kvalitu (vrátane definovania jasných KPIs, ktoré budú následne prijaté organizáciami) a neustále monitorovanie jej zavedenia v praxi a v pravidelných intervaloch vyhodnocovanie a aktualizácia.

Za účelom zvýšenia interoperability bol vytvorený CMÚ. Zatiaľ sú v ňom zdokumentované len ontológie, ale postupne majú pribúdať aj centrálné dátové modely a pravidlá. CMÚ však nie je určený na evidovanie KPI. Tie sú pre jednotlivé registre jedinečné podľa ich účelu a požiadaviek. OVM by si mali zvážiť ich potrebu a stanovenie samostatne.

Stav: Žiadny pokrok

Odporúčanie 22 (centrálna úroveň) – Zavedenie povinnosti vykonávania pravidelného merania a vyhodnocovania dátovej kvality so zverejňovaním výsledkov vo forme open data.

Táto povinnosť zatiaľ nie je zavedená. Pokiaľ aj jednotlivé OVM vykonávajú pravidelné merania DK a vyhodnocujú ich výsledky, reporty z meraní nie sú zverejňované.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 23 (lokálna úroveň) – Prehodnotenie a nastavenie pravidiel opráv chybných záznamov tak, aby bol proces čo najkratší, s jasne nastavenými KPIs vedúcimi k motivácii na jeho výkon.

Toto odporúčanie nebolo implementované v praxi a v súčasnosti nie je považované za prioritu. Po zavedení pravidelných meraní dátovej kvality a jej vyhodnocovania by však mal byť implementovaný aj tento krok s prihliadnutím na procesy jednotlivých OVM a metodiky ich riadenia dátovej kvality.

Stav: Žiadny pokrok

Odporúčanie 24 (lokálna úroveň) – Jasné definovanie povinných a nepovinných údajov pri jednotlivých záznamoch tak, aby bolo pri hodnotení údajov vidieť, aká je štatistika pre povinné aj pre nepovinné údaje.

Jednotlivé registre pri vstupe údajov definujú, ktoré údaje sú povinné a nepovinné. K dispozícii nie sú žiadne informácie o tom, nakoľko a akým spôsobom je táto požiadavka vyhodnocovaná.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 25 (lokálna úroveň) – Zabezpečenie lepšej osvety pre zadávanie nepovinných údajov a vysvetlenie pridanej hodnoty zadania nepovinného údaja do systému.

K dispozícii nie sú žiadne informácie o osvete v tejto oblasti. Je plne v kompetencii a v záujme každého vlastníka údajov posúdiť, ktoré údaje by mali byť povinné alebo nepovinné. V prípade nepovinných údajov by mali OVM jednotlivo posúdiť ich účel a dopad ich nevyplnenosti a na základe toho zvážiť povinnosť ich vyplnenia alebo zvýšenie osvety smerovanej na zdroje údajov.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 26 (lokálna úroveň) – Zavedenie jasných pravidiel pre aktualizáciu záznamov a ich atribútov, ako aj implementácia procesov monitoringu a dodržiavania pravidiel.

Hoci nie sú k dispozícii podrobné informácie o implementácii tohto odporúčania, je predpoklad, že pre každý register tieto pravidlá existujú, aj keď nie je jasné, nakoľko je ich dodržiavanie monitorované.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 27 (lokálna úroveň) – Nastavenie jednoduchých procedúr pre transformovanie údajov do foriem využiteľných pre realizáciu procesov podporujúcich dátovú kvalitu alebo procesov tvorby reportov a analýz.

Hoci nie sú z lokálnej úrovne k dispozícii informácie o implementácii tohto odporúčania, zisťovanie stavu DK prostredníctvom meraní a reportovanie ich výsledkov je nevyhnutnou súčasťou zlepšovania DK. S tým súvisí aj prípadné transformovanie údajov pre tento účel. Keďže OVM nevykonávajú pravidelné merania DK a nepublikujú ich výsledky, toto odporúčanie pravdepodobne nebolo implementované.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 28 (lokálna úroveň) – Nastavenie procesov prípravy údajov na analytické spracovanie (forma, štruktúra potrebná pre analýzy systémoch ako napr. R, SAS, Python, ...).

Z lokálnej úrovne nie sú k dispozícii žiadne informácie o nastavení procesov prípravy údajov na analytické spracovanie, ani o tom, nakoľko a akým spôsobom OVM využívajú údaje vo svojich datasetoch ako vstupy pre analytické procesy. Na mnohých rezortoch existujú analytické útvary⁴, napríklad Inštitút správnych a bezpečnostných analýz na MV SR alebo Útvar hodnoty za peniaze na MF SR, ale spracovateľom tohto dokumentu nie je známe, s akými nástrojmi pracujú a či pri vkladaní údajov do analytického nástroja využívajú transformačné procesy. Na centrálnej úrovni bol za týmto účelom spustený projekt KAV (Konsolidovaná analytická vrstva), ktorý ale zatiaľ nebol úspešne ukončený.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 29 (lokálna úroveň) – 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).

Rovnako ako v prípade odporúčania 27, keďže OVM nevykonávajú pravidelné merania DK a nepublikujú ich výsledky, nie je možné posúdiť, nakoľko bolo toto odporúčanie implementované. Platí však, že pokiaľ OVM riadi a meria kvalitu svojich údajov, je v jeho záujme čo najviac zjednodušiť prístup k meraným datasetom. Je v kompetencii každého OVM definovať kompetencie prístupu, kto a kedy môže vygenerovať potrebnú zostavu alebo export.

Stav: Žiadny pokrok nie je publikovaný

⁴ [Analytické jednotky a ich inštitucionálne zastrešenie, Ministerstvo vnútra SR - Verejná správa \(minv.sk\)](#)

2.3.4 Technológie

Odporúčanie 30 (centrálna úroveň) – Zabezpečenie príslušnej technologickej podpory pre riadenie kvality údajov pre všetky definované oblasti vychádzajúce zo stratégie riadenia kvality údajov.

Na centrálnej úrovni bolo sprístupnené jedno riešenie na meranie DK a jej zlepšovanie, postavené na platforme Talend. Licencie na prístup si musia OVM zabezpečiť na vlastné náklady, ale majú plne k dispozícii vyvinutú funkčnosť aj školenia a Dátová kancelária neustále pracuje na ďalšom rozvoji prostredia.

Stav: Prebieha

Odporúčanie 31 (lokálna úroveň) – Implementácia a zabezpečenie nástroja na riadenie dátovej kvality (v súčasnosti existuje platforma Talend, ktorá je zabezpečená na úrovni PaaS služby v štátnom cloude).

Toto odporúčanie je síce určené pre lokálnu úroveň, ale jeho implementácia je zabezpečená centrálnou poskytnutím riešenia, ktoré malo by byť implementované do v rámci rozšírenia CSRÚ (PaaS). Dátová kancelária zatiaľ neprijala rozhodnutie, na akej platforme bude riešenie nasadené.

Stav: Prebieha

Odporúčanie 32 (lokálna úroveň) – Nastavenie technologických pravidiel pre prácu s databázami (jasne definované pravidlá správy údajov, ktoré sa dajú podporiť existujúcimi technológiami).

Spracovatelia dokumentu nemali k dispozícii informácie o implementácii tohto odporúčania na rezortnej úrovni, ale je možné predpokladať, že tieto pravidlá boli definované. Bolo by vhodné, aby Dátová kancelária vytvorila metodiku, ktorá by zahŕňala aj odporúčania pre tieto nastavenia a vyzvala OVM na potvrdenie ich implementácie.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 33 (lokálna úroveň) – Aplikovanie systémov AI pre technologické riadenie kvality údajov.

Na centrálnej úrovni je v CMÚ implementované sémantické strojové usudzovanie. Nie sú k dispozícii informácie o tom, či sú systémy AI využívané na lokálnej úrovni. Ak áno, je to pravdepodobne len v prípade tých, ktoré využívajú CMÚ a LinkedPipes.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 34 (lokálna úroveň) – Nastavenie procesov profilovania a analyzovania údajov priamo v technologickom riešení.

Spracovatelia dokumentu nemali k dispozícii informácie o implementácii tohto odporúčania na lokálnej úrovni. Posúdenie potreby jeho implementácie je v kompetencii jednotlivých OVM a ich analytických útvarov.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 35 (lokálna úroveň) – Nastavenie proaktívnych kontrol pre zdrojové systémy – kontrola bude prebiehať už pri vstupe údajov na základe definovaných technologických a procesných požiadaviek.

Spracovatelia dokumentu nemali k dispozícii informácie o implementácii tohto odporúčania na lokálnej úrovni, ale je možné predpokladať, že kontroly na vstupe boli s najväčšou pravdepodobnosťou nastavené vo väčšine zdrojových registrov. Bolo by vhodné, aby Dátová kancelária vytvorila metodiku, ktorá by zahŕňala aj odporúčania pre tieto kontroly a vyzvala OVM na potvrdenie ich implementácie.

Stav: Prebieha

Odporúčanie 36 (lokálna úroveň) – Zabezpečenie technologickej podpory pre definované dátové pravidlá pri vstupe údajov do systému.

Spracovatelia dokumentu nemali k dispozícii informácie o tom, ako je zabezpečená technologická podpora pre dodržiavanie dátových pravidiel pri vstupe údajov do rezortných systémov. Odporúčanie je však naďalej platné.

Stav: Žiadny pokrok nie je publikovaný

Odporúčanie 37 (lokálna úroveň) – 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.

Spracovatelia dokumentu nemali k dispozícii informácie o tom, ako je toto odporúčanie implementované v rezortných systémoch, ale tieto kontroly s najväčšou pravdepodobnosťou prebiehajú. Bolo by vhodné, aby Dátová kancelária vytvorila metodiku, ktorá by zahŕňala aj odporúčania pre tieto kontroly a vyzvala OVM na potvrdenie ich implementácie.

Stav: Implementácia nebola overiteľná

Odporúčanie 38 (lokálna úroveň) – Implementácia nástroja / algoritmov na definovanie duplicity záznamov pri vstupe údajov do systému alebo na podporu procesu deduplikácie záznamov.

Spracovatelia dokumentu nemali k dispozícii informácie o tom, ako je toto odporúčanie implementované v rezortných systémoch, ale tieto kontroly s najväčšou pravdepodobnosťou prebiehajú. Bolo by vhodné, aby Dátová kancelária vytvorila metodiku, ktorá by zahŕňala aj odporúčania pre tieto kontroly a vyzvala OVM na potvrdenie ich implementácie.

Stav: Implementácia nebola overiteľná

Odporúčanie 39 (lokálna úroveň) – Identifikácia možností a následného zabezpečenia technologickej podpory pre elimináciu faktora „zlyhanie ľudského faktora“ vo všetkých procesoch a fázach práce s údajmi.

Spracovatelia dokumentu nemali k dispozícii informácie o tom, ako je toto odporúčanie implementované v rezortných systémoch. Bolo by vhodné, aby Dátová kancelária vytvorila metodiku, ktorá by zahŕňala aj odporúčania pre elimináciu zlyhania ľudského faktora pri práci s údajmi a vyzvala OVM na posúdenie alebo potvrdenie ich implementácie.

Stav: Žiadny pokrok nie je publikovaný

2.4 Stav v oblasti merania kvality dát

2.4.1 Doterajšie aktivity merania DK

Po vydaní Metodiky merania DK⁵ bolo v roku 2020 v spolupráci s niekoľkými OVM vykonané pilotné meranie. Meranie bolo vykonané na nasledujúcich datasetoch:

- Koncesie v taxislužbe, Ministerstvo dopravy a výstavby Slovenskej republiky,
- Geopriestorová žiadosť o podporu - zoznam žiadostí o priamu podporu spolu s deklaračnými tabuľkami z kampane 2019, Ministerstvo pôdohospodárstva a rozvoja vidieka Slovenskej republiky,
- Databáza diplomatov na území SR, Ministerstvo zahraničných vecí a európskych záležitostí Slovenskej republiky,
- Register právnických osôb, Štatistický úrad,
- Údaje z Modelu dotačných schém za rok 2018, Úrad podpredsedu vlády pre investície a informatizáciu,
- Centrálny register zmlúv, Úrad vlády SR,
- Geografické názvy nesídelných objektov, Úrad geodézie, kartografie a katastra SR.

Najprv bol v spolupráci s OVM určený zoznam všetkých kľúčových atribútov v meraní datasetu. Následne boli pre tieto atribúty definované biznis pravidlá s určením zdroja biznis pravidla a cieľa biznis pravidla. Pri určovaní biznis pravidiel sa vychádzalo zo zadaných pravidiel na povinnosť uvedenia údajov, ich prepojenie s číselníkmi a na ich formát. Potom boli definované biznis pravidlá na základe logických súvislostí medzi atribútmi. Na základe výsledkov vyhodnotenia biznis pravidiel boli nakoniec formulované odporúčania pre jednotlivé datasety.

Pilotné meranie dátovej kvality bolo síce vykonané v súlade s metodikou, ale len ako generické informatívne meranie. Neboli určené žiadni konzumenti dát, žiadne KPI a neboli merané žiadne ďalšie parametre dátovej kvality. Po skončení pilotného projektu prebiehalo do marca 2022 už len pravidelné kvartálne informatívne meranie

⁵ <https://datalab.digital/wp-content/uploads/Methodika-merania-d%C3%A1tovej-kvality-vo-verejnej-spr%C3%A1ve.pdf>

dátovej kvality na datasete Štatistického úradu. Po tomto termíne neboli publikované žiadne informácie o meraní dátovej kvality na žiadnom z OVM.

Okrem pilotného merania bolo v roku 2021 vykonané aj Hodnotenie dátovej kvality otvorených údajov⁶ na štyroch vybraných datasetoch:

- Centrálny register študentov,
- Register nehnuteľných národných pamiatok,
- Železničné trate,
- Objednávky a faktúry MIRRI SR.

V priebehu tohto roka je naplánované hodnotenie ďalších publikovaných datasetov otvorených údajov.

2.4.2 Požiadavky a politiky riadenia dátovej kvality

Riadenie dátovej kvality je dôležité pre meranie a dosahovanie potrebných štandardov dátovej kvality. V rámci orgánov verejnej moci (OVM) spadajú tieto aktivity do zodpovednosti dátových kurátorov s jednotlivými vymedzenými právomocami, pomocou ktorých majú pracovať na dosahovaní štandardov dátovej kvality.

V metodike merania DK sú špecifikované zodpovednosti a kompetencie dátových kurátorov. Po vydaní metodiky meranie dátovej kvality bolo vykonané pilotné Informatívne meranie dátovej kvality v rámci niektorých OVM, ktoré v ňom boli ochotné participovať. Do tohto merania boli zapojení dátoví kurátori príslušných OVM, ktorý spolupracovali s Ministerstvom investícií regionálneho rozvoja a informatizácie na realizácii pilotného merania. Aj napriek tomu, že v rámci metodiky je určený rozsah výkonu roly kurátora, v súčasnosti nie je v platnosti žiadna právne záväzná regulácia, z ktorej by vyplývala povinnosť vykonávať činnosti v plnom rozsahu určenom v metodike. Z tohto faktu následne aj pramenil rozsah, v ktorom sa jednotliví dátoví kurátori zapojili do pilotného merania dátovej kvality.

Keďže v rámci pilotného Informatívneho merania dátovej kvality boli dátoví kurátori zapojení z každého OVM v rôznom rozsahu, väčšina činností v rámci merania bola zastrešená Ministerstvom investícií regionálneho rozvoja a informatizácie. Po vykonaní pilotného Informatívneho merania dátovej kvality už dátoví kurátori ďalej nejavili záujem merať dátovú kvalitu, výnimkou bol len dátový kurátor štatistického úradu, ktorý zaviedol pravidelné Monitorovacie merania

V Metodike merania DK sú uvedené typy merania dátovej kvality rozdelené do jednotlivých činností. Pre tieto činnosti sú určené úrovne zodpovednosti dátového kurátora. Na základe týchto definovaných úrovní zodpovednosti doposiaľ čiastočne naplňal svoje povinnosti len dátový kurátor Štatistického úradu v rámci Monitorovacieho merania vykonávaného štvrťročne. V súčasnosti už nedochádza k vykonávaniu žiadneho typu meranie dátovej kvality na žiadnom z OVM, keďže Monitorovacie meranie na Štatistickom úrade bolo ukončené v dôsledku personálnych zmien.

⁶ [Výstup č. 2 – Dohľadanie nad kvalitou publikovaných datasetov a kontrola, či nedochádza k zneužitiu údajov – DataLab](#)

Súčasný stav nečinnosti dátových kurátorov v rámci jednotlivých OVM v riadení dátovej kvality je spôsobený nedostatkom legislatívy, ktorá by ustanovila ich kompetencie a zodpovednosti. V tejto oblasti sa aktuálne pripravuje Zákon o údajoch zameraný na komplexnú legislatívnu reguláciu dátovej kvality (*kapitola **Chyba! Nenašiel sa žiaden zdroj odkazov.**2.1.3*). V rámci tohto zákona bude taktiež ustanovený aj legislatívny rámec fungovanie roly dátového kurátora, ktorý je založený metodike merania dátovej kvality. Tento zákon teda pre dátových kurátorov vytvorí právny základ na napĺňanie ich povinností vyplývajúcich z metodiky merania dátovej kvality a teda dopomôže k zvyšovaniu dátovej kvality v rámci štátnej správy.

2.4.3 Predchádzajúce odporúčania a ich relevantnosť

Metodika merania DK stanovila štandardy a odporúčania pre zlepšenie a udržiavanie dátovej kvality a zároveň zdôraznila dôležitosť široko prijatých, centrálnych riadených a pravidelných procesov merania a monitorovania kvality dát najmä pre celý proces riadenia kvality dát. V tejto kapitole sa pozrieme na aplikáciu metodiky v praxi a jej relevantnosť v súčasnej dobe.

Meranie dátovej kvality nie je konzistentne zavedené do praxe OVM

Tabuľka nižšie sumarizuje stav implementácie procesov merania tak, ako sa nám ho podarilo zistiť z informácií poskytnutých objednávateľom. Hodnotenie relevantnosti je uvedené v príslušnom stĺpci pre povinné kroky procesu (znamienko plus). Odporúčania boli čiastočne použité v rámci pilotného projektu merania DK v 2020, ktorý viedol A. Fukas z MIRRI. Rozsiahlejšie merania dátovej kvality uskutočňoval v minulosti Štatistický úrad, ktorý zverejňoval výsledky meraní⁷ na svojom webovom sídle, tieto aktivity však skončili v polovici roka 2022. Žiadne ďalšie pravidelné merania kvality údajov u iných OVM sa nám žiaľ z verejne dostupných zdrojov nepodarilo potvrdiť.

Tabuľka 4 Hodnotenie 7-krokového procesu informatívneho merania DK

Krok	Popis	Relevantnosť	Zistený stav
Definovanie objektu	Identifikácia a definícia objektu na meranie	+	Čiastočne pilotované
Získanie dát	Získanie dát v rozsahu a formáte, definovanom pre vykonanie merania.	+	Čiastočne pilotované
Výber parametrov	Výber parametrov kvality, ktoré meriame a ktoré odrážajú biznis potreby príp. výsledky profilovania údajov.	+	Čiastočne pilotované
Výber ukazovateľov	Určenie ukazovateľov patriacich k vybraným parametrom (z kroku 3).	+	Čiastočne pilotované
Meranie	Vykonanie merania zvoleného súboru údajov vo vybranom nástroji	+	Čiastočne pilotované
Vytvorenie protokolu	Správa vygenerovaná z nástroja (protokol merania), v ktorom sa vykonalo meranie kvality údajov	+	Čiastočne pilotované

⁷ [Správy z merania dátovej kvality RPO \(statistics.sk\)](https://statistics.sk)

**Odoslanie
protokolu**

Správa sa odošle príslušným adresátom
(vlastník údajov, orgán dohľadu a pod.)

+

Žiadna informácia

Pre ďalšie procesy merania kvality dát, navrhnuté v rámci metodiky (monitorovacie meranie, komplexné meranie a kontrolné meranie), sa nám nepodarilo zavedenie do štandardizovanej praxe, okrem obmedzených iniciatív vykonávaných ad hoc ako pilotných aktivít. Taktiež sa nám nepodarilo zistiť žiadne zverejňované informácie o automatizácii procesov merania, dashboardoch alebo pravidelných správach používaných na monitorovanie kvality údajov.

Zlepšovanie kvality dát nie je systematické, dosiahnuté výsledky sa pravidelne nezverejňujú

Tabuľka nižšie sumarizuje stav implementácie procesov zlepšovania kvality údajov tak, ako sa nám ho podarilo zistiť z informácií poskytnutých objednávateľom príp. z verejne dostupných zdrojov. Podobne ako vyššie je uvedené aj hodnotenie relevantnosti odporúčaní pre súčasnosť (znamienko plus).

Tabuľka 5 Hodnotenie 10-krokového procesu na zlepšenie kvality údajov

Krok	Popis	Relevantnosť	Zistený stav
Definovanie potrieb a prístupu	Definovanie a odsúhlasenie hlavných problémov, príležitostí a cieľov, ako riadiť a vykonávať prácu počas projektu.	+	Nestanovené, obmedzený pokrok v rámci roztrieštených iniciatív
Analýza informačného prostredia	Zhromažďovanie, vytváranie a analyzovanie informácií o aktuálnom stave informačného prostredia. Dokumentovanie a overovanie životného cyklu informácií, ktoré poskytujú základ pre budúce akcie.	+	Nestanovené, obmedzený pokrok v rámci roztrieštených iniciatív
Hodnotenie kvality dát	Výsledky hodnotenia poskytujú základ pre budúce kroky (identifikácia príčiny problémov, potrebné zlepšenia a opravy údajov)	+	Nestanovené, obmedzený pokrok v rámci roztrieštených iniciatív
Hodnotenie vplyvu	Tento krok poskytuje zdôvodnenie a argumenty na zlepšenie kvality údajov, získanie podpory pre kvalitu údajov a určenie vhodných investícií do informačných zdrojov.	+	Nie je zavedené
Identifikácia príčiny problémov	Identifikovanie a prioritizácia dôležitosti príčin problémov s kvalitou údajov a vypracovanie odporúčaní na ich riešenie	+	Nie je zavedené
Vypracovanie plánov na zlepšenie	Vypracovanie akčného plánu na základe odporúčaní z hodnotenia kvality údajov, výsledkov z hodnotenia vplyvov a identifikácie príčiny problémov.	+	Nie je zavedené
Predchádzanie budúcim chybám	implementácia riešení, ktoré preventívne riešia hlavné príčiny problémov s kvalitou údajov.	+	Nie je zavedené
Oprava aktuálnej chyby	Vykonávanie krokov na opravu aktuálne zistených chýb údajov.	+	Nie je zavedené
Vykonávanie kontrol	Monitorovanie a overovanie dosiahnutých zlepšení.	+	Nie je zavedené
Komunikácia aktivít a výsledkov	Dokumentácia a oznamovanie výsledkov kontroly kvality a vykonaných zlepšení.	+	Nie je zavedené

Na základe posúdenia získaných informácií sme skonštatovali, že meranie a zlepšovanie kvality údajov sa v súčasnosti vykonáva ad hoc, na úrovni samotných OVM, centrálné schválené odporúčania, techniky a smernice sú implementované do bežnej praxe obmedzene alebo vôbec a o prebiehajúcich iniciatívach (ak sa konajú) nie sú verejne dostupné informácie. Najvýraznejší pokrok smerom k iniciatívam merania kvality dát podľa dostupných informácií dosiahol Štatistický úrad, ktorý sa v minulosti pravidelne pokúšal tento proces zaviesť, no aj tieto snahy boli pred viac ako rokom ukončené. Od roku 2019 až do marca 2022 pravidelne vykonával kvartálne merania DK a zverejňoval ich výsledky. Tieto aktivity však boli prerušené, zatiaľ nie je známu, či a kedy ŠÚ SR pripravuje ich opakované spustenie.

2.5 Centrálné projekty riadenia dát s väzbou na dátovú kvalitu

V súčasnosti neexistuje plne funkčné centrálné riešenie, ktoré by umožňovalo orgánom verejnej moci (OVM) riadiť dátovú kvalitu. Avšak existuje niekoľko projektov, ktoré budú poskytovať služby riadenia dátovej kvality:

- Dátová integrácia (Procedúry dátovej kvality)
- Centrálna integračná platforma (Služba PaaS)

Okrem centrálnych riešení poskytujúcich služby riadenia dátovej kvality bolo taktiež zrealizovaných aj niekoľko dopytových projektov, v rámci ktorých sa OVM v rôznom rozsahu venovali aj dátovej kvalite.

2.5.1 Dátová integrácia (DI)

Projekt DI⁸ je zameraný na naplnenie princípu “Jedenkrát a dost” tak, aby v čo najväčšej možnej miere odbremenil občana/podnikateľa od potreby dokladovania dokladov a údajov, ktoré už má VS k dispozícii. To má projekt DI dosiahnuť prostredníctvom zavedenia manažmentu údajov vo VS (data governance), implementácie master data managemet pohľadu na dáta VS, riešenia dátovej kvality a zabezpečenia požadovanej úrovne interoperability údajov verejnej správy. Minimálne požiadavky, ktoré má projekt DI naplniť sú navrhnúť efektívnejší administratívny proces dátovej integrácie súvisiaci s uzatváraním dohôd o poskytovaní údajov, dohôd o integračných zámeroch a vypracovať metodiku pre tvorbu, aktualizáciu a poskytovanie údajov pre Dátový slovník 1x a dost, navrhnúť proces a infraštruktúru pre špecifikovanie a následné zdieľanie informácií (Metadát) o štruktúrovaných a neštruktúrovaných dátach (napr. právnych noriem) súvisiacich s realizáciou 1x a dost a navrhnúť koncepčný a logický dátový model pre 1x a dost na úrovni verejnej správy. Projekt bol spustený v auguste 2021 s predpokladaným termínom dokončenia v októbri 2022. V súčasnosti projekt ďalej pokračuje.

Projekt dátovej integrácie sa v oblasti dátovej kvality zameriava na poskytnutie služieb a nástrojov na riadenie dátovej kvality. Riešenie je vystavané v aplikácii Talend, v rámci ktorého budú mať užívatelia možnosť využívať „Talend data stewardship console“.

Prístup k nástrojom na riadenie dátovej bude poskytovaný pre jednotlivé OVM, ktoré v aktuálnej fáze prejavujú záujem o riadenie dátovej kvality, v neskorších fázach sa predpokladá, že toto riešenie bude využívané väčšinou OVM, od ktorých sa po schválení zákona o údajoch bude požadovať riadenie a meranie dátovej kvality.

Pre dátových kurátorov OVM má byť pripravených päť procedúr na riadenie dátovej kvality. Jednotlivé procedúry boli založené na existujúcej Metodike merania dátovej kvality vo verejnej správe z roku 2019 a zároveň boli vytvorené tak, aby plne asistovali dátovým kurátorom pri riadení dátovej kvality v maximálnej možnej miere.

Dátoví kurátori budú mať k dispozícii tieto procedúry:

- DQ 1 Základná analýza dátovej kvality
- DQ 2 Validácia údajov voči OE
- DQ 3 Asistované čistenie údajov

⁸ [Projekt: Dátová integrácia: sprístupnenie údajovej základne VS vrátane otvorených údajov prostredníctvom platformy dátovej integrácie \(gov.sk\)](#)

- DQ 4 Stotožnenie údajov – automatické
- DQ 5 Stotožnenie údajov – manuálne
- DQ 6 Vstupná analýza datasetu pre transformáciu
- DQ 7 Mapovanie na ontologické modely
- DQ 8 Transformácia datasetu do formátu RDF
- DQ 9 Nasadenie transformovaného datasetu vo formáte RDF
- DQ 10 Custom služba DQ podľa špecifických potrieb konkrétneho OVM

DQ1: Základná analýza dátovej kvality

Základná analýza dátovej kvality bude poskytovať dátovým kurátorom informácie a kvalite ich dát pomocou rôznych analýz:

- Tabuľkové analýzy:
 - Analýza obsahu stĺpcov
 - Analýza unikátnych záznamov
- Analýza dát v stĺpcoch
 - Analýza zameraná na vyhľadávanie opakujúcich sa vzorov v stĺpci
 - Analýza dát v stĺpcoch na základe zvolených identifikátorov

Tieto analýzy bude možné vykonať pomocou identifikátorov, ktoré sú buď štandardnými a teda sú zahrnuté v prostredí Talend alebo špeciálnymi, ktoré bude nutné vytvoriť na základe špecifických potrieb dátových kurátorov alebo konzumentov dát. Špeciálne identifikátory by mali vychádzať z biznis pravidiel, z ktorých sa následne vytvoria identifikátory priamo v Talend alebo budú vytvorené a importované do prostredia Talend ako REGEX vo formáte csv.

Štandardné identifikátory v Talend:

- Základné štatistiky
 - Počet riadkov
 - Počet nulových záznamov
 - Celkový počet záznamov
 - Unikátny počet záznamov
 - Duplicitný počet záznamov
 - Počet prázdnych záznamov
- Textové štatistiky
 - Minimálna dĺžka
 - Minimálna dĺžka s nulovým záznamom
 - Minimálna dĺžka s prázdny záznamom
 - Minimálna dĺžka s nulovým a prázdny záznamom
 - Maximálna dĺžka
 - Maximálna dĺžka s nulovým záznamom
 - Maximálna dĺžka s prázdny záznamom
 - Maximálna dĺžka s nulovým a prázdny záznamom
- Štatistiky opakujúcich sa vzorov

- Frekvencia opakujúcich sa vzorov

Po vykovaní procedúry jej výstupom je jeden alebo viacero reportov podľa počtu vybraných analýz, zároveň je možné vytvoriť aj evolučný report záznamu, pre zobrazenie vývoja dátovej kvality pre vybrané identifikátory.

DQ2: Validácia údajov voči OE

Procedúra validácie údajov slúži na kontrolu záznamov voči vopred definovaným štandardom, ktoré sú určené pre jednotlivé stĺpce v rámci špecifického záznamu. Tieto štandardy sú zadané na základe biznis pravidiel, ktoré hovoria o tom ako by mali vyzerat' kvalitné dáta v špecifickom zázname. Definované biznis pravidlá sú následne vo forme REGEX ukladané v rámci Dátového slovníku. Pre zachovanie referenčnej integrity je v rámci dátového slovníku obsiahnutý aj Unikátny referenčný identifikátor (URI).

Aktuálne sa v rámci projektu DI rozširuje dátový slovník o nové objekty evidencie. Ako náhle sa objekt evidencie nachádza v dátovom slovníku, je možné využiť túto procedúru dátovej kvality na validáciu záznamu z daného objektu evidencie. Výstupom tejto procedúry je jeden alebo viacero reportov podľa rozsahu nastavenej validácie, zároveň je možné vytvoriť aj evolučný report záznamu, pre zobrazenie vývoja dátovej kvality voči nastaveným štandardom z dátového slovníku.

DQ3: Asistované čistenie údajov

Cieľom procedúry asistovaného čistenia údajov je poskytnúť dátovým kurátorom nástroj na prečisťovanie dátovej nekvality, ktorá bola zistená v rámci dvoch predchádzajúcich procedúr: Základná analýza dátovej kvality (DQ 1), Validácia údajov voči OE (DQ 2).

Procedúra Asistovaného čistenia kurátorom poskytuje kampaň, ktorú je možné využiť na riadenie dátovej kvality v prostredí Talend data Stewardship konzoly – **Kampaň na čistenie dát**. V rámci tejto kampane je úlohou dátového kurátora rozhodnúť a správnosti resp. chybe v zázname a upraviť ho do požadovaného formátu pre špecifické pole. Výstupom z kampane je vždy dátový súbor vo formáte CSV, v ktorom sa nachádzajú nové dáta s vykonanými úpravami uskutočnenými v kampani asistovaného čistenia údajov.

Okrem tejto základnej kampane je možné vytvoriť aj ďalšie kampane, ktoré je možné zostrojiť podľa špecifického prípadu použitia.

DQ4: Stotožnenie údajov – automatické

Úlohou procedúry automatického stotožnenia údajov je poskytnúť dátovému kurátorovi nástroj, pomocou ktorého si bude môcť overiť vecnú pravdivosť údajov. Toto stotožnenie bude realizované porovnaním záznamu so záznamom z objektu evidencie. V rámci tejto procedúry dôjde k deduplikácii záznamov ich stotožnením v rámci jedného dátového súboru.

Výstupom z kampane je vždy dátový súbor vo formáte CSV, v ktorom sa nachádzajú nové dáta, ktoré prešli deduplikáciou v rámci stotožnenia.

DQ5: Stotožnenie údajov – manuálne

Procedúra manuálneho stotožnenia údajov má rovnakú úlohu ako automatické stotožnenie údajov, a to overiť vecnú pravdivosť údajov na základe stotožnenia s objektom evidencie. Rozdielom je, že v tomto prípade dátový kurátor vykonáva deduplikáciu manuálne v rámci Talend data Stewardship konzoly.

Výstupom z kampane je vždy dátový súbor vo formáte CSV, v ktorom sa nachádzajú nové dáta, ktoré prešli deduplikáciou v rámci stotožnenia.

Procedúry DQ 6 až DQ 10

V súčasnosti sa pripravuje špecifikácia procedúr DQ 6 až DQ 9, ktoré sú zamerané na služby súvisiace s transformáciou údajov. Procedúry by mali poskytovať nasledujúce služby.

DQ 6 Vstupná analýza datasetu pre transformáciu

Procedúra bude poskytovať analýzu vstupného XML/XSD datasetu z hľadiska jeho biznis povahy, spoločne s analýzou ontológií súvisiacich s dátovým modelom transformovaného datasetu.

DQ 7 Mapovanie na ontologické modely

V rámci tejto procedúry bude vykonané mapovanie štruktúr XML/XSD datasetu na ontologické modely.

DQ 8 Transformácia datasetu do formátu RDF

Procedúra vykoná mapovanie jednotlivých elementov, vrátane zápisu transformácie jazykom XSLT.

DQ 9 - Nasadenie transformovaného datasetu vo formáte RDF

V rámci procedúry DQ 9 bude vykonané:

- nasadenie (novej transformácie) upraveného datasetu na testovacie prostredie,
- testovanie/ladenie transformácie,
- validácia výstupov z testovania (využitie SHACL, ak je k dispozícii) a schválenie transformácie,
- nasadenie transformovaného datasetu vo formáte RDF na produkciu.

DQ 10

V rámci procedúry DQ 10 budú poskytované služby DK podľa špecifických potrieb konkrétnych OVM.

2.5.2 Centrálna integračná platforma (CIP)

Projekt centrálnej integračnej platformy⁹ (CIP) prinesie do IS CSRÚ nástroje na riadenia dátovej kvality v rámci služieb „Platforma ako služba“ (PaaS). Aktuálne ešte nie je jasné aké softvérové riešenie bude vybrané na realizáciu riadenia dátovej kvality, avšak vybrané riešenie bude podľa detailného návrhu riešenia (DNR). V rámci dokumentácie sa uvádza, že služba dátovej kvality bude poskytovať grafické rozhranie pre sprístupnenie a tvorbu reportov dátovej kvality vytvorených dátovým kurátorom v procesoch dátovej kvality - profilácii dát.

Cieľom projektu CIP je ďalší rozvoj platformy dátovej integrácie a riešenie manažmentu osobných údajov. Centrálné služby a funkcie platformy dátovej integrácie (CIP) budú zahŕňať PaaS pre manažment údajov, distribúciu údajov, obslužnú zónu, podporné služby pre poskytovateľov a konzumentov údajov a webový prístup. Výsledkom časti CIP má byť rozšírenie a vylepšenie GUI pre komunikáciu s registrami, vďaka čomu bude princíp „jeden-krát“ a časť služieb CIP prístupná aj pre používateľov s neintegrovanými systémami alebo chýbajúcimi informačnými systémami, GUI umožní používateľsky zmysluplný a plnohodnotný ekvivalent G2G služieb CIP, služba pre priamy „push model“ distribúcie údajov, čo umožní pilotné nasadenie proaktívnych služieb a zrušenie oznamovacej povinnosti, zápisová služba pre komunikáciu zdrojového a referenčného registra ako súčasť zjednodušenia distribúcie a synchronizácie údajov, platformové služby PaaS pre manažment údajov (integračná platforma ako služba a master data management ako služba) – všetky inštitúcie verejnej správy môžu realizovať manažment svojich údajov s využitím týchto nástrojov, vytvorenie rozhrania na Portál otvorených údajov (data.gov.sk) umožňujúca OVM zjednodušiť publikovanie otvorených údajov pomocou CIP, publikovanie a evidencia oprávnení pre získavanie údajov a žiadostí pre získavanie údajov pre zníženie administratívnej náročnosti procesu integrácie nástroje pre obsluhu konzumentov a poskytovateľov údajov.

Z hľadiska merania DK projekt CIP prináša službu PaaS („Platforma ako služba“), v rámci ktorej budú môcť jednotlivé OVM vykonávať meranie DK. Podľa detailného návrhu riešenia má služba dátovej kvality poskytovať grafické rozhranie pre sprístupnenie a tvorbu reportov dátovej kvality vytvorených dátovým kurátorom v procesoch dátovej kvality - profilácii dát. Riešenie by taktiež malo poskytovať nástroje na vykonanie týchto činností v oblasti riadenia dátovej kvality:

- Tvorba reportov dátovej kvality a stanovovanie biznis pravidiel na sledovanie kvality a čistenie dát
- Analýzu a čistenie dát ich štandardizáciou, elimináciou duplícít
- Monitoring a reportovanie stavu a vývoja zmien pri zlepšovaní kvality dát
- Sledovanie dátovej kvality s cez preddefinované alebo vyvinuté reporty dátovej kvality cez riadené GUI prostredie

Súčasťou projektu CIP je aj časť Manažment osobných údajov (MOU), ktorá prináša ako konzument novú sadu požiadaviek na dátovú kvalitu (popísané v kapitole [4.1.4.1.14-2-1-1](#) Moje údaje (MOU)).

Z hľadiska integrácia dát pre konzumenta MOU je táto činnosť realizovaná prostredníctvom IS CSRÚ, kde budú niektoré dáta uložené v perzistentnej vrstve, pre zvyšné dáta bude IS CSRÚ slúžiť ako nástroj na ich prenos od jednotlivých OVM.

⁹ [Projekt: Rozvoj platformy integrácie údajov \(centrálna integračná platforma\) a Manažment osobných údajov \(gov.sk\)](#)

Keďže MOU vyžaduje dáta vo formáte JSON-LD je IS CSRÚ jediným bodom, cez ktorý budú prechádzať dáta pre potreby MOU. Dôvodom za týmto typom integrácie je skutočnosť, že jednotlivé OVM nedisponujú dátami vo formáte JSON-LD a nemajú kapacity na transformáciu dát do tohto formátu. Práve kvôli tomu do dátového toku medzi OVM a MOU vstupuje IS CSRÚ, ktorý poskytuje transformačnú službu štruktúrovaných dát do formátu JSON-LD.

2.5.2.1 O meraní kvality v nástroji Talend

Informačný systém Centrálnej správy referenčných údajov (IS CSRÚ) je súčasťou modulu procesnej integrácie a integrácie údajov, ktorý patrí podľa zákona o e-governmente medzi tzv. „spoločné moduly“. Umožňuje zdieľanie a vymieňanie si dát vo verejnej správe navzájom medzi jednotlivými organizáciami.

Na riadenie kvality dát používa riešenie postavené na platforme Talend. Talend Data Fabric je obľúbená open-source platforma pre dátovú integráciu, ktorá poskytuje sadu nástrojov určených na riadenie dátovej kvality:

- Talend Data Integration: Poskytuje široké možnosti pre extrakciu, transformáciu a načítanie (ETL) dát. Umožňuje integráciu dát z rôznych zdrojov, vrátane relačných databáz, súborových systémov, cloudových služieb a ďalších.
- Talend Data Quality: Komplexný modul, ktorý poskytuje možnosti profilovania, štandardizácie, obohacovania, čistenia a monitorovania na zabezpečenie kvality údajov z rôznych zdrojov. Umožňuje používateľom analyzovať, štandardizovať, čistiť a obohacovať svoje údaje. Môžu v ňom vykonávať rôzne kontroly dátovej kvality údajov, napríklad identifikáciu duplicitných údajov, overenie údajov podľa vopred definovaných pravidiel a zisťovanie anomálií. Umožňuje im definovať vlastné pravidlá dátovej kvality na základe ich špecifických požiadaviek.

Talend Data Integrity and Governance: Pomáha organizáciám riadiť dátovú kvalitu, ochranu osobných údajov a dodržiavanie regulácií. Poskytuje prostriedky na sledovanie a správu prístupov k dátam, nastavovanie politík a zabezpečenie zhody s legislatívou a vnútornými pravidlami. Talend Application and API Integration: Poskytuje jednotnú platformu na vývoj API, integráciu aplikácií a údajov a kvalitu údajov na zvýšenie produktivity tímu a rýchlejšie dodanie riešení na trh. Zahŕňa všetko potrebné na splnenie požiadaviek rozhraní API a udalosťami riadenej architektúry. Organizácie, ktoré v rámci projektu CSRÚ alebo autonómne využívajú platformu Talend môžu vykonávať manuálnu alebo automatizovanú kontrolu dátovej kvality v dvoch prostrediach - Talend Data Steward a Talend Open Studio, ktoré slúžia na rôzne účely a používajú ich rôzne roly v rámci organizácie. Talend Open Studio je nástroj na integráciu údajov, ktorý používajú vývojári na navrhovanie, vývoj, testovanie a nasadzovanie úloh integrácie údajov. Poskytuje grafické používateľské rozhranie na vytváranie pracovných postupov integrácie údajov a podporuje širokú škálu prípadov použitia integrácie údajov, ako je migrácia údajov, synchronizácia údajov, dátová kvalita a správa zdrojových údajov. Open Studio primárne používajú vývojári, dátoví inžinieri a IT profesionáli. Na druhej strane Talend Data Stewardship Console je webová aplikácia, ktorú správcovia údajov používajú na správu a monitorovanie dátovej kvality údajov. Poskytuje správcovi údajov centralizované rozhranie zamerané na úlohy správy údajov, ako je kontrola a validácia údajov, spolupráca a obohacovanie údajov. Používajú ho predovšetkým správcovia údajov, analytici údajov a biznis používatelia, u ktorých sa nepredpokladajú rozsiahle vývojárske zručnosti.

Z hľadiska automatizovaného (lepšie však skôr strojového) merania kvality poskytuje Talend niekoľko možností vytvárania reportov, ktoré možno použiť na vytváranie správ o kvalite údajov. Reportovací nástroj v Talende je súčasťou Talend Data Preparation a umožňuje používateľom vytvárať, upravovať a zdieľať reporty. Je to cloudový nástroj. Tu sú niektoré z možností:

- 1 Šablóny reportov: Talend poskytuje niekoľko vopred vytvorených šablón, ktoré možno použiť na generovanie správ o kvalite údajov. Tieto šablóny zahŕňajú správy o úplnosti, presnosti, konzistentnosti a jedinečnosti údajov. Tieto šablóny možno prispôbiť pridaním alebo odstránením metrík a výberom zdrojov údajov a filtrov.
- 2 Vlastné reporty: Talend tiež umožňuje používateľom vytvárať vlastné reporty o kvalite údajov. Používatelia si môžu vybrať metriky, pre ktoré chcú vytvárať prehľady, ako aj zdroje údajov a filtre. Používatelia môžu tiež vytvárať vlastné vizualizácie, ako sú tabuľky a grafy, na zobrazenie metrík kvality údajov.
- 3 Zdieľanie reportov: Talend umožňuje používateľom zdieľať správy s inými používateľmi alebo zainteresovanými stranami. Správy možno zdieľať prostredníctvom e-mailu, vkladať na webové stránky alebo exportovať ako súbory PDF alebo Excel. To umožňuje používateľom spolupracovať v oblasti kvality údajov a zdieľať poznatky.
- 4 Plánovanie reportingu: Talend umožňuje používateľom naplánovať automatické generovanie správ v určitých intervaloch. To môže ušetriť čas a zabezpečiť pravidelné generovanie prehľadov. Používatelia si môžu zvoliť frekvenciu a načasovanie správ, ako aj príjemcov.
- 5 Integrácia s inými nástrojmi: Talend je možné integrovať s inými nástrojmi na vytváranie prehľadov, ako sú Tableau alebo PowerBI. Používatelia tak môžu využívať možnosti tvorby prehľadov týchto nástrojov a zároveň používať Talend na integráciu údajov a kvalitu údajov.

Hoci Talend poskytuje komplexnú sadu nástrojov a modulov na meranie kvality údajov, automatické merania kvality údajov majú určité obmedzenia. Tu sú niektoré možné obmedzenia:

- Nepresné profilovanie údajov: Funkcia profilovania údajov spoločnosti Talend sa spolieha na štatistické algoritmy na identifikáciu vzorov a anomálií v údajoch. Tieto algoritmy však nemusia vždy presne identifikovať všetky problémy, najmä ak sú údaje zložité alebo zašumené.
- Obmedzené chápanie sémantiky: Automatické merania kvality údajov môžu byť obmedzené nedostatočným pochopením kontextu a sémantiky údajov. Napríklad pravidlá kvality údajov možno bude potrebné prispôbiť na základe špecifických biznis požiadaviek alebo znalostí domény. Vyžaduje to teda ľudský zásah.
- Neúplné kontroly kvality údajov: Hoci Talend poskytuje komplexný súbor kontrol kvality údajov, nemusí pokrývať všetky možné problémy s kvalitou údajov. Niektoré problémy s kvalitou údajov môžu vyžadovať vlastné kontroly alebo manuálnu kontrolu.
- Ťažkosti so spracovaním neštruktúrovaných údajov: Moduly kvality údajov Talendu môžu mať problémy so spracovaním neštruktúrovaných údajov, ako sú text alebo obrázky. Na meranie kvality údajov pre neštruktúrované údaje môžu byť potrebné ďalšie nástroje alebo prispôbenie nástroja.
- Závislosť od zdrojov údajov: Merania kvality údajov Talendom závisia od kvality samotných zdrojov údajov. Ak sú zdroje údajov neúplné alebo nepresné, môže to ovplyvniť kvalitu automatizovaných meraní.

Nasleduje príklad postupu, ktorým vytvoríme report z merania kompletnosti dát na určenej databáze, najprv uvádzame jeho popis, potom príklad zodpovedajúceho scriptu. Upozorňujeme však, že sa jedná len o ilustračný príklad.

- 1 Vytvoriť nový job v Talende a pridať vstupný komponent tInput, ktorý prečíta data zo spracovávanej databázy
- 2 Pridať komponent tAggregateRow a nakonfigurovať ho, aby spočítal počet záznamov v databáze
- 3 Pridať komponent tJavaRow a použiť ho na výpočet kompletnosti dát. Napríklad sa dá spočítať percento záznamov, ktoré majú vyplnené všetky povinné polia.
- 4 Pridať komponent tFileOutputDelimited a zapísať ním výpočet kompletnosti do CSV súboru
- 5 Vytvoriť nový vzor reportu v Talend Data Preparation a pridať nový zdroj dát, ktorý smeruje na CSV súbor vytvorený v predošlom kroku
- 6 Pridať novú vizualizáciu z vzoru reportu, napríklad stĺpcový alebo koláčový diagram a nakonfigurovať ho na zobrazenie výpočtu kompletnosti.
- 7 Upraviť vzor reportu podľa špecifických požiadaviek alebo zvyklostí napr. pridaním nadpisov, hlavičiek a pätičiek apod.
- 8 Uložiť vzor reportu a naplánovať automatické generovanie reportov v pravidelných intervaloch

2.5.3 Konsolidovaná analytická vrstva (KAV)

Projekt KAV je financovaný z OP II a jeho realizátorom je MIRRI. Cieľom projektu je podporiť rozhodovanie na základe využívania dát („data-driven-state“) a využívať tieto údaje pre zlepšenie fungovania inštitúcií verejnej správy. To vyžaduje zabezpečenie kvalitných dát pre potreby analytických jednotiek, vytvorenie základnej dátovej infraštruktúry pre podporu analytického spracovania údajov vo verejnej správe a vytvorenie transparentného logovania všetkých prístupov k údajom. KAV by mala poskytnúť analytické výstupy nielen pre podporu strategického riadenia štátu a tvorbu politik, ale aj pre operatívne rozhodovanie a určenie „hodnoty za peniaze“. Predpokladom je, že KAV poskytne inštitúciám verejnej správy prístup k potrebným údajom a nástroje na ich analýzu, vrátane matematických a štatistických nástrojov, nástrojov BI a ETL, podporu strojového učenia (machine learning) a samozrejme nástroje na vizualizáciu a publikáciu výstupov.

Stav realizácie projektu ani integrácia dátových tokov pre KAV nie je v súčasnosti vyjasnená. Preto nie je v súčasnosti možné presne identifikovať akým spôsobom budú integrované dátové toky medzi OVM a KAV. Avšak dá sa predpokladať, že časť dátových tokov bude integrovaná prostredníctvom IS CSRÚ, zvyšok bude integrovaný priamym prepojením s OVM cez API, alebo cez iné nešpecifikované riešenie. **Dopad situácie na meranie dátovej kvality je preto zatiaľ nejasný** a až do vyjasnenia v ďalšom nebude uvažovaný.

2.5.4 Životné situácie (ŽS)

Projekty Životné situácie 1 a 2 a Životné situácie 3 a 4 sú financované z Fondu obnovy a odolnosti a ich realizátorom je tiež MIRRI. Hlavným zámerom oboch projektov je zjednodušiť občanom a podnikateľom elektronickú komunikáciu so štátom pri riešení životných situácií. V rozsahu projektu je zahrnutých 52 formulárov pre životné situácie v kategóriách Strata zamestnania, Kúpa a vlastníctvo nehnuteľnosti na bývanie, Začatie

podnikania a Kúpa auta. V rámci projektu už boli špecifikované množiny údajov pre jednotlivé formuláre, ale zatiaľ nebolo určené vlastníctvo formulárov so zodpovednosťou za ich publikovanie. Preto ani **zatiaľ neboli definované požiadavky na dátovú kvalitu** pre podkladové datasety potrebné pre vyplnenie formulárov. Avšak niektoré požiadavky na DK vyplývajú aj priamo z jednotného dizajnu manuálu elektronických služieb Slovenska „ID-SK“¹⁰. Podľa dostupných informácií sú však tri zo štyroch v súčasnosti realizovaných životných situácií len v počiatočných fázach riešenia.

¹⁰ [ID-SK Frontend \(gov.sk\)](https://gov.sk/)

3 Návrh úprav a aktualizácia štandardu pre meranie dátovej kvality

3.1 Návrh na úpravu odporúčaní

V rámci Metodiky bola navrhnutá skupina odporúčaní na zlepšenie dátovej kvality verejnej správy na Slovensku. Navrhované odporúčania sa zameriavali na opatrenia, ktoré je potrebné realizovať na centrálnej alebo na lokálnej úrovni. Tieto odporúčania boli síce čiastočne implementované v praxi, ale ich aplikácia je nedostatočná a od roku 2019 došlo aj k zmene priorít, ktoré boli v odporúčaní zahrnuté. Táto kapitola na základe vyhodnotenia odporúčaní z Metodiky (kapitola 2.) navrhuje ich zmeny.

3.1.1 Dátové štandardy

Odporúčanie 1 (lokálna úroveň) – Zavedenie jasných biznis pravidiel pre riadenie dátovej kvality vrátane väzby biznis pravidiel na legislatívu upravujúcu dotknuté údaje.

Toto odporúčanie je naďalej platné. Je čiastočne implementované v praxi a má jasne definované a naplánované ďalšie aktivity. Pokiaľ vznikne na lokálnej úrovni potreba definovať nové biznis pravidlá mali by byť zahrnuté do centrálného zoznamu biznis pravidiel publikovaného na datalab.digital.

Odporúčanie 2 (lokálna úroveň) – Implementácia biznis pravidiel a 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.

Toto odporúčanie je naďalej platné. Je čiastočne implementované v praxi, ale nemá jasne definované a naplánované ďalšie aktivity. Odporúčame pripraviť príslušné plány a zabezpečiť ich implementáciu tak, aby boli definované súvisiace KPI a vykonávali sa pravidelné merania kvality dát vrátane podávania správ o výsledkoch tohto merania.

Odporúčanie 3 (lokálna úroveň) – Definovanie presného formátu pre každý zadávaný údaj, aby bola zabezpečená konzistencia všetkých zdrojov, ktoré sú v organizácii využívané.

Toto odporúčanie je naďalej platné. Je čiastočne implementované v praxi a má jasne definované ďalšie aktivity. Dôraz by mal byť naďalej kladený na zabezpečenie konzistencie údajov priamo pri ich vzniku.

Odporúčanie 4 (centrálna úroveň) – Vypracovanie jednotného dátového slovníka pre verejnú správu ako komponent centrálného dátového modelu.

Odporúčanie 5 (centrálna úroveň) – Vytvorenie centrálného dátového modelu verejnej správy a definovanie kompetencií a organizačného zabezpečenia dátového modelu.

Obe tieto odporúčania ostávajú naďalej platné, ale mali by byť spojené do jedného odporúčania. Sú čiastočne implementované v praxi a majú jasne definované ďalšie aktivity.

Odporúčanie 6 (lokálna úroveň) – 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.

Toto opatrenie by malo byť preformulované, naplánované a implementované tak, aby sa zabezpečila integrácia s CMÚ, prioritne pre referenčné registre (RPO a jeho zdrojové registre, RFO). Je možné zvážiť prevzatie realizácie tohto opatrenia pre referenčné registre pod zvýšený dohľad resp. priamo pod gesciu centrálnej úrovne t. j. správcu NSZÚ (Dátová kancelária).

Odporúčanie 7 (centrálna úroveň) – 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.

Odporúčanie je naďalej platné. Je potrebné naplánovať, ako sa zabezpečí referencovanie a stotožňovanie údajov medzi registrami a zabezpečiť jeho implementáciu (finančné, organizačné, technologické aspekty a aspekty, zamerané na riadenie zmien – change management).

Odporúčanie 8 (lokálna úroveň) – Zabezpečenie referencovania získavaných údajov organizáciou alebo údajov vedených v databázach organizácií na existujúce referenčné registre.

Toto odporúčanie je naďalej platné a čiastočne implementované v praxi.

Odporúčanie 9 (lokálna úroveň) – Zabezpečenie súladu súčasného stavu s existujúcimi štandardami dátovej kvality.

Odporúčanie je naďalej platné a Dátová kancelária by mala stanoviť a realizovať aktivity, ktoré podpora využívanie metodiky na strane OVM.

Odporúčanie 10 (centrálna úroveň) – Definovanie oblastí dátovej kvality pre zabezpečenie líderstva v tejto oblasti na úrovni EU.

Toto odporúčanie je vhodné nahradiť odporúčaním, aby sa Slovensko inšpirovalo riešeniami a metodickým smerovaním ostatných členských štátov EÚ. Jednou z ciest je posudzovať pri príprave projektov aj zahraničné riešenia a vziať ich do úvahy ako alternatívu. Súčasťou každej pripravovanej štúdie by mala byť aj povinná časť, ktorá popisuje, ako bolo riešenie realizované v iných členských štátoch s dopadom na SWOT analýzu. Pokiaľ sa žiadateľ aj napriek už realizovaným riešeniam rozhodne postupovať iným spôsobom, mal by byť povinný to zdôvodniť.

Odporúčanie 11 (lokálna úroveň) – 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.

Odporúčanie je naďalej platné a Dátová kancelária by mala stanoviť a realizovať aktivity, ktoré podporia využívanie metodiky na strane OVM.

Odporúčanie 12 (centrálna úroveň) – 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.

Odporúčanie je naďalej platné. Je čiastočne implementované v praxi a má jasne definované ďalšie aktivity. Na podporu zlepšenia riadenia dátovej kvality môže MIRRI naďalej využívať pozitívnu motiváciu prostredníctvom dotačných výziev.

3.1.2 Organizácia

Opatrenie 13 (centrálna úroveň) – Definovanie jasných kompetencií, právomocí a zodpovednosti pre pozície dátových kurátorov.

Bolo by vhodné opatrenie preformulovať alebo definovať nové opatrenie týkajúce sa štandardizácie procesov dátovej kvality OVM a všetkých rolí (napr. dátový architekt, analytik a p.), ktoré majú byť do procesov zapojené. Zároveň je nutné viac úsilia venovať riadeniu strategickej zmeny nad rámec školení, aby sa navrhované opatrenia dostali aj do reálnej praxe.

Odporúčanie 14 (lokálna úroveň) – Posilnenie kapacít pre organizačné zabezpečenie riadenia dátovej kvality vo forme dátových kurátorov.

Odporúčanie je naďalej platné, ale posilnenie organizačného zabezpečenia by sa nemalo vzťahovať len na dátových kurátorov, ale aj na ďalšie roly v procesoch riadenia a merania dátovej kvality naprieč OVM. Bolo by tiež vhodné rozšíriť poskytovanie školení a metodickej podpory pre už existujúcich dátových kurátorov. Zároveň je nutné viac úsilia venovať riadeniu strategickej zmeny nad rámec školení, aby sa navrhované opatrenia dostali aj do reálnej praxe.

Opatrenie 15 (centrálna úroveň) – Zabezpečenie školení pre oblasť riadenia a správy dátovej kvality v rezorte a vybudovanie „call centra“ pre zabezpečenie poradenstva v oblasti dátovej kvality.

Dátová kancelária by mala zlepšiť komunikáciu a spoluprácu s OVM delegovaním komunikačných manažérov pre jednotlivé OVM, ktorí budú mať prehľad o tom, aké projekty na OVM prebiehajú. Opatrenie je naďalej platné, ale nakoľko Dátová kancelária nezvažuje vybudovanie call-centra, opatrenie by malo byť preformulované nasledovne: „Zlepšenie komunikácie s OVM a zabezpečenie školení pre oblasť riadenia a správy dátovej kvality v rezorte“.

Odporúčanie 16 (centrálna úroveň) – Posilnenie kompetencie Dátovej kancelárie v oblasti dátovej kvality tak, aby v súčasnosti existujúce pravidlá vedela kontrolovať a podmieňovať realizáciu projektov napr. z dopytových výziev vyvodzovať prípadné konsekvencie.

Opatrenie je naďalej platné, ale hoci Dátová kancelária má už dnes kompetencie kontroly v oblasti dátovej kvality, vrátane dodržiavania publikačného minima, nemá žiadne kompetencie z hľadiska vyvodzovania následkov. Keďže Dátová kancelária potvrdila, že udeľovanie sankcií za nedodržiavanie povinností nepovažuje za efektívne, je potrebné formulovať stratégiu, ako motivovať OVM na dodržiavanie svojich povinností.

Odporúčanie 17 (lokálna úroveň) – 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.

Toto opatrenie nebolo implementované, ale je naďalej dôležité a Dátová kancelária by mala zvážiť kroky vedúce k jeho dôslednej implementácii. Je potrebné naplánovať, ako sa zabezpečí referencovanie a stotožňovanie údajov medzi registrami a zabezpečiť jeho implementáciu (finančné, organizačné, technologické aspekty a aspekty, zamerané na riadenie zmien – change management).

Odporúčanie 18 (centrálna úroveň) – Nastavenie pravidiel pre opravy identifikovaných nezrovnalostí v záznamoch tak, aby bol proces čo najkratší a aby bol realizovateľný v čase vzniku alebo odhalenia konzistentnosti v údajoch (viď. opravy záznamov v obchodnom registri, ktoré sú zdrojom pre RPO – chyba sa zistí v RPO).

Opatrenie je naďalej platné. Je potrebné metodicky posúdiť aktuálne nastavené procesy opráv v prioritných registroch (napríklad RPO a Katastrálny portál UGKK), navrhnúť ich zmeny a zabezpečiť ich implementáciu (finančné, organizačné, technologické aspekty a aspekty, zamerané na riadenie zmien – change management). V konkrétnych prípadoch je nutné implementovať najvhodnejší spôsob na zvýšenie dátovej kvality registra, či už zmenou legislatívy alebo napríklad opakovanými kampaňami „čistenia dát“, ako to urobil správca Registra adries.

Odporúčanie 19 (centrálne úroveň) – Jasné nastavenie pravidiel hodnotenia a odmeňovania pre zodpovedné osoby za dátovú kvalitu reflektujúcu 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“.

Toto odporúčanie je vhodné zrušiť alebo preformulovať a zamerať sa na konkrétne aktivity pozitívne, nie negatívne politiky.

Odporúčanie 20 (lokálna úroveň) – Implementovanie zmeny v oblasti dátovej kultúry a interpretácie údajov, ktoré má organizácia vo „vlastníctve“ vrátane zavedenia biznis metadát.

Odporúčanie je naďalej platné, ale v porovnaní s inými odporúčaniami, ktoré smerujú k vyriešeniu akútnych a kľúčových problémov, je jeho priorita nízka. Na zlepšenie dátovej kultúry a interpretácie údajov by sa malo Slovensko zamerať po vyriešení väčších identifikovaných problémov.

3.1.3 Proces

Odporúčanie 21 (centrálne úroveň) – Definovanie stratégie pre dátovú kvalitu (vrátane definovania jasných KPIs, ktoré budú následne prijaté organizáciami) a neustále monitorovanie jej zavedenia v praxi a v pravidelných intervaloch vyhodnocovanie a aktualizácia.

Toto odporúčanie je potrebné upraviť tak, aby sa zameriavalo na monitorovanie implementácie a aktualizáciu stratégie pre dátovú kvalitu. Zároveň je potrebné posilniť zabezpečenie nielen implementácie ale aj udržateľnosti zavedených zmien (aspekty zamerané na riadenie zmien – change management, teda finančné, organizačné, technologické aspekty implementácie a udržateľnosti).

Odporúčanie 22 (centrálne úroveň) – Zavedenie povinnosti vykonávania pravidelného merania a vyhodnocovania dátovej kvality so zverejňovaním výsledkov vo forme open data.

Odporúčanie je naďalej platné a zatiaľ nebolo implementované, povinnosť vykonávať pravidelné merania a vyhodnocovanie DK zatiaľ nie je zavedená. Dátová kancelária by mala túto povinnosť zapracovať do pripravovaného Zákona o údajoch.

Odporúčanie 23 (lokálna úroveň) – 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).

Toto odporúčanie nebolo implementované v praxi a v súčasnosti nie je považované za prioritu. Po zavedení pravidelných meraní dátovej kvality a jej vyhodnocovaní by však

mal byť implementovaný aj tento krok s prihliadnutím na procesy jednotlivých OVM a metodiky ich riadenia dátovej kvality.

Odporúčanie 24 (lokálna úroveň) – Jasné definovanie povinných a nepovinných údajov pri jednotlivých záznamoch tak, aby bolo pri hodnotení údajov vidieť, aká je štatistika pre povinné aj pre nepovinné údaje.

Toto odporúčanie je naďalej potrebné. Dátová kancelária by mala prehodnotiť postupy ako plnenie odporúčania monitorovať, vyhodnocovať jeho presadenie do praxe a na základe toho prijať rozhodnutie o ponechaní odporúčania alebo o jeho zrušení.

Odporúčanie 25 (lokálna úroveň) – Zabezpečenie lepšej osvetly pre zadávanie nepovinných údajov a vysvetlenie pridanej hodnoty zadania nepovinného údaja do systému.

Toto odporúčanie zatiaľ nebolo implementované v praxi a je naďalej platné. OVM by mali jednotlivo posúdiť účel nepovinných údajov a dopad ich nevyplnenosti a na základe toho zvážiť povinnosť ich vyplnenia alebo zvýšenie osvetly smerovanej na zdroje údajov.

Odporúčanie 26 (lokálna úroveň) – Zavedenie jasných pravidiel pre aktualizáciu záznamov a ich atribútov, ako aj implementácia procesov monitoringu a dodržiavania pravidiel.

Odporúčanie je naďalej platné a z centrálnej úrovne by mali byť podniknuté kroky na podporu jeho dodržiavania v súlade s odporúčaním 18.

Odporúčanie 27 (lokálna úroveň) – Nastavenie jednoduchých procedúr pre transformovanie údajov do foriem využiteľných pre realizáciu procesov podporujúcich dátovú kvalitu alebo procesov tvorby reportov a analýz.

Odporúčanie by malo byť súčasťou implementácie odporúčaní zameraných na meranie stavu DK a jeho reportovanie a nie je nutné ho evidovať samostatne. Opatrenie by potom mohlo byť zrušené.

Odporúčanie 28 (lokálna úroveň) – Nastavenie procesov prípravy údajov na analytické spracovanie (forma, štruktúra potrebná pre analýzy systémoch ako napr. R, SAS, Python, ...).

Odporúčanie je naďalej platné, ale jeho implementácia nemusí byť vždy potrebná. Posúdenie tejto potreby je v kompetencii jednotlivých OVM a ich analytických útvarov prípadne ako súčasť projektu KAV.

Odporúčanie 29 (lokálna úroveň) – 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).

Odporúčanie by malo byť súčasťou implementácie odporúčaní zameraných na meranie stavu DK a jeho reportovanie a nie je nutné ho evidovať samostatne. Opatrenie by mohlo byť potom zrušené.

3.1.4 Technológie

Odporúčanie 30 (centrálna úroveň) – Zabezpečenie príslušnej technologickej podpory pre riadenie kvality údajov pre všetky definované oblasti vychádzajúce zo stratégie riadenia kvality údajov.

Odporúčanie je čiastočne implementované v praxi a je naďalej platné. Dátová kancelária sa musí rozhodnúť, či jediným nástrojom na meranie DK bude naďalej Talend alebo zväziť aj nástroje na iných platformách a je potrebné vyriešiť prístup OVM do Talendu alebo iného nástroja, aby nedostupnosť finančných prostriedkov na licenčné pokrytie nebránila OVM v prístupe.

Odporúčanie 31 (lokálna úroveň) – Implementácia a zabezpečenie nástroja na riadenie dátovej kvality (v súčasnosti existuje platforma Talend, ktorá je zabezpečená na úrovni PaaS služby v štátnom cloude).

Odporúčanie je síce naďalej platné, ale malo by byť smerované na centrálnu úroveň alebo preformulované ako povinnosť využívať centrálnu riešenie alebo implementovať vlastné na lokálnej úrovni. Je na Dátovej kancelárii zväziť, či popri poskytnutí centrálného nástroja, ktorý vynucuje využívanie centrálnu stanovených štandardov, umožní implementácie vlastných nástrojov na lokálnej úrovni.

Odporúčanie 32 (lokálna úroveň) – Nastavenie technologických pravidiel pre prácu s databázami (jasne definované pravidlá správy údajov, ktoré sa dajú podporiť existujúcimi technológiami).

Odporúčanie je naďalej platné. Dátová kancelária by mala vytvoriť metodiku, ktorá by zahŕňala aj odporúčania pre tieto nastavenia a vyzvala OVM na potvrdenie ich implementácie.

Odporúčanie 33 (lokálna úroveň) – Aplikovanie systémov AI pre technologické riadenie kvality údajov.

Odporúčanie je naďalej platné, ale zatiaľ nebolo implementované v praxi. Využitie AI by malo byť zvažované jednotlivo podľa potrieb a dostupných zdrojov jednotlivých OVM.

Odporúčanie 34 (lokálna úroveň) – Nastavenie procesov profilovania a analyzovania údajov priamo v technologickom riešení.

Odporúčanie je naďalej platné. Dátová kancelária by mala vytvoriť metodiku, ktorá by zahŕňala aj odporúčania pre tieto nastavenia a vyzvala OVM na potvrdenie ich implementácie.

Odporúčanie 35 (lokálna úroveň) – Nastavenie proaktívnych kontrol pre zdrojové systémy – kontrola bude prebiehať už pri vstupe údajov na základe definovaných technologických a procesných požiadaviek.

Odporúčanie je naďalej platné. Dátová kancelária by mala vytvoriť metodiku, ktorá by zahŕňala aj odporúčania pre tieto nastavenia a vyzvala OVM na potvrdenie ich implementácie.

Odporúčanie 36 (lokálna úroveň) – Zabezpečenie technologickej podpory pre definované dátové pravidlá pri vstupe údajov do systému.

Odporúčanie je naďalej platné.

Odporúčanie 37 (lokálna úroveň) – 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.

Odporúčanie je naďalej platné. Dátová kancelária by mala vytvoriť metodiku, ktorá by zahŕňala aj odporúčania pre tieto nastavenia a vyzvala OVM na potvrdenie ich implementácie.

Odporúčanie 38 (lokálna úroveň) – Implementácia nástroja / algoritmov na definovanie duplicity záznamov pri vstupe údajov do systému alebo na podporu procesu deduplikácie záznamov.

Odporúčanie je naďalej platné. Dátová kancelária by mala vytvoriť metodiku, ktorá by zahŕňala aj odporúčania pre tieto nastavenia a vyzvala OVM na potvrdenie ich implementácie.

Odporúčanie 39 (lokálna úroveň) – Identifikácia možností a následného zabezpečenia technologickej podpory pre elimináciu faktora „zlyhanie ľudského faktora“ vo všetkých procesoch a fázach práce s údajmi.

Odporúčanie je naďalej platné. Dátová kancelária by mala vytvoriť metodiku, ktorá by zahŕňala aj odporúčania pre tieto nastavenia a vyzvala OVM na potvrdenie ich implementácie.

3.2 **Nové odporúčania**

Odporúčania z roku 2019 boli všeobecné a zaoberali sa tým, ako zvýšiť kvalitu údajov a zlepšiť jej meranie. Z pohovorov s Dátovou kanceláriou ale vyplynula aj potreba nových odporúčaní, ktorých zavedenie by viedlo k zlepšeniam a na ktoré je takisto potrebné sa zamerať. Niektoré z nich boli síce formulované ako súčasť navrhovanej aktualizácie už existujúcich odporúčaní, ale uvádzame ich aj na tomto mieste pre prehľadnosť.

1 Zvýšiť počet OVM integrovaných na CMÚ

V záujme interoperability by mali byť všetky modely údajov a pravidlá 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 ďalej by mali pribúdať ďalšie registre. Prioritne by nimi mali byť referenčné registre (RPO a jeho zdrojové registre, RFO). Je možné zvážiť prevzatie realizácie tohto opatrenia pre referenčné registre pod zvýšený dohľad, resp. priamo pod gesciu centrálnej úrovne, správcu NSZÚ (Dátová kancelária), t. zn. 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 byť 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.

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. Odporúčania k reportovaniu sú uvedené v kapitole 6 tohto dokumentu.

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 2020¹¹ 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 2021¹². 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 nezverejňujú 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 centrálny 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 politik naprieč OVM, ale aj ich vhodnosť pre každé OVM a ich dodržiavanie. Viac k tejto téme bude vo výstupe 1.2.1. Opatrenia pre zvýšenie dátovej kvality.

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 neprávym 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 efektivitu 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

¹¹ [Detail uznesenia | Portal OV \(gov.sk\)](#)

¹² [Metodicke-usmernenie-009417-2021-oSBAA-1-v3.pdf \(gov.sk\)](#)

dôkladnú plánovaciú a komunikačnú stratégiu. Ak sa však správne uchopí a dôsledne realizuje, môže významne prispieť k zvýšeniu dátovej kvality vo verejnej správe. Hoci sa tejto téme bude venovať aj výstup 1.2.1, tu je niekoľko kľúčových úloh riadenia zmien pri zvyšovaní dátovej kvality vo verejnej správe:

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 priorití.

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 Návrhy na zlepšenie merania kvality dát

Zlepšenie merania dátovej kvality je dôležité pre zlepšenie kvality údajov, ktorá je nevyhnutná, aby mohla organizácia robiť rozhodnutia na základe spoľahlivých a relevantných údajov alebo v prípade poskytovateľov údajov, aby mohli poskytovať spoľahlivé referenčné údaje pre svojich konzumentov. Základnou podmienkou zlepšenia merania DK je organizačná podpora a zapojenie kľúčových hráčov (stakeholderov), najmä konzumentov údajov. Zlepšovanie merania DK je priamo podmienené tým, akú dôležitosť prikladajú stakeholderi dátovej kvalite. Mali by v prípade potreby predefinovať alebo modifikovať ciele, ich dátové potreby a ich dopad na ciele merania. Vo všeobecnosti je ale možné zlepšovať meranie DK nasledovne:

- Pravidelne prehodnocovať biznis pravidlá, KPI a ich prahové hodnoty a nevnímať ich ako statickú množinu nemenných nastavení.
- Vykonávať meranie opakovane a pravidelne a sledovať zmeny v kvalite údajov, ktoré poukazujú na oblasti potenciálneho zlepšenia.
- Vzdelávať a školiť zamestnancov, aby pochopili dôležitosť dátovej kvality a zoznámili sa s odporúčanými postupmi na jej meranie. Vysvetliť im dôležitosť ich roly a zodpovedností, ktoré so sebou prináša.
- Zabezpečiť konzistentnosť údajov a meraní, najmä ak je meranie vykonávané nad rôznymi zdrojmi údajov.
- Využívať automatizované nástroje, ktoré zrýchľujú meranie a eliminujú pravdepodobnosť chýb spôsobených ľudským faktorom.

V priebehu prípravy tohto dokumentu bola vykonaná revízia odporúčaní z Metodiky merania DK (kapitola [02.2.1](#) ▲

Formátov

Formátov
Automati

~~Podrobnejšie vyhodnotenie odporúčaní~~~~Podrobnejšie vyhodnotenie odporúčaní~~) a v rámci diskusií s Dátovou kanceláriou boli navrhnuté aj ďalšie konkrétne odporúčania na zvýšenie dátovej kvality vo verejnej správe na Slovensku. V tabuľke nižšie je uvedený súhrn odporúčaní, ktoré sa **špecificky viažu k zlepšeniu merania dátovej kvality**.

Tabuľka 6 Pôvodné a nové odporúčania pre meranie DK

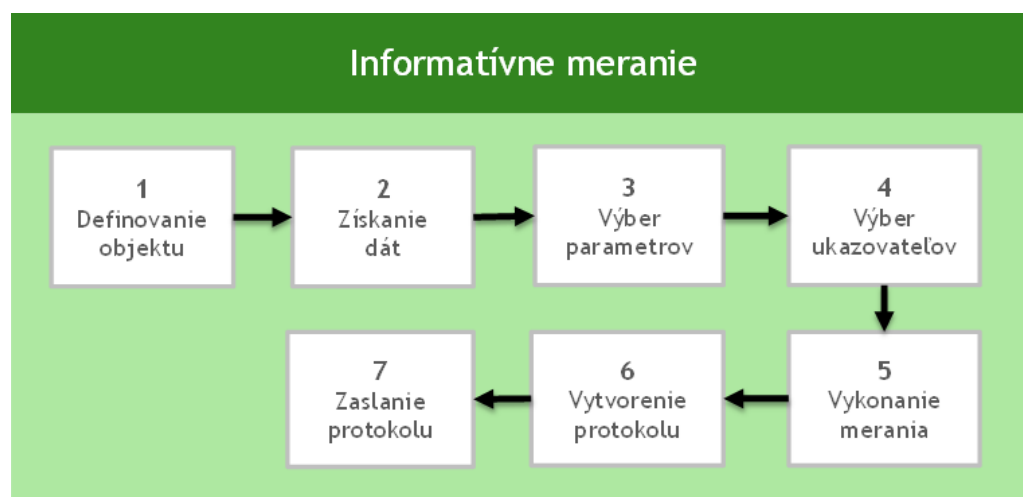
Odporúčania pre meranie DK	
Odporúčania z Metodiky merania DK	<p>Odporúčanie 2 – Implementácia biznis pravidiel a pravidiel pre meranie dátovej kvality (vrátane nastavenia jednoznačných KPI a ich prahových hodnôt ako aj stanovenie, pre ktoré atribúty / záznamy je potrebné merať dátovú kvalitu) v organizáciách a nastavenie ich riadenia. (odporúčanie pre OVM)</p> <p>Odporúčanie 22 – Zavedenie povinnosti vykonávania pravidelného merania a vyhodnocovania dátovej kvality so zverejňovaním výsledkov vo forme open data. (odporúčanie pre Dátovú kanceláriu)</p> <p>Odporúčanie 29 – 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). (odporúčanie pre OVM)</p>
Nové odporúčania	<p>Zaviesť vynútiteľnú povinnosť pravidelne merať DK a jednotným spôsobom zverejňovať výsledky meraní (odporúčanie pre Dátovú kanceláriu)</p> <p>Zverejňovať pravidelné reporty dátovej kvality jednotným spôsobom (odporúčanie pre OVM)</p> <p>Zvýšiť osvetu, ktorá podporí dodržiavanie štandardov DK na úrovni OVM (odporúčanie pre Dátovú kanceláriu)</p>

4 Definícia a spôsob merania pre kľúčové výkonnostné indikátory

Postup definovania kľúčových parametrov a opis spôsobu ich merania popíšeme v tejto kapitole na príkladoch dvoch referenčných registrov, a to Registra právnických osôb (RPO) a Registra adries (RA), jednak preto, že považujeme opis konkrétnych prípadov za názornejší ako popis generického postupu a tiež preto, lebo budeme vedieť poukázať na to, ako sme riešili jednotlivé praktické situácie. Čitateľ/ka potom bude ľahko vedieť odvodiť postup aplikovateľný v jej/jeho situácii.

Vzhľadom na to, že informačný systém Centrálnej správy referenčných údajov (IS CSRÚ), ktorý je súčasťou modulu procesnej integrácie a integrácie údajov používa na riadenie kvality dát riešenie postavené na platforme Talend, bolo dohodnuté, že tento nástroj bude uvažovaný aj pre účely tohto výstupu.

Použitie 7-krokového procesu pre informatívne meranie dátovej kvality podľa Metodiky bolo vcelku priamočiarym rozhodnutím aj s ohľadom na kontinuitu s meraním z r. 2020 (pozri kap. [2.4.12-3-1](#)). Proces je vizualizovaný na nasledovnom diagrame.



Obrázok 2 7-krový proces pre informatívne meranie dátovej kvality

4.1 Príprava na meranie kvality

Prípravná fáza sa skladá z prvých štyroch krokov procesu: definovanie objektu, získanie dát, výber parametrov a výber ukazovateľov.

4.1.1 Definovanie objektu

Okrem využitia postupov z Metodiky boli pri definovaní objektu merania posudzované aj faktory ako prínos merania, dostupnosť údajov, prácnosť prípravy datasetu pre meranie a možnosť opakovaného merania, ako i čas, ktorý bol k dispozícii na realizáciu. V prvom kroku boli uvažované všetky dostupné datasety verejnej správy. Za účelom maximalizácie prínosu merania bol zoznam zúžený na datasety, ktoré slúžia ako zdroje pre niektorého z centrálnych konzumentov (centrálne projekty riadenia dát s väzbou na dátovú kvalitu, pozri podkapitolu 2.5). Z uvedených potenciálnych konzumentov má definované dátové požiadavky projekt MOU, preto zvolená podmnožina zahŕňala jeho dátové zdroje. Následne bola posúdená prácnosť prípravy / dostupnosť datasetu. Výsledkom procesu postupného posudzovania vhodnosti pre meranie DK bol skrátený zoznam piatich registrov, ktoré slúžia ako zdroje dát pre konzumenta MOU a nachádzajú sa v perzistentnej vrstve CSRÚ. Pri finálnom výbere dvoch registrov boli zvolené tie registre z tohto zoznamu, ktoré boli aj súčasťou pilotného merania DK (kapitola [2.4.12-4 Doterajšie aktivity merania DK](#) ~~Doterajšie aktivity merania DK~~), teda Register adries v správe Ministerstva vnútra SR a Register právnických osôb v správe Štatistického úradu SR.

Dodatočným prínosom nového merania je aj možnosť posúdenia vývoja v dátovej kvalite týchto registrov od roku 2019. Obaja správcovia registrov boli tiež oslovení so žiadosťou a poskytnutie aktuálneho zoznamu biznis pravidiel, ktoré v súčasnosti merajú alebo by chceli merať, ale nemajú na to dostatočné technické podmienky.

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.

Register osôb bol vybraný na meranie aj kvôli jeho komplexnosti a tomu, že je často vnímaný ako problémový z hľadiska kvality. Štatistický úrad ako správca registra spája informácie zo zdrojových registrov, avšak nemá kompetenciu pozmeňovať informácie z týchto registrov aj napriek tomu, že môžu byť nesprávne. Kvality údajov je teda priamo spojená s kvalitou v zdrojových registroch a jej zlepšovanie je výrazne komplikovanejšie, keďže vyžaduje koordináciu mnohých dotknutých organizácií a ich správcov dát. Cieľom merania okrem posúdenia vývoja kvality je pokúsiť sa získať podklady pre návrh opatrení na zvýšenie dátovej kvality (výstup 1.2.1).

Register adries (RA)

Informačný systém registra adries je zdrojom dát o adresách všetkých fyzických budov na území Slovenskej republiky, ktorým bolo určené súpisné číslo. Do registra adries zapisujú údaje obce a mestá, v ktorých kompetencii je určovanie súpisných a orientačných čísiel ako aj určovanie názvov ulíc a iných verejných priestranstiev.

Údaje o územnom členení akými je: kraj, okres, obec, časť obce zapisuje do registra Ministerstvo vnútra Slovenskej republiky, ktoré je aj vlastníkom tohto registra.¹³

Register adries sa v súčasnosti javí ako register s dobrou dátovou kvalitou, keďže za posledné roky bolo venované množstvo úsilia na jeho zlepšenie. Okrem toho, že je veľmi dôležitým registrom, bol zvolený práve z tohto dôvodu s cieľom zhodnotiť, ako bolo toto zlepšenie dosiahnuté, ako je možné DK naďalej zlepšovať a zároveň využiť register adries ako pozitívny príklad na zlepšovanie dátovej kvality ostatných registrov.

4.1.2 Získanie dát

Z hľadiska prístupu k údajom je možné vykonať meranie nad exportom údajov zo zdrojovej databázy alebo využiť priamy prístup do databázy. Pri rozhodnutí o spôsobe získania údajov je potrebné zvážiť aj integráciu objektu evidencie do prostredia Talend. Pokiaľ je objekt evidencie integrovaný v rámci riešenia Talend a teda sú preň určené sémantické popisy atribútov, ktoré sa v ňom nachádzajú, je pripravený na meranie dátovej kvality. Zároveň sa v rámci tohto scenára predpokladá, že vlastníkom údajov má zadané biznis pravidlá, pre ktoré boli v rámci projektu DI vytvorené funkcie na ich meranie. Pokiaľ sú naplnené tieto dva predpoklady a vlastníkom objektov evidencie nemá záujem o úpravu alebo vytvorenie nových funkcií, môže sa pristúpiť k vykonaniu merania dátovej kvality v riešení Talend. Ak objekt evidencie síce má určené sémantické popisy všetkých atribútov, a teda je integrovaný v rámci riešenia Talend, ale nemá vytvorené všetky potrebné funkcie na meranie biznis pravidiel nie je hneď možné vykonať meranie dátovej kvality. Vlastník registra musí prostredníctvom výzvy požiadať Dátovú kanceláriu o vytvorenie týchto funkcií na meranie zadaných pravidiel. Po ich vytvorení bude možné vykonať meranie v riešení Talend.

Využitie registrov umiestnených v perzistentnej vrstve CSRÚ umožnilo znížiť prácnosť pri získaní dát, keďže merací nástroj má priamo prístup k dátam a nebude nutné vykonať export na strane správcu registra ani import na strane nástroja. Ani jeden z registrov umiestnených v perzistentnej vrstve však nemal v Talende vytvorené potrebné funkcie na meranie biznis pravidiel.

4.1.3 Výber parametrov

Výber parametrov na meranie dátovej kvality je úzko spätý s biznis požiadavkami na register, na ktorom bude vykonávané meranie. Ak aj na niektorom objekte evidencie nebudú na meranie dátovej kvality využité všetky parametre dátovej kvality, využitie všetkých parametrov DK je odporúčané. V prípade merania dátovej kvality na registri RPO a RA existujú biznis požiadavky, na základe ktorých boli vykonávané merania už v minulosti, a teda sú už k dispozícii biznis pravidlá na meranie dátovej kvality. Preto sme pre určenie, ktoré parametre sa majú použiť na meranie dátovej kvality postupovali takto:

- **Revízia existujúcich biznis pravidiel pre vybraný dátový súbor** – V prípade RPO a RA existujú už zadané biznis pravidlá, ktoré však treba validovať. (V prípade iných registrov je možné očakávať, že biznis pravidlá nebudú zadané v dostatočnej kvalite a forme, a teda bude nutné ich preformulovať). V tomto kroku dôjde zároveň aj k identifikácii parametrov DK, ktoré sú v meraniach na dátovom objekte využívané.

¹³ [Register adries, Ministerstvo vnútra SR - Verejná správa \(minv.sk\)](http://Register.adries.Ministerstvo.vnutra.SR-Verejna.sprava.minv.sk)

- **Doplnenie alebo vytvorenie nových biznis pravidiel a KPI** – z validácie existujúcich pravidiel môže vyplynúť potreba vytvoriť nové pravidlá, keďže požiadavky na dátovú kvalitu sa vyvíjajú. (Pre iné dátové súbory, ktoré nemajú zadefinované biznis pravidlá by ich v tomto kroku bolo treba vytvoriť). Pracujeme tu ako s východiskom so zoznamom parametrov na meranie dátovej kvality, ktoré sú zadefinované v Metodike. Dôležitým krokom je aj určenie prahových hodnôt pre vytvorené biznis pravidlá. Určenie prahovej hodnoty by malo vždy vychádzať z potreby dosahovať túto hodnotu na základe: technických požiadaviek, funkčných požiadaviek alebo požiadaviek, ktoré vníma konzument alebo vlastník dát ako dôležité.
- **Výber parametrov** - v prípade oboch predmetných registrov (RA a RPO) nebolo v minulosti realizované meranie dátovej kvality pomocou všetkých parametrov dátovej kvality. Aj z dôvodov porovnateľnosti budú pre meranie DK využité tieto parametre:
 - **Presnosť** – miera, do akej predmet dôkazu predstavuje skutočný svet, vyjadrená zhodou s referenčnými údajmi. Meraná bude syntaktická presnosť.
 - **Konzistentnosť** – Vzájomné logické vzťahy v rámci objektu záznamu sú korektne kompatibilné a v súlade so stanovenými biznis pravidlami.
 - **Správnosť** – Súlad údajov s kritériami, ktoré určujú formát údajov.
 - **Kompletnosť** – kompletnosť údajov znamená, že v údajovom prvku sú prítomné všetky údaje z objektu evidencie, považované alebo označené za povinné. Meraná bude teda „vyplnenosť“ povinných atribútov.
 - **Unikátnosť** – Vyhodnotenie duplicity údajov vo vzťahu k jedinečnému odkazovateľnému identifikátoru.

Výstupom prípravnej fázy bude zoznam biznis pravidiel v tabuľke, ktorá bude mať rovnakú štruktúru ako súčasný zoznam centrálnych biznis pravidiel publikovaný na stránkach Dátovej kancelárie¹⁴. Zoznam biznis pravidiel bude obsahovať tieto atribúty: Skratka orgánu verejnej moci, Názov orgánu verejnej moci, Názov datasetu, ID biznis pravidla, Znenie biznis pravidla, Zdroj biznis pravidla, Platnosť, Platnosť od, Platnosť do, Poznámka, Cieľ biznis pravidla, KPI, Názov KPI.

Pri výbere biznis pravidiel pre meranie na RA bolo potrebné zvážiť účel merania. Cieľom bolo zhodnotiť vývoj DK v registri od pilotného merania a identifikovať priestor na ďalšie zlepšenie. MV SR vykonáva nepravidelné merania DK prostredníctvom selectov nad IS RA za účelom zistenia chybných záznamov. RA pôvodne vznikol migráciou údajov o pobytoch z REGOB, pri ktorej boli do registra prenesené záznamy s vysokou chybovosťou bez kontroly na vstupe. V priebehu niekoľkých rokov boli na vstupných formulároch implementované kontroly, ktoré zamedzujú napríklad vzniku duplicít alebo chybám formátu, vďaka čomu sa výrazne zvýšila dátová kvalita novovytváraných záznamov. Dátová kvalita historických záznamov bola zvyšovaná postupne. Pri identifikácii biznis pravidiel pre vykonanie tohto merania DK boli v spolupráci s dátovým kurátorom MV SR boli zvolené dve skupiny biznis pravidiel:

- Jednoduché pravidlá (napr. dodržanie vyžadovaného formátu), ktoré sú v súčasnosti nastavené ako kontroly pri vstupe údajov do registra a ako jednoduché systémové kontroly na existujúcich údajoch. Účelom ich použitia v tomto meraní je identifikovať prípadné zostávajúce historické záznamy, ktoré nespĺňajú definované podmienky. Pri týchto pravidlách je predpokladaná nízka chybovosť.
- Komplexné pravidlá, ktoré nebolo možné nastaviť ako automatizované kontroly pri vstupe údajov, ani ako systémové kontroly v informačnom systéme registra. Ide

¹⁴ datalab.digital

o pravidlá zložené z viacerých podmienok naviazaných na viaceré atribúty. Je predpoklad, že počet záznamov, ktoré nevyhovujú týmto pravidlám je vysoký.

Príkladom jednoduchého pravidla je hodnota atribútu effectiveDate, ktorá musí byť menšia než aktuálny dátum (biznis pravidlo BP_MVSR_RA_005). Komplexné pravidlo je napríklad podmienka, že ak je atribút PropertyRegistrationNumber nenulový, musí existovať väzba na nadradený objekt MUNICIPALITY, čiže hodnota atribútu municipalityIdentifier na zázname nesmie byť NULL (biznis pravidlo BP_MVSR_RA_097).

Štatistický úrad vykonával pravidelné kontroly DK a publikoval ich reporty do marca 2022. Z dôvodu personálnej zmeny (odchod pracovníka zodpovedného za meranie, vyhodnotenie a reportovanie DK) boli tieto aktivity prerušené. V súčasnosti ŠÚ SR realizuje dopytový projekt vytvorenia nástroja na meranie DK, ktorý sa nachádza v analytickej fáze. Podľa informácií, poskytnuté autorom tohto dokumentu sa projektový tím venuje výberu parametrov merania a je značne vyťažovaný, preto zatiaľ nebolo možné poskytnúť aktualizovaný zoznam biznis pravidiel mapovaný na atribúty. Z týchto dôvodov bolo definovanie meraných biznis pravidiel na RPO problematické. Po konzultácii s Dátovou kanceláriou bol zvolený prienik biznis pravidiel meraných v pilotnom meraní, parametrov z historických meraní do marca 2023 a atribútov, ktoré sa nachádzajú v perzistentnej vrstve v CSRÚ. Všetky biznis pravidlá v zozname budú len jednoduché pravidlá, napríklad atribút Activities musí obsahovať aspoň jeden aktívny predmet činnosti.

4.1.4 Výber ukazovateľov (kľúčové výkonnostné ukazovatele, KPI)

Výber ukazovateľov pre meranie DK vybraných registrov (RA a RPO) sa zakladá na biznis požiadavkách v rámci týchto registrov a je úzko spätý priamo so stanovovaním biznis pravidiel, pomocou ktorých bude vykonané samotné meranie.

V prípade registra RA už existovali základné biznis pravidlá, ktoré boli odvodené od základných požiadaviek na dátovú kvalitu registra a nachádzajú sa ako filtre priamo na vstupe do registra. Ukazovatele teda vyplynuli priamo z požiadaviek na dátovú kvalitu registra.

Ukazovatele pre meranie na RPO boli použité z pilotného merania.

Metodika merania DK uvádza všeobecné parametre dátovej kvality, pre ktoré by mali byť ďalej vybraté ukazovatele dátovej kvality. Zároveň definuje, ako by mali byť merateľné ukazovatele aplikované a či by mali byť využité na meranie DK na lokálnej alebo centrálnej úrovni v rámci dátového toku. Určuje aj prahové hodnoty, ktoré sú nastavené všeobecne podľa príkladov dobrej praxe. Prahové hodnoty merateľných ukazovateľov je dobré nastavovať a upravovať na základe špecifickej požiadavky dátového konzumenta alebo vlastníka. Ďalej sa teda pozrieme na požiadavky centrálnych konzumentov (centrálne projekty riadenia dát s väzbou na dátovú kvalitu, pozri podkapitolu 2.5).

4.1.4.1 Požiadavky na ukazovatele kvality z hľadiska konzumentov dát

Pre účely tohto výstupu berieme do úvahy najmä potreby používateľov dát, uložených v IS CSRÚ. Každý z týchto centrálnych projektov (pozri kapitolu 2.5) sa aktuálne nachádza v inej fáze realizácie a integrácie dát do budovaného softvérového

prostredia, čo priamo ovplyvňuje, do akej miery je daný projekt schopný definovať svoje špecifické požiadavky.

4.1.4.1.1 Moje údaje (MOU)

Konzument MOU je zatiaľ jediný, ktorý má špecifické požiadavky na dátovú kvalitu. Ich dôležitosť je podčiarknutá aj tým, že dáta sa majú využívať na právne úkony.

Konzument MOU (počas konzultácií s vlastníkami projektu) identifikoval ukazovatele, ktoré považuje za prioritné a ich prahové hodnoty by mali byť v ideálnom prípade stanovené vyššie, ako boli zadefinované v metodike.

Konzument MOU špecifický tím, že ako doposiaľ jediný vyžaduje dáta vo formáte JSON-LD. Preto je aj meranie dátovej kvality odlišné. Keďže dáta v takomto formáte nie sú dostupné, je potrebné použitie transformačného nástroja, ktorý zabezpečí vytvorenie dátového formátu JSON-LD. Transformácia priamo ovplyvňuje, ako je možné merať dátovú kvalitu a zároveň po jej uskutočnení vznikajú nové prvky, na ktorých je tiež potrebné merať dátovú kvalitu. Avšak najzásadnejším rozdielom je, že po transformácii dát z xml alebo csv do JSON-LD už nie je možné zmerať v tomto formáte dátovú kvalitu určitých dátových prvkov zo zdrojových dátových súborov. Do špecifikácie ukazovateľov teda pridávame aj požiadavku, či sa má merať pred alebo po vykonaní transformácie, prípadne aj pred aj po nej.

Tabuľka 7 Prioritné merateľné ukazovatele

Merateľný ukazovateľ	Nová prahová hodnota	Pôvodná prahová hodnota	Meraný pred/po transformácii
Syntaktická presnosť hodnoty	>90%	N/A	Pred
Dodržiavanie formátu atribútu	>99%	>99%	Pred a po
Design v súlade so štandardom	>99%	>99%	Pred a po
Výplnenosť povinného údaja	>99%	>99%	Pred a po
Unikátnosť identifikátora	>99%	>90%	Pred a po
Unikátnosť záznamov	>90%	>80%	Pred a po
Rýchlosť aktualizácie	V reálnom čase / raz za deň	N/A	Pred a po
Kompletnosť referenčného identifikátora	100%	100%	Pred a po
Podiel referencovaných objektov	Menej prísny v závislosti od objektu	>90%	Po

Nad rámec všeobecných požiadaviek na dátovú kvalitu, ktorá vyplynuli z Metodiky, konzument MOU stanovil merateľné ukazovatele v týchto kategóriách:

WEBID

- Je unikátny identifikátor, musí sa nachádzať pri každom zázname, a teda jeho prahová hodnota vyplnenosti musí byť 100%, inak sa jedná o systémovú chybu.
- Musí mať formát správneho IRI, správnosť formátu musí byť pri každom zázname, a teda jeho prahová hodnota musí byť 100%, inak sa jedná o systémovú chybu
- Obsahuje PČO (Počítačové číslo občana) – malo by byť obsiahnuté pri 100% záznamoch, ale aj 90% môže byť dostačujúcich. Nižšia miera existencie PČO

v záznamoch obmedzí rozsah funkcionalít, ktoré bude MOU schopné poskytnúť občanovi so záznamom bez PČO.

Identifikátor grafu

- Identifikátor grafu jednotlivých datasetov identifikuje zdroj údajov v grafe, je niektorou z číselníkových hodnôt ISVS, tento identifikátor by mala byť vyplnený pri každom údaj, ktorý ma pridelenú číselníkovú hodnotu v ISVS. Do MOU budú vstupovať aj údaje tretích strán z privátneho sektora, ktoré nemajú pridelené číselníkové hodnoty, a teda nie je možné stanoviť prahovú hodnotu na toto meranie.

Identifikátor osoby RFO

- Záznam fyzickej osoby obsahuje iný jednoznačný identifikátor osoby (aspoň jeden záznam "adms:identifier") - vyplnenosť aspoň 95 %. V RFO by nemali existovať osoby bez identifikátorov, tieto prípady ukazujú buď na systémovú chybu alebo zlú kvalitu zdrojového systému.

Číselníková hodnota

- Číselníková hodnota do transformácie vstupuje ako jednoznačný kód (kľúč) a nie text (hodnota) / teraz tomu tak nie je a je to skôr požiadavka na kvalitu vstupných dát, teda zdroja, alebo formátu dát zdroja. Vyplnenosť jednoznačného kódu (kľúča) by mala byť aspoň 80 %. Pokiaľ atribút nie je vyplnený, musia sa kľúče mapovať späť z hodnôt, čo predstavuje potenciálny zdroj nepresností.
- Po transformácii obsahuje kód v atribúte "skos:notation" – vyplnenosť aspoň 95 %, v opačnom prípade to znamená, že nebude možné priradiť platný kód hodnote na vstupe transformácie
- Text v atribúte "skos:prefLabel" pre slovenskú lokalizáciu - vyplnenosť aspoň 95 %, aplikácie ako MOU klient ich používajú na zobrazenie čitateľnej hodnoty pre koncového používateľa.
- Text v atribúte "skos:prefLabel" pre anglickú lokalizáciu – vyplnenosť aspoň 50 %, interoperabilita pre výmenu dát aj mimo hraníc SR. Musí byť zabezpečená aj "zrozumiteľnosť" dát pre iných koncových používateľov, aktuálny stav je 0 (alebo jednotky percent), bolo by potrebné dopracovať lokalizácie aspoň do anglického jazyka.
- Po transformácii sa nachádza referencia na zdrojový číselník v atribúte "skos:inScheme" – vyplnenosť aspoň 95 %. V opačnom prípade sa porušuje princíp štvorhviezdikovej dátovej kvality.
 - Referencia na zdrojový číselník obsahuje názov číselníka minimálne v slovenskej lokalizácii – vyplnenosť aspoň 95 %, aplikácie používajú atribút ako ľudsky čitateľné pomenovanie pre koncového používateľa.

4.1.4.1.2 Životné situácie (ŽS) & Konsolidovaná analytická vrstva (KAV)

Dátoví konzumenti Životné situácie a Konsolidovaná analytická vrstva v súčasnosti nešpecifikovali žiadne špecifické požiadavky na dátovú kvalitu nad rámec tých, ktoré boli zadané v rámci Metodiky. U každého z týchto dvoch konzumentov je dôvod iný:

- Projekt Životné situácie má za cieľ odstraňovať bariéry v rámci komunikácie medzi štátom a občanom v rôznych životných situáciách. Dátová kvalita je riešená rôzne

v jednotlivých situáciách, avšak ŽS nemajú na dátovú kvalitu žiadne špecifické požiadavky nad rámec Metodiky. Zaujímavým aspektom, ktorý je potrebné spomenúť, je riešenie dátovej kvality vo vstupných online formulároch prostredníctvom dizajnu manuálu ID-SK, ktorý predpisuje jednotný manuál vzhľadu formulárov a hovorí aj o tom, aké filtre by mali byť nastavené na jednotlivé polia formuláru, aby sa zabezpečila vstupná dátová kvalita. Okrem iného existuje aj požiadavka, ktorá hovorí o nutnosti referencie na údaje, ktoré sú vypíňané, teda aby nebolo do formulára možné vyplniť napríklad inú adresu akú pozná referenčný register. Zdá sa však, že tento predpis v súčasnosti nie je implementovaný, a teda referencovanie údajov v rámci formulárov nie je funkčné.

- Projekt Konsolidovaná analytická vrstva sa nachádza len v počiatočných fázach realizácie, v súčasnosti ešte nie je jasné, ani ktoré dátové súbory budú vstupovať do KAV. Preto špecifické požiadavky na dátovú kvalitu neboli definované.

4.2 Vykonanie merania

Táto podkapitola opisuje, ako bude vykonané strojové meranie dátovej kvality na vybranej vzorke. Vzhľadom na komplexnosť prípravného procesu a termínu odovzdania tohto výstupu bude vyhodnotenie vykonaného strojového merania a navrhnutie súboru odporúčaní pre zlepšenie dátovej kvality aj na základe zistených výsledkov merania obsahom výstupu 1.2.1 Opatrenia pre zvýšenie dátovej kvality, kam aj vecne patrí.

Meranie RPO a RA

Meranie dátovej kvality vybraných registrov prebehne v nástroji Talend, ktorý je integrovaný v rámci projektu Dátová integrácia. Po ukončení všetkých prípravných krokov pre obidva existujúce zdrojové registre vzniknú dva zoznamy biznis pravidiel, ktoré budú následne spracované do jednotlivých meraní v rámci procedúry DQ2 Validácie údajov voči OE. Meranie dátovej kvality na základe biznis pravidiel definovaných pre RA a RPO prebehne na atribútoch týchto registrov umiestnených v perzistentnej vrstve IS CSRÚ.

Meranie dátovej kvality prebehne s využitím dvoch z dostupných procedúr v Talende, vyššie spomenutej procedúry DQ2 a procedúry DQ4.

DQ2 – Validácia údajov voči OE

Táto procedúra je zameraná na overenie špecifických biznis pravidiel vzťahujúcich k špecifickému dátovému atribútu meraného registra, výstupom tohto merania bude report s výsledkami merania jednotlivých biznis pravidiel. Na základe daného reportu bude následne vo výstup 1.2.1 Opatrenia pre zvýšenie dátovej kvality vyhodnotená dátová kvalita pre daný register a budú v ňom navrhnuté konkrétne opatrenia na jej zvýšenie. Tento typ merania bude vykonaný na oboch registroch.

DQ3 – Asistované čistenie údajov

Využitím tejto procedúry dôjde k overeniu špecifických biznis pravidiel vzťahujúcich sa k špecifickému dátovému atribútu podobne ako pri procedúre DQ2, avšak s tým rozdielom, že pri tomto meraní výstupom nebude report so štatistickým súborom o dátovej kvalite nad jednotlivými atribútmi podľa biznis pravidiel. Výstup tejto procedúry poskytne report so zoznamom konkrétnych chýb. Ten bude slúžiť na identifikovanie nekvality a možnosti jej následnej nápravy pre jednotlivé záznamy. Táto procedúra dátovej kvality bude aplikovaná iba pre Register adries, keďže pre Register právnických osôb tento typ výstupu na rozdiel od RA nevyužije. Hlavným dôvodom nerealizovania tohto typu merania pre RPO je teda fakt, že podobné výstupy sa pokúša dosiahnuť Štatistický úrad ako správca tohto registra interne, v rámci dopytového projektu zameraného na zlepšovanie dátovej kvality RPO.

4.3 Výstup z merania

Táto časť popisuje kroky 6 Vytvorenie protokolu a 7 Zaslanie protokolu sedemkrokového procesu merania DK.

Výsledkom merania na Registri adries budú dva protokoly:

- analytický protokol, ktorý uvádza počet chybných záznamov a percentuálnu úroveň dosiahnutie stanovených parametrov pre definované ukazovatele,
- podrobný protokol, t.t. zoznam všetkých záznamov, na ktorých boli zistené nezhody voči definovaným biznis pravidlám.

Zatiaľ čo analytický report popisuje úroveň dátovej kvality registra na základe definovaných parametrov, podrobný report bude slúžiť správcovi registra ako podklad na komunikáciu so zdrojmi údajov so žiadosťou o opravu nekvalitných záznamov. Správca registra si bude môcť opakovane spúšťať meranie DK na aktualizovanej časti registra v perzistentnej vrstve CSRÚ a merať postup čistenia údajov.

Keďže správca Registra právnických osôb nemal záujem o vytvorenie podrobného reportu, výsledkom merania na RPO bude len analytický protokol.

Protokoly z merania budú zaslané správcom registra, Dátovej kancelárii (MIRRI) a tvorcami tohto dokumentu, ktorí použijú informácie v nich uvedené ako podklad pre výstup 1.2.1 Opatrenia pre zvýšenie dátovej kvality.

5 Popis alternatívnych nástrojov pre analýzu kvality dát

5.1 O prístupe k výberu nástroja pre dátovú kvalitu

Prístup k výberu nástroja na správu údajov (vrátane DK) by nemal byť riadený len technologickými úvahami, mal by byť definovaný všeobecnou stratégiou správy údajov. To znamená, že pred vykonaním prieskumu trhu alebo pred testovaním a pilotným nasadením konkrétnych produktov by mali byť vypracované, zhromaždené, prioritizované a štruktúrované požiadavky a pravidlá a mala by byť vypracovaná cieľová architektúra údajov, ako aj plán na dosiahnutie požadovaného stavu. To vyžaduje analýzu nasledujúcich faktorov a ich zosúladenie s cieľmi organizácie.

Tabuľka 8 Faktory, ktoré ovplyvňujú výber nástroja pre analýzu DK

Faktor	Popis	Poradie/ Dôležitosť
Stratégia a plán aktivít	Aká je schválená a prijatá stratégia dosiahnutia cieľového stavu v dátovej kvalite? Ako vyzerá jej zodpovedajúci podrobný plán konkrétnych iniciatív/aktivít?	1/kľúčový
Biznis pravidlá a požiadavky	Aké sú požiadavky na cieľové dátové prostredie?	2/kľúčový
Legislatíva a nariadenia	Všetky relevantné požiadavky a obmedzenia stanovené slovenskou a medzinárodnou legislatívou.	3/kľúčový
Bezpečnosť	Všetky relevantné aspekty týkajúce sa kybernetickej bezpečnosti a kontinuity podnikania.	4/kľúčový
Pokrytie	Pokrytie jednotlivých pilierov riadenia kvality dát, podrobne popísaných v tabuľke nižšie.	5/významný
Integrácie/IS ekosystém	Technické požiadavky alebo obmedzenia, vrátane vplyvu ostatných v organizácii už nasadených riešení a/alebo doplnujúcich (alternatívnych) systémov.	6/kľúčový
Zdroje a podpora dodávateľom	Aké sú naše dostupné ľudské kapacity – používatelia, experti, technická podpora? Aké nároky na používateľa má zvažovaný nástroj (t. j. aké máme požiadavky na jednoduchosť používania)? Aké sú nábehové krivky učenia používania nástroja? Vyžaduje sa lokalizácia (napr. podpora slovenského jazyka)? Aká kvalitná je dostupná dokumentácia? Akú podporu používateľom resp. implementátorom poskytuje dodávateľ?	7/kľúčový
Nákupná stratégia	Akú máme nákupnú stratégiu? Preferujeme väčší univerzálnejší balík pre správu dát alebo izolované špecializované nástroje? Preferujeme viazanosť na dodávateľa alebo skôr viazanosť na vývojárov produktu na mieru (buy vs. build)?	8/kľúčový
Analýza prínosov a nákladov	Nákupná cena, náklady na podporu, ponúkaný typ licencie (kapitálové vs. prevádzkové náklady)? Aká je návratnosť investície (ak vieme určiť)?	8/kľúčový

Správa údajov je postavená na 11 pilieroch, ktoré je potrebné v organizácii vybudovať a pri zvažovaní nástrojov alebo riešení pre dátovú kvalitu je dôležité vziať na vedomie, ako a do akej miery vybraný nástroj pokrýva jednotlivé piliere. Popis jednotlivých pilierov je v nasledujúcej tabuľke.

Tabuľka 9 Pilieri správy údajov

Pilier	Hlavný cieľ alebo účel
Stratégia a správa údajov	Určenie stratégie za údajmi a zavedenie správnej štruktúry a politik (typicky nebýva pokryté automatizovanými nástrojmi)
Dátová architektúra a modelovanie	Vytvorenie, modelovanie a dokumentácia dátovej architektúry a jej komponentov
Metaúdaje	Správa rôznych typov metadát naprieč aplikáciami, procesmi prostredníctvom centralizovaného prístupu
Dátová kvalita	Zabezpečenie, že údaje sú vhodné na zamýšľaný účel, a to zameraním sa na úplnosť, presnosť, konzistentnosť a aktuálnosť údajov
Zdrojové a referenčné údaje	Vytvorenie jedinej verzie pravdy v celej organizácii
Dátové operácie	Operatívne riadenie dát od získania, cez využitie až po vyradenie
Správa dokumentov a obsahu	Štandardizácia organizácie a ukladania obsahu (predovšetkým dokumentov) v rámci celej organizácie
Dátová integrácia a interoperabilita	Zabezpečenie zdieľania, integrácie a univerzálneho používania údajov v rámci organizačných útvarov a aplikácií
Bezpečnosť údajov	Zabezpečenie správnej autentifikácie, autorizácie, monitorovania a auditovania údajov
Business Intelligence/reporting	Transformácia údajov na informácie na podporu analýzy a rozhodovania
Pokročilá analytika	Použitie pokročilých techník na modelovanie a predpovedanie budúceho správania

Pri výbere a implementácii nástroja dátovej kvality sú k dispozícii tri základné možnosti:

- výber jedného dodávateľa na celé riešenie,
- výber špecializovaných nástrojov pre jednotlivé požiadavky časti riešenia, alebo
- vývoj na mieru.

Prvá možnosť predpokladá pokrytie automatizácie a podpory všetkých (viacerých) procesov a pilierov implementáciou jediného riešenia od konkrétneho dodávateľa. Druhá možnosť predstavuje výber rozličných dodávateľských riešení pre konkrétne funkcie s využitím výhody definovania konkrétnych cieľov a požiadaviek pre každé z nich (best-of-breed). Nakoniec, procesy dátovej kvality je možné riadiť interne alebo dodávateľsky vyvinutými skriptami a aplikáciami. Každý z týchto prístupov má svoje výhody a nevýhody a organizácie zvyčajne využívajú určitú mieru ich kombinácie. Tieto výhody a nevýhody sú popísané v tabuľke nižšie.

Tabuľka 10 Výhody a nevýhody prístupov k výberu riešenia riadenia DK

Prístup	Výhody	Nevýhody
Výber jedného riešenia / dodávateľa (single vendor)	Integrovanosť Podpora OPEX Existuje ekosystém	Viazanosť na jedného dodávateľa (vendor lock-in) Prispôbenie môže byť ťažšie a/alebo nákladnejšie
Výber viacerých špecializovaných nástrojov / riešení (best-of-breed)	Prispôbené pre dané odvetvie Dá sa dosiahnuť lepšia funkčnosť (výber najlepších funkcií) OPEX Nižšia viazanosť na jedného dodávateľa (vendor lock-in)	Integrovanosť Celkové náklady Vysoké požiadavky na odbornosť personálu (najmä technického)
Vlastný vývoj	Prispôbené pre odberateľa CAPEX Nie je viazanosť na jedného dodávateľa (vendor lock-in)	OPEX Integrovanosť Bezpečnosť/prevádzková kontinuita Škálovateľnosť Viazanosť na vývojárov

Samozrejme, výhody a nevýhody jednotlivých prístupov sa môžu výrazne líšiť v rozličných prípadoch použitia. Napríklad licenčné politiky pre komerčných zákazníkov sa môžu odlišovať od tých pre verejný sektor a môžu byť definované samostatne pre jednotlivé obchodné prípady, takže výsledná výška investičných a prevádzkových nákladov môže byť určená až po rokovaniach s dodávateľmi. Na druhej strane, pred výberom nástroja na riadenie DK je potrebné analyzovať ďalšie parametre a faktory, dostupnosť skúsených zamestnancov a vývojárov na podporu konkrétnej technológie, prístup k odporúčaným postupom, miesto nasadenia (cloud alebo v mieste prevádzky), ďalšie požiadavky na infraštruktúru, škálovateľnosť rozličných uvažovaných riešení a podobne. Vo všeobecnosti však platí, že v rámci prostredia riadenia dátovej kvality musí nasadzované riešenie pokrývať nasledujúce oblasti a funkčnosť:

- **Pripojiteľnosť:** schopnosť prístupovať k dátam naprieč množstvom zdrojov údajov, vrátane externých a interných, statických alebo streamovaných, relačných a nerelačných, umiestnených lokálne alebo v cloude.
- **Profilovanie:** schopnosť štatistickej analýzy rôznych datasetov, vrátane štruktúrovaných i neštruktúrovaných, umiestnených lokálne alebo v cloude. Analýza musí používateľom priniesť poznatky o dátovej kvalite a umožniť im identifikovať súvisiace problémy.
- **Analytika a interaktívne vizualizácie:** Interaktívne analytické pracovné toky a vizuálne výstupy zo štatistickej analýzy, ktoré pomôžu vecným aj IT používateľom identifikovať, pochopiť a monitorovať problémy dátovej kvality a odhaliť vzorce a trendy prostredníctvom reportov a najrôznejších typov výstupov, prehľadových panelov (dashboard) i na mobilných zariadeniach.

- **Monitorovanie a detekcia:** Monitorovanie dátovej kvality na základe predkonfigurovaných, na mieru vytvorených alebo adaptívnych pravidiel a upozornení na porušenie pravidiel. Automatické zisťovanie prekročenia hodnôt, anomálií, vzorov a posunov. Takisto poskytnutie monitorovacích panelov, protokolovacích súborov (log files) a auditných záznamov pre preukázanie dodržiavania súladu s nariadeniami.
- **Rozklad, štandardizácia a čistenie údajov:** rozklad (parsing) a formátovanie rozličných datasetov založený na všeobecných štandardoch pre verejnú správu alebo interných štandardoch, biznis pravidlách, znalostných bázach a metaúdajoch prípadne doplnený o strojové učenie (Machine Learning, ML). Úprava hodnôt údajov tak, aby boli v súlade s doménovými obmedzeniami, ako aj s obmedzeniami z hľadiska ich integrity a ďalšími pravidlami organizácie.
- **Priradenie, prepájanie a zlučovanie:** Priraďovanie, prepájanie a zlučovanie súvisiacich dátových záznamov v rámci jedného datasetu alebo naprieč rozličnými datasetmi s využitím rozličných tradičných alebo nových prístupov, napríklad pravidiel, algoritmov, metaúdajov prípadne umelej inteligencie (AI) a ML.
- **Metaúdaje a sledovanie pôvodu dát:** zber a odhaľovanie metaúdajov alebo ich import z nástrojov tretích strán. Zostavenie alebo import línie pôvodu dát (data lineage) na vykonanie rýchlej analýzy základných príčin problémov s kvalitou údajov a analýzy vplyvu jej prípadnej nápravy. Aplikovanie aktívneho vyhľadávania príp. extrakcie metadát s využitím odporúčaní a asociácií pre odhalenie vzorcov a pravidiel. Funkcie vyhľadávania a katalogizácie údajov, zobrazenie metrík v porovnaní s kritickými dátovými prvkami.
- **Overenie adresy/geokódovanie:** Podpora pre štandardizáciu a čistenie údajov súvisiacich s polohou. Dopĺňanie čiastkových údajov v reálnom čase alebo v dávkovom procese.
- **Spracovanie a obohacovanie údajov:** možnosť integrovať údaje z externých zdrojov pre zlepšenie úplnosti a hodnoty údajov.
- **Správa pravidiel a validácia údajov:** funkcie pre návrh, tvorbu a nasadenie biznis pravidiel pre konkrétne hodnoty údajov. Pravidlá je možné aplikovať v rámci použitého riešenia alebo ich volať z aplikácií tretích strán na účely overenia údajov, čo možno vykonať v dávkovom režime alebo v reálnom čase.
- **Pracovné toky (workflow) pre riešenie problémov:** možnosť definovať a používať pracovné toky (workflows) na automatizáciu resp. riadenie toku procesov riešenia problémov s kvalitou údajov (stewardship workflow), ktoré umožnia ľahkú identifikáciu, sanitáciu, priraďovanie na riešenie, eskaláciu a vyriešenie problémov s kvalitou údajov, uľahčenú nástrojmi pre spoluprácu, priebežné monitorovanie a správa jednotlivých prípadov (case management).
- **Podpora DataOps:** Spolupráca v oblasti riadenia kvality dát so zameraním na zlepšovanie komunikácie, integrácie a automatizácie tokov dát medzi manažérmi dát a užívateľmi dát v celej organizácii.
- **Architektúra a integrácia:** podpora pre integráciu, previazanosť, konzistenciu a interoperabilitu medzi rôznymi komponentmi sady nástrojov pre kvalitu údajov (vrátane nástrojov tretích strán) a iných riešení alebo komponentov na správu údajov.
- **Automatizácia a rozšírenie:** možnosť využívať pokročilé technológie (ako sú AI/ML, znalostné grafy, aktívne metaúdaje alebo spracovanie prirodzeného jazyka) na automatizáciu, aby sa minimalizovalo manuálne úsilie a zvýšila rýchlosť spracovania.

- **Podpora viacerých rolí:** možnosť využívania nástroja viacerými rolami, vrátane netechnickými používateľskými rolami, ako sú napr. dátoví inžinieri, správcovia dát (data steward), analytici DK, dátoví architekti, biznis analytici, príprava dát atď.

Nie všetky funkcie uvedené vyššie sú povinné alebo rovnako dôležité pre výber nástroja na analýzu a riadenie dátovej kvality, avšak vybrané riešenie by malo byť schopné zachovať interoperabilitu s inými riešeniami, ktoré podporujú niektoré z požadovaných funkcií.

Rozhodujúci význam majú podrobné a štruktúrované biznis pravidlá a požiadavky. Vo väčšine prípadov je schopnosť ich podpory kľúčovým rozdielom, ktorý umožňuje zúženie výberu.

Pôvodný predpoklad v roku 2019 bol, že väčšina funkcií súvisiacich s dátovou kvalitou, vrátane merania a analýzy, sa bude vykonávať v prostredí Talend. V rámci aktivít zameraných na dátovú kvalitu v projektoch DI a CIP však boli uvažované a interne testované alternatívne riešenia. Talend Enterprise bol porovnaný s niekoľkými nástrojmi z hľadiska počtu schopností/funkcií, vrátane profilovania údajov, čistenia údajov, vykazovania kvality údajov, integrácie údajov a zdrojov (databázy, REST API, súborové systémy atď.), správcovstva, správy používateľov a prístupu, ako aj dostupnosti cloudovej verzie. U niektorých funkčností bola kontrolovaná iba ich dostupnosť v nejakej forme (bližšie nešpecifikovanej resp. špecifikovanej ako funkcia poskytnutá/neposkytnutá).

Medzi testovanými nástrojmi boli

- Great Expectations,
- OpenRefine,
- DataCleaner Community Edition,
- Open Source Data Quality and Profiling a
- DataHub (len na správu metadát).

Na základe vykonaných porovnaní bol na ďalšie použitie navrhnutý nástroj DataCleaner Community Edition.

Z nám dostupných informácií MIRRI uprednostňuje riešenia s otvoreným zdrojovým kódom (open source). Navrhujeme teda pridať do zoznamu uvažovaných ďalšie nástroje ako

- Quadient DataCleaner,
- DataMatch Enterprise,
- Apache Griffin pre špecializované úlohy a procesy.

Väčšinu z nich je však možné nasadiť iba v modeli best-of-breed alebo vlastný vývoj. Množstvo dôležitých funkcií riadenia DK je potrebné pokryť inými riešeniami.

Je potrebné zdôrazniť, že v súčasnosti nie je dokončené mapovanie procesov a dátových tokov v celej infraštruktúre vrátane požiadaviek zapojených OVM. Preto je naozaj fundované a kvalitné posúdenie riešení v oblasti kvality údajov podľa spoľahlivých a relevantných kritérií alebo skutočných potrieb všetkých zainteresovaných strán neriešiteľným problémom.

5.2 Rozšírenie výberu odporúčaných nástrojov

V prípade prístupu s jedným dodávateľom (komplexné riešenie pokrývajúce väčšinu úloh správy údajov, single vendor), ktorý posudzuje rôzne komplexné nástroje v porovnaní s požiadavkami a funkčnosťami opísanými vyššie v tejto kapitole, ale aj s ohľadom na predošlé skúsenosti by mali byť do zoznamu zahrnutí aj títo dodávatelia a riešenia, čo je podporené aj trendom ich umiestnenia v hodnotení v rokoch 2020-2022 (v magickom kvadrante firmy Gartner):

- Ataccama¹⁵,
- Experian¹⁶,
- Informatica¹⁷,
- Innovative Systems¹⁸,
- Talend¹⁹,
- TIBCO Software²⁰.

V rámci prístupu single vendor môžu byť ako hlavní kandidáti na užší výber nominovaní Talend a Ataccama (obe riešenia sú open-source), vďaka ich rozsiahlej funkčnosti a osvedčeným postupom pre verejný sektor.

Je zrejmé, že všetky pokusy o vyhodnotenie (bez ohľadu na prístup – single vendor, best-of-breed, vlastný vývoj, hybrid) si nakoniec vyžadujú výpočet celkových nákladov na vlastníctvo, t. j. investičných a prevádzkových nákladov, nákladov na vývoj, integráciu, migráciu a podporu. Tieto výpočty nemožno vykonať bez úplného zoznamu biznis požiadaviek, plánu kapacity a škálovateľnosti, odhadu počtu používateľov pre licenčné pokrytie, aby bolo možné určiť náklady na celkovú a nielen základnú funkčnosť.

V tomto momente teda nie je možné navrhnúť konkrétne nástroje, pretože ich výber a obstarávanie nespadá do rozsahu tohto dokumentu.

¹⁵ [Data Quality | Ataccama](#)

¹⁶ [Data Quality Management Solutions & Services | Experian \(edq.com\)](#)

¹⁷ [Data Quality Tools and Services | Informatica](#)

¹⁸ [Innovative Systems: Enterprise Data Quality Solutions](#)

¹⁹ [Talend Data Quality: Trusted Data for the Insights You Need | Talend](#)

²⁰ [Data Quality | TIBCO Software](#)

6 Definícia systému pravidelnej aktualizácie a reportovania

Dátová kancelária je zodpovedná za zabezpečovanie metodologickej podpory v riadení dátovej kvality a mala by spolupracovať s OVM na prijatí spoločného rámca na podávanie správ o kvalite datasetov a odvodených informácií, ktoré šíria. Spôsob poskytovania správ o kvalite údajov závisí do značnej miery od účelu dátového produktu a používateľov informácií a dokumentovaného dátového produktu. Za vytvorenie jednotného rámca, na základe ktorého budú vytvárané správy o kvalite údajov, by malo byť zodpovedné MIRRI, avšak na úrovni OVM by prirodzene mala byť ponechaná určitá autonómia, aby sa vytvoril priestor pre rezortné špecifiká.

Nasledujúcich šesť bodov slúži ako podklad pre tvorbu a nastavenie efektívneho a funkčného systému reportovania:

- 1 Vytvorenie podrobnej **technickej dokumentácie** o všetkých aspektoch dátového produktu pre zachovanie kontinuity a potreby šírenia produktu. Dokumentácia by mala popisovať hlavné črty dátového súboru, vrátane jeho účelu hlavných použití a očakávanej periodicity aktualizácie. Mala by zahŕňať realizované technické operácie, teda podrobnosti o jeho tvorbe a dátové slovníky.
- 2 **Zohľadnenie kvality informačných systémov a z toho vyplývajúcej kvality údajov**, ktoré poskytujú vrátane identifikovaných hrozieb pre kvalitu alebo jej meranie a kompromisov, ktoré museli byť urobené na zmiernenie hrozieb. Pokiaľ je k dispozícii meranie širokej škály dimenzií a atribútov je možné internú dokumentáciu použiť na prispôbenie dodatočným požiadavkám na používateľské správy o dátovej kvalite pre špecifických používateľov dátových súborov.
- 3 **Zohľadnenie kvality každého zdroja vstupných údajov** pri dokumentovaní integrovaných údajov vrátane všetkých krokov ich spracovania. Dokumentácia by mala obsahovať informáciu o metódach použitých na dodatočné spracovanie zdrojových údajov potrebných na integráciu, vrátane informácie o prípadných úpravách a relevantnú metodickú informáciu o integrácii. Dokumentácia o dátovej kvalite by mala zahŕňať podrobné hodnotenia identifikovaných hrozieb pre integrovanú kvalitu údajov (podľa úrovne relevantnosti). Kvalita dát pre každý vstupný zdroj by mala byť zdokumentovaná pre jeho použitie v integrovanom dátovom produkte.
- 4 Používateľom údajov je dôležité **poskytnúť prehľad o kvalite údajov**, pre niektoré výstupy údajov možno na parafrázovanie takýchto informácií použiť sekciu „často kladené otázky“ (FAQ). Používatelia údajov by mali byť schopní rýchlo a prehľadne sa z týchto informácií dozvedieť najrelevantnejšie problémy s kvalitou údajov, vrátane silných a slabých stránok údajov a informácie o akýchkoľvek prístupoch, ktoré boli použité na zmiernenie problémov a prípadne o dôsledkoch používania takto modifikovaných údajov.
- 5 Správa o kvalite údajov môže pracovať s rôznou úrovňou spresnenia/zrnitosti – podrobnosti dátovej kvality potrebné pre komplexné integrované súbory údajov sa môžu líšiť od podrobností potrebných pre malé dátové súbory. Vo všetkých prípadoch by mali byť uvádzané odkazy na ďalšie dostupné relevantné podrobné dokumentácie k dátovej kvalite v spojitosti so špecifickým dátovým súborom.
- 6 Súhrn musí zahŕňať kľúčové otázky kvality týkajúce sa súboru údajov/informačného systému a ich primárnych používateľov. **Súhrn by mal poskytnúť prehľad o pôvode údajov a opísať ich vhodnosť na konkrétne použitie** alebo použitia a aká je pravdepodobnosť, že kľúčové výstupy údajov môžu viesť k zavádzajúcim informáciám.

Zrelosť riadenia dátovej kvality

Pri navrhovaní efektívneho a funkčného systému monitorovania a reportovania je potrebné vziať do úvahy aj zrelosť DK. Úroveň zrelosti je kvalitatívne určenie založené na hodnote prameniacej z daného stupňa a úrovni jednotlivých dimenzií riadenia kvality v danom čase. Návrh systému musí odzrkadľovať nielen úroveň zrelosti, na akej sa DK práve nachádza, ale najmä úroveň, na akú ho má systém posunúť. Neočakáva sa však najvyššia úroveň zrelosti pri každej schopnosti, vždy je potrebné vziať do úvahy kompromis medzi ich prínosom a nákladmi, ktoré sú s nimi spojené. Rastúca úroveň zrelosti DK so sebou prináša vyššiu strategickú hodnotu. Úrovne zrelosti DK a schopnosti v nich zahrnuté sú bližšie opísané v tabuľke nižšie.

Tabuľka 11 Zrelosť riadenia DK

Úroveň zrelosti	Schopnosti zahrnuté v danej úrovni
Reaktívne	<p>Problémy sú identifikované ad-hoc, nie je zavedený ani iniciovaný žiadny formálny program na vyriešenie zistení merania/auditov dátovej kvality</p> <p>Nedefinované alebo nedostatočne definované štandardy kvality údajov, ad-hoc postupy a nejasné určenie zodpovedností</p> <p>Nie sú k dispozícii žiadne nástroje na zachytávanie a sledovanie problémov s údajmi a ich kvalitou</p>
Kontrolované	<p>Štandardy a procesy dátovej kvality sú definované, ale ešte nie sú konzistentne aplikované v celej organizácii</p> <p>Obmedzený počet pravidiel dátovej kvality definovaných pre kľúčové údaje</p> <p>Miestne tímy spravujú a sledujú niektoré problémy s DK</p> <p>Zodpovednosti za údaje sú len obmedzene definované so slabým mandátom na riešenie problémov</p>
Proaktívne	<p>Štandardy dátovej kvality sú uplatňované na všetky nové iniciatívy v oblasti dát a IT</p> <p>Pravidlá dátovej kvality definované pre všetky kľúčové údaje sú vhodné na daný účel</p> <p>Údaje sú dôsledne spravované s jasne definovanou zodpovednosťou</p> <p>Štandardné reporty sledujú problémy dátovej kvality, ktoré boli identifikované vďaka pravidelnému využívaniu vyhodnocovania kvality</p> <p>Dodržiavajú sa postupy nápravy problémov</p>
Prediktívne	<p>Automatizované profilovanie údajov a generovanie pravidiel na základe definícií údajov</p> <p>Robustné pravidlá DK definované pre všetky kritické dátové prvky so schopnosťou predchádzať problémom dátovej kvality priamo pri zdroji</p> <p>Automatizované informačné panely a metriky na úrovni organizácie umožňujú spravovať a zlepšovať DK v reálnom čase</p> <p>Existuje rozhodovací strom nápravy problémov DK založený na historických opravách</p> <p>Existujúce údaje boli dodatočne vybavené tak, aby spĺňali štandardy kvality</p>

6.1 Zavedené postupy hlásenia a aktualizácií dátovej kvality a podmienok hodnotenia

Ako sa uvádza vyššie v predchádzajúcej kapitole, predpokladom k úspešnému riadeniu dátovej kvality je správne nastavenie reportovania a monitorovania aktualizácií. Jednotlivé OVM k tejto úlohe pristupujú rôzne, avšak od marca 2022 už žiadny z nich nevykonáva merania svojich objektov evidencie, ktorých výsledky (správy z merania) by boli aj následne zverejňované. Tieto informácie sa týkajú predovšetkým datasetov typu register (objektov evidencie), u datasetov typu štatistický produkt (môžeme zovšeobecniť na analytický produkt alebo výstup) je situácia iná, ako je poznamenané nižšie v texte, kde je výber príkladov prístupu k meraniu a vyhodnocovaniu dátovej kvality vybraných oblastí a OVM.

6.1.1 Štatistický úrad – RPO

Štatistický úrad ako jediný OVM doposiaľ pravidelne zverejňoval reporty z meraní. Zverejňované reporty informovali o dátovej kvalite registra právnických osôb (RPO). Prvá správa z merania bola zverejnená v marci 2020, obsahovalo 67 biznis pravidiel, pomocou ktorých bola skontrolovaná dátová kvalita jednotlivých atribútov RPO. Následne boli zverejňované reporty z meraní každý štvrtrok až do marca 2022, počas tohto obdobia sa vyvíjalo aj množstvo biznis pravidiel v správach z merania.

Štruktúra správ z merania dátovej kvality RPO bola nasledovná:

- úvod obsahujúci základné informácie o RPO a meraní,
- zoznam biznis pravidiel, s ktorými bolo vykonané meranie,
- výsledky vykonaného merania po biznis pravidlách,
- odporúčania na zlepšenie dátovej kvality na základe výsledkov merania,
- porovnanie vývoja dátovej kvality na jednotlivých zdrojových registroch RPO podľa biznis pravidiel v čase.

Na Štatistickom úrade aktuálne prebieha dopytový projekt Manažment údajov Registra právnických osôb, podnikateľov a orgánov verejnej moci²¹, tento projekt má zmeniť, ako reportovanie dátovej kvality RPO bude v budúcnosti prebiehať. Hlavnou zmenou je zmena kritérií dátovej kvality, ktoré sa budú používať pre meranie a ktoré sa budú publikovať v reporte o dátovej kvalite. Taktiež dôjde k zmene frekvencie publikovania z intervalu raz štvrtročne na raz za rok. Ďalšou novinkou zavádzanou v dopytovom projekte je rozdelenie merania dátovej kvality podľa biznis pravidiel do 3 kategórií, ktoré určujú, kedy má byť meranie uskutočnené. Týmto skupinami kontrol dátovej kvality sú: Kontrola pri zápise, Kontrola po zápise a Kontrola na vyžiadanie.

Okrem úprav v reportovaní, ktoré sú v súčasnosti pripravované Štatistickým úradom pre RPO je dôležité do štruktúry budúcich reportov zapracovať aj odporúčania z kapitoly 6.2. Zapracovanie odporúčaní je dôležité najmä pre dosiahnutie takej úrovne reportov, aby poskytovali podstatné informácie pre konzumentov údajov a zároveň na základe nich bolo možné prijímať rozhodnutia na zlepšovanie dátovej kvality.

²¹ <https://metais.vicpremier.gov.sk/detail/Projekt/33f4a9e2-9984-460e-a1e1-70557f2a5faa/cimaster?tab=detailForm>

Je však dôležité zdôrazniť, že vyššie uvedené informácie sa týkajú najmä RPO, ktorý je typickým predstaviteľom dátového súboru typu register (register objektov evidencie). Situácia je iná pre dátové súbory typu štatistický produkt, kde je orientácia na kvalitu zakotvená v nariadení Európskeho parlamentu a Rady (ES) 223/2009 o európskej štatistike (a v jeho prípadných novelizáciách). Nariadenie ustanovuje štatistické zásady, ktorými sa riadi rozvoj, tvorba a šírenie európskej štatistiky, kritériá kvality s cieľom garantovať kvalitu štatistických výstupov. Orientáciu na kvalitu ďalej výrazne podporuje Spoločný rámec kvality Európskeho štatistického systému (EŠS), ktorý obsahuje Vyhlásenie EŠS o kvalite, Kódex postupov pre európsku štatistiku²² a Rámec pre zabezpečenie kvality EŠS. Od uvedených dokumentov je odvodený aj prístup ku riadeniu kvality Národného štatistického systému (NŠS) v SR. Zavedený systém manažérstva kvality je stelesnený v Príručke kvality²³, ktorá obsahuje opis systému manažérstva kvality a naplnenie požiadaviek na kvalitu EŠS a tiež normy ISO 9001. Uplatnenie príručky v praxi zabezpečuje, že všetky činnosti, ktoré majú vplyv na kvalitu vytváraných produktov sú plánované, riadené, preskúmané, vyhodnocované a spĺňajú požiadavky akceptované v objednávke odberateľa (konzumenta) produktu.

Integrálnou súčasťou riadenia kvality sú aj Správy o kvalite²⁴, ktoré ŠÚ vydáva k svojim štatistickým produktom a sú chápané ako metadáta k nim (tzv. referenčné metadáta). Tu uvádzame pre názornosť a za účelom referencie štruktúru takej správy, hoci s vedomím, že všetky položky nemusia byť vo všeobecnosti aplikovateľné pre iné prípady použitia:

- Kontakt (organizácia)
- Aktualizácia metaúdajov
- Popis štatistiky
- Merná jednotka
- Referenčné obdobie
- Inštitucionálny mandát
- Štatistická dôvernosť
- Politika zverejňovania
- Periodicita zverejňovania
- Dostupnosť a zrozumiteľnosť
- Riadenie kvality
- Relevantnosť
- Presnosť a spoľahlivosť
- Včasnosť a dochvilnosť
- Porovnateľnosť a koherentnosť
- Náklady a záťaž respondentov
- Revízia údajov
- Štatistické spracovanie
- Poznámka

²² <https://www.susr.sk/wps/portal/ext/aboutus/key.documents/code.of.practise/>

²³ <https://www.susr.sk/wps/portal/ext/aboutus/key.documents/book.of.quality/>

²⁴ [Správy o kvalite \(statistics.sk\)](https://www.susr.sk/wps/portal/ext/aboutus/key.documents/book.of.quality/)

6.1.2 Otvorené údaje

V prípade otvorených dát nedochádza k publikovaniu „tradičných“ reportov dátovej kvality. Jednorazovo bolo v roku 2021 vykonané hodnotenie dátovej kvality vybraných datasetov otvorených údajov²⁵. Pri otvorených dátach sa v správach dátovej kvality hodnotí kvalita objektov evidencie na 5 hviezdčkovej stupnici, ktorá hovorí o použiteľnosti otvorených dát a ich vlastnostiach, a teda z tohto pohľadu o ich kvalite. Ďalej sa ešte vyhodnocujú metadáta publikovaných otvorených údajov, ktoré sú publikované na európskej úrovni aj za Slovensko na webe data.europa.eu²⁶.

Analýza zverejňovania datasetov ústredných orgánov štátnej správy na portáli otvorených dát²⁷ sa venuje zhodnoteniu stavu otvorených údajov z rôznych hľadísk. Časť, ktorá sa pojednáva o kvalite otvorených dát, je časť hodnotenia kvality na základe

5-hviezdčkovej stupnice a pozostáva z týchto častí:

- Vysvetlenie metodológie pre hodnotenie na 5 hviezdčkovej stupnici
- Vyhodnotenie publikovaných otvorených dát na tejto stupnici.

Vzhľadom na to, že otvorené údaje majú slúžiť na mnohé ciele, pri ktorých dátová kvalita je dôležitým aspektom, je dôležité, aby konzumenti týchto dát boli informovaní o ich kvalite. Odporúčame, aby okrem súčasného reportu zahŕňajúceho 5-hviezdčkovú stupnicu kvality otvorených dát, bolo zavedené aj merania iných tradičných ukazovateľov dátovej kvality. Rozsah širšieho merania dátovej kvality a následného reportovania o kvalite otvorených údajov bol navrhnutý aj v rámci projektu Zlepšenia využívania údajov vo verejnej správe vo výstupe č.4²⁸, ktorý je naďalej relevantný a mal by byť využívaný ako podklad pre tvorbu reportov.

6.1.3 Ministerstvo vnútra – Register adries

MV SR spravuje niekoľko desiatok registrov, medzi ktoré patria registre verejnej správy, ale aj evidencie policajného zboru. Podľa nám poskytnutých informácií nie je síce vypracovaná a prijatá interná metodika na meranie DK a zostavovanie reportov z merania DK, vykonávajú sa však merania nad viacerými vybratými registrami, vrátane RA. Pri meraní sa riadia centrálnou Metodikou merania DK. Výsledky merania (správa o kvalite) sa nezverejňujú.

MV SR nevyužíva reportovací nástroj na vykazovanie výsledkov merania DK. Získanie údajov na meranie prebieha najčastejšie formou selektov priamo z údajov jednotlivých registrov s následným manuálnym posúdením a vyhodnotením kvality. Reportovanie, evidovanie, ale aj vyhodnotenie merania kvality sa vykonáva v bežne dostupných kancelárskych nástrojoch s využitím výstupov získaných zo selektov, ktoré zvyčajne definuje dodávateľ daného registra v zmysle požiadavky gestora, ktorý týmto spôsobom definuje resp. kontroluje správnosť aplikovaných biznis pravidiel.

Vyhodnocovanie prebieha dvojfázovo. V prvej fáze výsledky merania posudzuje gestor z hľadiska súladu s biznis pravidlami, resp. s pravidlami, ktoré definuje legislatíva pre

²⁵ [Výstup č. 2 – Dohľadanie nad kvalitou publikovaných datasetov a kontrola, či nedochádza k zneužitiu údajov – DataLab](#)

²⁶ <https://data.europa.eu/mga/catalogues/data-gov-sk/?locale=sk>

²⁷ https://www.minv.sk/swift_data/source/rozvoj_obcianskej_spolocnosti/otvorene_vladnutie/2022/2022-07_26_analyza_datasetov_2022.pdf

²⁸ <https://datalab.digital/wp-content/uploads/Dataset-parametrov-d%C3%A1tovej-kvality-vo-verejnej-spr%C3%A1ve-vo-form%C3%A1te-otvoren%C3%BDch-d%C3%A1t.pdf>

daný register. Následne sa posudzuje chybovosť, dôvod vzniku danej chyby, možnosti jej nápravy (technické a legislatívne) a definujú sa nové pravidlá/návrh opatrení ako predchádzať vzniku nových chýb toho istého druhu. Aj napriek absentujúcej internej metodike sa pre každý register aplikujú nové pravidlá pre predchádzanie chybovosti a hľadajú sa spôsoby ako opraviť zistené nedostatky, pokiaľ je to technicky a legislatívne možné.

Pokiaľ ide o zverejňované množiny údajov typu open data MV SR, kontrola a vyhodnotenie ich dátovej kvality prebieha interným spôsobom. V publikovaných datasetoch sú prezentované len zmenené alebo opravené hodnoty objektov evidencie a atribútov, nie informácie o náraste dátovej kvality. Všetky aktivity merania a reportovania sú vykonávané len na úrovni zdrojových registrov, nie na úrovni publikovaných datasetov.

V priebehu prípravy tohto výstupu sa MV SR zapojilo do aktivity merania DK na RA. Jedným z prínosov tejto aktivity je aj vytvorenie opakovane použiteľných mechanizmov na meranie DK v nástroji Talend s možnosťou generovania analytických a podrobných reportov, ktoré môže MV SR v budúcnosti použiť ako podklady pre aktivity zlepšovania DK a publikovať ich pre verejnosť.

6.1.4 Ministerstvo školstva, vedy, výskumu a športu – Rezortný informačný systém

MŠVVaŠ SR je správcom Rezortného informačného systému (RIS), ktorý má postupne nahradiť papierový systém štatistického vykazovania škôl a školských zariadení zberom údajov do centrálného rezortného systému. RIS zabezpečí konsolidáciu získaných údajov o zamestnancoch a o deťoch, žiakoch a poslucháčoch (DŽP), ich sumarizáciu a sprístupnenie užívateľským organizáciám. Účelom centrálného registra je znížiť súčasnú administratívnu záťaž učiteľov a manažmentu základných a stredných škôl, ako aj školských zariadení, urýchliť proces zberu dát, aktualizáciu údajov a znížiť náklady spojené s ich zasielaním.

Aktualizácia údajov v RIS prebieha na ročnej a na mesačnej báze. Na začiatku školského roka prebehne po ukončení zápisov DŽP na školskú dochádzku jednorazový zber údajov za daný rok. Následne systém prijíma pravidelné mesačné aktualizácie zmien. Tieto údaje sú online stotožňované voči Registru fyzických osôb. Školy vykonávajú aktualizáciu údajov prostredníctvom svojich školských informačných systémov (ŠIS) aScAgenda a eŠkola. Školy, ktoré nemajú tieto ŠIS, si môžu nainštalovať ich bezplatnú verziu určenú len na odoslanie údajov do RIS. RIS nezaznamenáva časové rezy a každou novou aktualizáciou sú pôvodné údaje za jednotlivé školy prepísané.

MŠVVaŠ SR nerealizuje pravidelné merania podľa Metodiky merania DK, a teda ani nezverejňuje žiadne reporty z vyhodnotenia DK. Vykonáva však kontroly vyplnenosti a kvality údajov (aj stotožnením). Nedostatočné vyplnenie údajov a zistené nezahody sú riešené priamo so školami. Motiváciou škôl pre zvyšovanie dátovej kvality je to, že podklady z RIS slúžia ako podklad pre ich normatívne financovanie. Správnosť údajov je čiastočne overovaná vďaka tomu, že MŠVVaŠ SR poskytuje údaje DŽP na základe princípu „jedenkrát a dost“ niektorým ďalším štátnym orgánom, ktoré ich využívajú na účely svojej agendy (napr. výplata rodinných prídavkov, daňový bonus, atď.). V prípade zlej evidencie žiakov môže dôjsť ku nesprávnym akciám týchto orgánov (napr. nevyplateniu rodinných prídavkov zákonným zástupcom žiaka). V prípade, ak zákonný zástupca zistí, že boli organizácii poskytnuté nesprávne údaje, je nutné opraviť

zdrojové údaje v ŠIS školského zariadenia a údaje aktualizovať manuálnym spustením aktualizácie do RIS.

MŠVVaŠ SR pravidelne v ročných intervaloch publikuje 24 datasetov na portáli otvorených údajov data.gov.sk. Kvalita týchto datasetov nie je vyhodnocovaná.

6.2 Model reportovania

Jednotlivé druhy reportov o kvalite dát a informácie v nich obsiahnuté sa líšia podľa typov merania dátovej kvality:

- a) **Profilovanie údajov** – slúži na podrobné pochopenie dát. Výsledky získané z tohto merania sa používajú **v prehľadových správach z profilovania údajov**.
- b) **Informatívne merania kvality údajov** – technika na určenie úrovne kvality údajov v akomkoľvek objekte evidencie (dátový súbor s vybranými atribútmi) s určením ľubovoľnej kombinácie parametrov a indikátorov kvality údajov pre účely merania. Tento typ merania kvality dát možno použiť kedykoľvek a kdekoľvek podľa potreby. Výsledky získané z tohto merania sa používajú **v správach z informatívneho merania kvality**.
- c) **Monitorovanie/vyhodnotenie zmeny stavu kvality dát** – sledovanie zmien na meranom objekte evidencie po ukončení projektu na zlepšenie dátovej kvality na zvolených atribútoch je činnosťou, ktorá sa bežne vykonáva po ukončení projektov zlepšovania dátovej kvality. Merania by sa mali vykonávať pravidelne a súbor parametrov a ukazovateľov dátovej kvality údajov by sa nemal výrazne líšiť od tých z predchádzajúceho súvisiaceho projektu na zlepšenie dátovej kvality. Každé monitorovacie meranie je teda súčasťou skupiny monitorovacích meraní patriacich pod ukončený projekt zameraný na zvyšovanie dátovej kvality. Výsledky získané z tohto merania sa používajú **v hodnotiacich správach projektov** zlepšovania kvality údajov.

Odporúčaný postup pri tvorbe reportov dátovej kvality by sa mal skladať z nasledujúcich 3 krokov:

Zhromaždenie podkladov na hodnotenie

Zhromaždiť všetky dostupné podklady, ktoré budú slúžiť pre účely tvorby reportu o dátovej kvalite. Rozsah potrebných podkladov je daný samotnou štruktúrou reportu, ktorého príklad sa nachádza v kapitole 6.2.1. Podklady môžu mať podobu aj informácií mimo informačného systému, v podobe vlastných poznámok.

Za minimálne potrebné podklady na vytvorenie reportu dátovej kvality sú považované:

- Výsledky z merania
- Vzorový report dátovej kvality (vytvorený dátovou kanceláriou MIRRI)
- Report z predchádzajúceho merania (ak existuje)
- Zoznam zmien zavedených v objekte evidencie od posledného merania
- Zoznam použitých biznis pravidiel
- Základné informácie o zdrojových registroch a objekte evidencie
- Rôzne iné podklady potrebné pre vytvorenie reportu špecifického objektu evidencie

Vypracovanie reportu dátovej kvality

Analyzovať zhromaždené podklady pre účely hodnotenia, pričom rozsah vykonávanej analýzy je daný požadovanou štruktúrou reportu dátovej kvality. Pri vypracovaní je potrebné analyzovať všetky výsledky z merania a taktiež aj zohľadniť reporty z predchádzajúcich meraní pokiaľ existujú. Report bude vytváraný na základe všeobecne odporúčanej štruktúry a obsahu vypracovaného dátovou kanceláriou (MIRRI).

V reporte bude taktiež potrebné sa zamerať na kritické prípady nekvality, na základe ktorých je dôležité navrhnuť opatrenia na zlepšenie dátovej kvality. Počas vypracovania reportu je zároveň dôležité zvážiť jednotlivé obsahové časti daného reportu na základe typu merania, z ktorého je pripravovaný a meraného objektu evidencie.

Vytvorené reporty by mali pokrývať všetky hlavné parametre kvality údajov – presnosť, konzistentnosť, správnosť, kompletnosť, unikátnosť, aktuálnosť a strojová spracovateľnosť. Avšak v špecifických prípadoch môže byť rozhodnuté, že pri meraní dátovej kvality vybraného objektu evidencie nie je potrebné vykonať meranie so všetkými hlavnými parametrami kvality údajov. K takémuto rozhodnutiu môže dôjsť práve vtedy, keď neexistuje potreba žiadnej zo zainteresovaných strán na realizovanie merania podľa špecifického parametra, takéto rozhodnutie by malo byť priamo obsiahnuté v reporte.

Sumarizácia a odsúhlasenie aktivít na zlepšenie dátovej kvality a ostatného obsahu reportu dátovej kvality

Pred zverejnením reportu dátovej kvality vybraného objektu evidencie je potrebné jeho schválenie. Zhotoviteľ reportu zasiela vytvorený report na schválenie na to určenej osobe (v rámci OVM), ktorá rozhodne, či je možné report zverejniť, alebo ho vráti zhotoviteľovi na prepracovanie. V prípade, že je správa vrátená musia byť špecifikované konkrétne pripomienky, ktoré majú byť zapracované. Po zapracovaní pripomienok a následnom schválení reportu môže byť tento odovzdaný na zverejnenie.

Zverejňovanie

Z hľadiska vyššie uvedených typov reportov rozlíšime reporty pravidelné a ad-hoc. **Pravidelné reporty** sú typicky reportami z informatívneho merania kvality alebo monitorovacieho merania, či už sa jedná o pravidelné reportovanie kvality dát v registroch (objektoch evidencie, OE) alebo správy o kvalite špecifických dátových produktov (štatistické produkty, analytické produkty a pod.). Tieto sa spravidla zverejňujú v pravidelných intervaloch, daných periódou zverejňovania dátových produktov alebo dohodnutou periódou monitorovacieho merania kvality OE. **Ad-hoc reporty** sú spravidla viazané na míľniky (udalosti) resp. aktivity cyklu zlepšovania kvality údajov. Jedná sa teda o projektovo orientované reportovanie – reporty z profilácie a hodnotiace reporty.

Určiť frekvenciu pravidelného monitorovacieho merania kvality a zverejňovania príslušných reportov pre OE všeobecne nie je možné, nakoľko môžu existovať špecifiká pre určitý objekt evidencie, na základe ktorých určitá frekvencia môže byť primárna alebo naopak priveľká na to, aby malo vykonané meranie a vytvorenie reportu požadovaný efekt. Z toho teda vyplýva, že by mala byť **určená spoločná perióda**, ktorá bude platiť všeobecne pre všetky objekty evidencie (**napr. jedenkrát ročne**),

avšak gestor v rámci riadenia dátovej kvality (MIRRI) môže určiť, že meranie môže byť vykonané aj s vyššou frekvenciou, pokiaľ je to opodstatnené, na základe biznis alebo technických potrieb a možností.

Zo súčasných skúseností vieme, že pravidelné reporty dátovej kvality vykonávané štatistickým úradom a jednorazové reporty z pilotného merania vykonané MIRRI boli uverejňované na datalab.digital²⁹ a na príslušnom webe OVM, ktoré meranie vykonalo. Nateraz je tento model umiestnenia reportov dostačujúci, avšak v budúcnosti môžu vzniknúť ďalšie požiadavky na umiestnenie zverejňovania reportov. Odporúčame pripraviť používateľsky prívetivejšie rozhranie na prezeranie reportov v rámci datalab.digital.

6.2.1 Návrh obsahu reportov dátovej kvality

Obsah reportu sa môže čiastočne líšiť pre jednotlivé objekty evidencie a zároveň sa ich obsah môže líšiť aj podľa druhu merania, z ktorého sú vytvárané. Vyššie sú zadané 3 typy merania, z ktorých priamo vyplývajú druhy reportov a každý z týchto reportov môže vyžadovať trochu odlišné zloženie. V prípade reportu z profilovania údajov a monitorovania/vyhodnotenia zmeny stavu kvality dát je potrebné obsiahnuť prevažnú väčšinu z nižšie uvedených bodov obsahového zloženia reportov. Pri reporte z informatívneho merania nie je nutné obsiahnuť všetky obsahové body zloženia reportov, v tomto type reportu by teda mali byť najmä obsiahnuté tie informácie, ktoré poskytnú pohľad na zhodnotenie špecifických oblastí dátovej kvality, na základe ktorých bolo inicializované informatívne meranie.

Obsahové zloženie dátovej kvality:

- informácie o zdrojových registroch, z ktorých sa skladá,
- metodológia tvorby registra a podrobné informácie o údajoch, ktoré sa v ňom nachádzajú,
- periodicita aktualizácii registra a jeho úprav,
- informácie o využívaných dátových slovníkoch,
- informácie o problémoch kvality spojených so zdrojovými registrami,
- vyhodnotenie dátovej kvality objektu evidencie na základe vybraných biznis pravidiel a merateľných ukazovateľov – zohľadnenie nárokov konzumentov dát, na základe ktorých boli merané biznis pravidlá a ukazovatele podľa ich potrieb,
- vyhodnotenie merania dátovej kvality v rôznych úrovniach detailu a spresnenia podľa vopred zadaných požiadaviek,
- informácie o možných hrozbách, ktoré môžu vzniknúť pri používaní objektu evidencie pri rozsahu nameranej nekvality,
- vytvorenie odpovedí na často kladené otázky v spojitosti s daným objektom evidencie a jeho dátovou kvalitou,
- odporúčania a návrhy na zlepšenie dátovej kvality z dátového a vecného hľadiska.

Karty kvality údajov (informačné panely alebo tiež dashboardy)

Dashboardy sú jedným zo spôsobov, ktorým je možné zobraziť výsledky z merania kvality údajov v interaktívnej forme a podporiť zainteresované strany (vlastníkov

²⁹ <https://datalab.digital/hodnotiace-spravy/>

zdrojových registrov) pri riadení kvality pomocou grafických vizualizácií metrík. Vizualizované metriky pomáhajú pri prijímaní rozhodnutí a dopomáhajú aj pri:

- Prispôsobovaní reportov pre potreby každého zainteresovaného subjektu, vzhľadom na rozsah údajov, za ktoré je subjekt zodpovedný a/alebo ich konzumuje.
- Jasnom zdôraznení rozsahu, výsledkov a časových bodov.
- Zvýraznení problémových oblastí v dátovej kvalite
- Realizovaní hĺbkovej analýzy, ktorá v prípade potreby poskytne viac detailov.
- Zobrazovaní výsledkov v časovom bode a aj časovom vývoji.

6.2.2 Návrh prístupu k aktualizácii a reportovaniu

Dosiahnutie reportovania dátovej kvality v štandardizovanej a efektívnej forme vyžaduje zapojenie do procesu tvorby reportov gestora dátovej kvality (MIRRI) v pozícii aktéra poskytujúceho metodologickú podporu a jednotlivé OVM, ktoré budú vykonávať merania dátovej kvality a vytvárať reporty z týchto meraní. Zodpovednosti a povinnosti pri takomto prístupe by boli rozdelené nasledovne:

MIRRI

- V pozícii gestora agendy pre dátovú kvalitu **stanoví rámec pre reporty dátovej kvality**, ktoré budú vytvárané jednotlivými OVM.
- **Vytvorí vzorový report**, ktorý bude spĺňať požiadavky reportov dátovej kvality zadefinovaný v prvom kroku podľa odporúčaní v tejto kapitole.
- **Komunikuje** s jednotlivými OVM **požiadavky** a poskytuje im metodickú podporu pri tvorbe reportov.
- **Usmerňuje** jednotlivé **OVM** pri zavádzaní modifikácii do štruktúry reportov dátovej kvality pre zachovanie konzistentnosti reportov.
- Kontroluje zverejňovanie reportov OVM a zverejňuje ich na datalab.digital.
- **Vyhodnocuje celkový stav dátovej kvality** vo verejnej správe na základe reportov jednotlivých OVM a **pripravuje akčné plány** v koordinácii s OVM na ich **zlepšenie**

OVM

- **Pripravuje reporty dátovej kvality** na základe vykonaných meraní na jednotlivých objektoch evidencie podľa rámca stanoveného MIRRI
- **Určuje frekvenciu vykonávania merania** a následného zverejňovania reportov, podľa požiadaviek vyplývajúcich z existujúcich biznis požiadaviek.
- **Zverejňuje reporty dátovej kvality na svojich webových stránkach** a informuje MIRRI o zverejnení nového reportu.
- **Vytvára systém na publikovanie vytvorených reportov a ich schvaľovania**
- **Koordinuje svoje úsilie s MIRRI** pri: zmenách štruktúry reportov a v otázkach súvisiacich s metodológiou prípravy reportov
- **Kooperuje s MIRRI pri príprave akčných plánov** na zlepšenie dátovej kvality

7 Definícia požiadaviek na ISVS pre automatizované meranie kvality

7.1 Všeobecné požiadavky automatizácie merania DK

Automatizácia merania kvality dát v prostredí verejnej správy je dôležitým nástrojom na zabezpečenie dôveryhodnosti a presnosti informácií. Zavádzanie automatizovaných procesov umožňuje rýchle a systematické vyhodnocovanie kvality dát v reálnom čase, čo vedie k zlepšeniu efektívnosti a účinnosti správy dát. Automatizácia tiež minimalizuje manuálne zásahy a znižuje riziko ľudských chýb, čím šetrí čas a prostriedky. Je však dôležité k automatizácii pristupovať vyvážené a vziať do úvahy existujúce IS, želaný cieľový stav a súvisiace obmedzenia. Nie všetky procesy je možné automatizovať a plná automatizácia nemusí byť nákladovo efektívna. Navyše, akýkoľvek systém automatizácie merania DK by mal spĺňať minimálne požiadavky uvedené v nasledujúcej tabuľke.

Tabuľka 12 Zoznam 10 základných požiadaviek na automatizovaný výpočet parametrov kvality dát

Schopnosť	Popis
Správa metadát	Systém by mal poskytovať možnosti na dokumentovanie metadát vrátane mien, typov, zdieľaných referenčných údajov, aby sa mohli konzistentne používať a spravovať.
Profilovanie dát	Systém by mal poskytovať možnosti na analýzu údajov, s cieľom nájdenia vzorov/pravidiel, chýbajúcich hodnôt a iných základných štatistických charakteristík (uvedené vyššie). Mal by poskytovať prehľady, ktoré pomôžu identifikovať problémy s kvalitou údajov. Dá sa použiť aj pri hodnotení/meraní kvality dát.
Monitorovanie	Systém by mal poskytovať možnosti implementovať kontroly na zabezpečenie súladu so zavedenými biznis pravidlami, ktoré definujú kvalitu údajov a merať mieru ich súladu.
Dashboarding	Systém by mal poskytovať sadu informačných panelov (dashboards), ktoré dokážu vizualizovať agregovaný stav monitorovaných údajov, generovať upozornenia na problémy a poskytnúť rozsiahle možnosti usporiadania na potreby používateľa (customizácia)
Verifikácia dát	Systém by mal poskytovať možnosti overovať údaje na základe definovaných pravidiel a integrovať sa na centrálné slovníky a dátové modely, obsahujúce tieto pravidlá tak, aby boli prepustené iba tie údaje, ktoré spĺňajú požadované kritériá overenia.
Matching	Systém by mal poskytovať možnosti identifikácie a zlučovania zhodných záznamov na účely deduplikácie nadbytočných záznamov na základe definovaných kritérií zhody medzi nimi alebo v rámci nich

Schopnosť	Popis
Obohatenie	Systém by mal poskytovať možnosti zvyšovať hodnotu interných údajov pridaním atribútov z externých zdrojov (napríklad demografických alebo geografických údajov, pridanie URI liniek napr. na referenčné registre a pod.)
Čistenie	Systém by mal poskytovať možnosti upravovať hodnoty tak, aby spĺňali obmedzenia domény, obmedzenia integrity alebo iné biznis pravidlá, ktoré definujú dostatočnú kvalitu údajov (napr. oprava neplatných hodnôt, pokiaľ to je vecne a technicky možné).
Štandardizácia a parsing	Systém by mal poskytovať možnosti na rozklad textových polí na časti podľa noriem, užívateľom definovaných pravidiel, zadaných vzorov alebo vzorových hodnôt.
Napojenie na dátové zdroje	Systém by mal poskytovať možnosti na napojenie na rôzne typy dátových zdrojov s možnosťou flexibilnej transformácie a úpravy dát pre analytické potreby a čistenie (funkcia ETL).

7.2 Požiadavky na automatizáciu merania DK v kontexte ISVS

V rámci plánovania a vybudovania informačného systému verejnej správy je pri vytváraní postupov automatizácie procesov merania DK potrebné vykonať nasledujúcich **5 krokov**.

- 1 Inventarizácia a audit existujúcich techník a riešení používaných na účely súvisiace s dátovou kvalitou v zainteresovaných OVM (napr. použité nástroje, postupy, štandardy a pod.).
- 2 Posúdenie používaných riešení vzhľadom na všeobecné (bežné resp. centrálné stanovené) požiadavky na procesy dátovej kvality, vrátane kritérií merania dátovej kvality.
- 3 Posúdenie súčasného stavu používaných riešení a nástrojov a stanovenie možností ich ďalšieho využitia (modernizácia, migrácia, vyradenie z prevádzky a pod.).
- 4 Vyhodnotenie technologického prostredia – napríklad požadované integrácie a rozhrania API, prekrývanie riešení a naopak odhalené medzery v automatizácii.
- 5 Stratégia smerom k spoločnému riadeniu kvality dát a automatizácii merania, vrátane komplexného mapovania stavu pri zavádzaní ISVS, napr. využitie centrálnych ponúkaných možností (napr. Platform-as-a-Service – PaaS v rámci projektu CIP) alebo využitie existujúcich riešení v OVM, pokiaľ také existujú, a ich napojenie v rámci spoločného (centrálneho) dátového modelu a kritérií kvality.

Zároveň je potrebné vziať do úvahy špecifické požiadavky vyplývajúce z odporúčaných postupov a merania dátovej kvality v prostredí verejnej správy.

Validácia a kontrola údajov

Informačné systémy by mali byť schopné vykonávať automatizovanú validáciu a kontrolu údajov na základe definovaných pravidiel a preddefinovaných štandardov.

Tieto pravidlá a štandardy by mali zahŕňať overenie formátu, hodnoty a vzťahov medzi údajmi prioritne vo väzbe na referenčné registre (RPO, RA, RFO) prípadne ďalšie zdrojové registre.

Detekcia chýb a nekonzistencií

Systémy by mali byť schopné identifikovať chyby, nekonzistencie a duplicity v dátach a automaticky generovať upozornenia alebo opravné akcie. Môžu napríklad kontrolovať jedinečné identifikátory, platnosti dátumu a času, správne napojenie na referenčné údaje a podobne.

Monitorovanie a správa kvality

Systémy by mali byť doplnené nástrojmi na monitorovanie a správu kvality dát. To zahŕňa sledovanie kľúčových ukazovateľov výkonnosti (KPI), generovanie správ a prehľadov o kvalite dát a poskytovanie informácií o aktuálnom stave kvality dát.

Automatická korekcia a doplnenie

V prípade identifikácie chýb alebo nekonzistencií by systémy mali byť schopné vykonávať automatickú korekciu a doplnenie údajov na základe definovaných pravidiel a algoritmov. To znižuje potrebu manuálnej intervencie a zvyšuje rýchlosť a presnosť procesu. Samozrejme, táto požiadavka môže mať k konkrétnym situáciám jednak legislatívne limity a takisto limity, vyplývajúce z prípadnej neexistencie jednoznačných pravidiel na korekciu/doplnenie údajov. Aj týmto je zdôraznená potreba zabezpečiť kvalitu dát na vstupe, ako je poznamenané vyššie.

Integrácia s existujúcimi systémami

Systémy a nástroje pre správu kvality dát by mali byť schopné integrovať sa s existujúcimi systémami verejnej správy, aby mohli získavať prístup k relevantným dátam a poskytovať integrované riešenia aj pre meranie kvality.

Auditovateľnosť a transparentnosť

Systémy by mali byť schopné zaznamenávať a uchovávať informácie o vykonaných kontrolách, opravách a zmenách v dátach aj v záujme auditovateľnosti procesov a transparentnosti voči vnútornej a externej kontrole.

Bezpečnosť dát

Požiadavky na zabezpečenie dát, ako je šifrovanie, prístupové práva a ochrana pred neoprávneným prístupom, sú kľúčové pre informačné systémy verejnej správy aj z hľadiska automatizácie merania a správy kvality dát.

7.3 Odporúčania pre automatizáciu merania DK v ISVS

Okrem vyššie uvedených požiadaviek na ISVS pre automatizovaný výpočet parametrov kvality dát je potrebné zvážiť aj niekoľko dôležitých aspektov špecifických pre informačné systémy merania DK vo verejnej správe.

Primárne hodnotenie stavu by malo byť zamerané na stav kmeňových/referenčných dát a metadát v prostredí OVM a na zmysluplné mapovanie požiadaviek na výmenu dát v rámci centrálného dátového modelu na požiadavky na dáta v prostredí OVM, ktoré typicky býva dostatočne špecifické na to, aby sa dali požiadavky aplikovať bez prispôbenia.

Vo väčšine prípadov je veľmi ťažké navrhnúť všeobecné požiadavky na sémantickú kvalitu údajov relevantných pre väčšinu OVM, preto by sa mala najprv venovať pozornosť syntaktickej kvalite údajov a sémantickú kvalitu riešiť prioritne pri vzniku/úprave údajov priamo na vstupe.

V rámci akéhokoľvek nového nasadenia alebo modernizácie transakčného systému v OVM by sa mali povinne zahrnúť požiadavky na automatickú kontrolu kvality údajov (súlady s pravidlami) a jej meranie s primárnym cieľom zabezpečenia kvality (sémantickej i syntaktickej) čo najbližšie vstupu údajov do systému.

Súbor pravidiel a KPI pre dátovú kvalitu by mal postihovať tak špecifické biznis požiadavky konkrétneho OVM, ako aj zdieľané požiadavky súvisiace s výmenou údajov (požiadavky konzumentov). Špecifické biznis požiadavky OVM na údaje vyplývajú z vecnej podstaty spravovanej agendy (štrukturálna i sémantická kvalita údajov pre použitie v konkrétnej inštitúcii). OVM je zodpovedný a riadi tieto procesy vrátane merania kvality, identifikácie, realizácie a vyhodnocovania úspešnosti nápravných opatrení. Zdieľané (spoločné) požiadavky na dátovú kvalitu sú potrebné na výmenu údajov, ich ďalšie spracovanie inými spotrebiteľmi (konzumentmi) a prípravu otvorených údajov. Zoznam takýchto požiadaviek je pod dohľadom centrálného orgánu, správcu NSZÚ (Dátová kancelária MIRRI) a meranie kvality by sa malo vykonávať tak na strane OVM (aby dáta spĺňali tieto požiadavky), ako aj centrálné (ako prevencia šírenia chýb v rámci spracovateľského reťazca).

8 Registre

8.1 Register tabuliek

Tabuľka 1 Celkový stav iniciatív podľa oblastí	<u>76</u>
Tabuľka 2 Centrálna úroveň - vyhodnotenie realizácie opatrení	8
Tabuľka 3 Rezortná úroveň - vyhodnotenie realizácie opatrení	9
Tabuľka 4 Hodnotenie 7-krokového procesu informatívneho merania DK	<u>2415</u>
Tabuľka 5 Hodnotenie 10-krokového procesu na zlepšenie kvality údajov	<u>2616</u>
Tabuľka 6 Pôvodné a nové odporúčania pre meranie DK	<u>4736</u>
Tabuľka 7 Prioritné merateľné ukazovatele	<u>5342</u>
Tabuľka 8 Faktory, ktoré ovplyvňujú výber nástroja pre analýzu DK	<u>5847</u>
Tabuľka 9 Piliere správy údajov	<u>5948</u>
Tabuľka 10 Výhody a nevýhody prístupov k výberu riešenia riadenia DK	<u>6049</u>
Tabuľka 11 Zrelosť riadenia DK	<u>6554</u>
Tabuľka 12 Zoznam 10 základných požiadaviek na automatizovaný výpočet parametrov kvality dát.....	<u>7463</u>

8.2 Register obrázkov

Obrázok 1 Proces zlepšovania dátovej kvality	6
Obrázok 2 7-krokový proces pre informatívne meranie dátovej kvality	<u>4837</u>

Contact us

Rudolf Sedmina
partner

E 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

© 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.