Operačný program
Efektívna
verejná správa

Európska únia
Európsky sociálny fond

Metodika merania dátovej kvality vo verejnej správe

Projekt

Zlepšenie využívania údajov vo verejnej správe

Zmluva o dielo č. 321/2018

Výstup č.1

KPMG

ÚRAD PODPREDSEDU VLÁDY SR
PRE INVESTÍCIE
A INFORMATIZÁCIU

Tento projekt je podporený z Európskeho sociálneho fondu.

© 2019 KPMG, Slovensko Advisory, k.s., the Slovak member firm of KPMG International Cooperative ("KPMG International"), a Swiss entity.
All rights reserved.

Zoznam skratiek

UPVII	Úrad podpredsedu vlády pre investície a informatizáciu Slovenskej republiky
RACI	Matica zodpovednosti (R – zodpovedný vykonávateľ, A – zodpovedný vlastník, C – konzultuje, I – je informovaný)
SPOC	Single point of contact (centrálny komunikačný uzol)
ERP	Enterprise resource planning (Systémy plánovania zdrojov)
SQL	Structured query language (štruktúrovaný dotazovací jazyk)
AS-IS	Súčasný východiskový stav
TO-BE	Cieľový stav
OVM	Orgán verejnej moci
ISVS	Informačný systém verejnej správy
RFO	Register fyzických osôb
RPO	Register právnických osôb
RÚ	Register úpadcov
RA	Register adries
IČO	Identifikačné číslo organizácie
NA	Neaplikovateľné
GDPR	Global data protection regulation (Všeobecné nariadenie na ochranu osobných údajov)
HW	Hardvér
SW	Softvér
DK	Dátová kvalita

© 2019 Firm name. All rights reserved.

Obsah

© 2019 Firm name. All rights reserved.

© 2019 Firm name. All rights reserved.

1 Manažérske zhrnutie

Prečo riešime dátovú kvalitu ?

Cieľom zlepšovania dátovej kvality nie je dátová kvalita sama o sebe, ale vyššia kvalita života a lepšie kvalitnejšie služby verejnosti - ktoré sú silno prepojené a závislé práve na primeranej dátovej kvalite. Dátová kvalita nie je abstraktný izolovaný cieľ, ale cesta a základňa pre zvyšovanie kvality služieb, produktov a komunikácie s ňou spojených.

Aký je zmysel dátovej kvality ?

Dáta sú realita a holé fakty. Čím kvalitnejšie dáta máme, tým hodnovernejšie sú informácie a znalosti z nich odvodené. Cieľom je nielen zvýšenie prvotných ukazovateľov dátovej kvality, ako sú presnosť, aktuálnosť a úplnosť dát, ale aj ich celková konzistentnosť, vierohodnosť, použiteľnosť, spoľahlivosť a integrita.

Prečo meriame dátovú kvalitu ?

To, čo chceme systematicky zlepšovať, musíme aj nejakým spôsobom vyhodnocovať merať a kvantifikovať. Aby sme mohli objektivizovať súčasný aj plánovaný stav a vedeli čo najobjektívnejšie sledovať trendy a dynamiku vývoja.

Prečo práve teraz ?

Dátová kvalita posledné roky získava stále väčšiu pozornosť a dôležitosť pre exponenciálny nárast objemu štruktúrovaných aj neštruktúrovaných dát a pre práve prebiehajúcu digitálnu transformáciu. V tomto prostredí už nestačí doteraz obvyklý intuitívny subjektívny ad hoc prístup ku kvalite dát, ale je nevyhnutné zaviesť koncepčný prístup založený na overenej metodike.

V prvej fáze programu zvyšovania dátovej kvality ide hlavne o diagnózu : objektivizáciu, kvantifikáciu a merateľnosť súčasného stavu a sledovanie postupnej riadenej konvergencie do žiadaného cieľového stavu dátovej a informačnej kvality.

V druhej fáze programu sa priorita bude presúvať aj na terapiu : čistenie a konsolidáciu dát, deduplikáciu duplicitných záznamov, postupné reálne zlepšovanie parametrov a ukazovateľov dátovej kvality.

V tretej fáze programu sa ťažisko bude presúvať na pro-aktivitu a trvalú udržateľnosť koncepcie dátovej kvality. Budú nastavené pravidlá pre kvalitu dátových vstupov do systému na samotnom začiatku procesu ako aj princípy dátovej integrácie, validácie a transformácie.

Riadenie dát a ich kvality je tímový „šport“. Je to nielen o dátových technológiách a procesoch, ale hlavne o ľuďoch, o ich prístupe a zodpovednostiach. Je to o orchestrácii a správnom nastavení rolí a ich kompetencií v celom ekosystéme riadenia a spracovania dát a informácií.

© 2019 Firm name. All rights reserved.

2 Riadenie dátovej kvality v informačných systémoch verejnej správy

2.1 Úvod k riadeniu dátovej kvality v informačných systémoch verejnej správy

Pre riadenie dátovej kvality v informačných systémoch verejnej správy je v nasledujúcich častiach popísaná metodika, ktorá reprezentuje systematický prístup ako zaviesť, merať a zároveň zlepšovať dátovú kvalitu. Metodika kombinuje konceptuálny framework pre pochopenie dátovej kvality a 10 krokový proces, ktorý poskytuje inštrukcie, techniky a osvedčené postupy. Slúži ako ucelený framework poskytujúci procesy, aktivity a techniky zabezpečujúce dátovú kvalitu.

2.2 Biznis požiadavky na riadenie, správu a meranie dátovej kvality

V nasledujúcej tabuľke (Tabuľka 1) sú rozpísané biznis požiadavky, ktoré boli definované v počiatkoch tvorenia metodiky merania dátovej kvality vo verejnej správe. Ich definícia sa prioritne zameriava na samotné meranie a doplňuje ju dodatočný dôraz na správu a riadenie dátovej kvality.

P.č.	Oblasť	Popis biznis požiadavky
1	Objekt merania dátovej kvality	Pri rozsahu požiadaviek majú byť presne špecifikované data-sety, ktoré budú predmetom skúmania, merania a riadenia
2	Objekt merania dátovej kvality	K daným data-setom (objekt merania) je potrebné poskytnúť všetky dostupné business a technické metadáta (dáta o dátach) a potrebné dokumentácie
3	Procesy dátovej kvality	Potreba pomenovania konkrétneho business problému a cieľa, ktorý sa má projektovo alebo programovo riešiť
4	Procesy dátovej kvality	Je potrebné definovať mieru detailu merania (či pôjde o meranie na úrovni jednotlivých záznamov, riadkov, stĺpcov, agregovaných celkov...)
5	Procesy dátovej kvality	Je dôležité špecifikovať, či ide len o merania stavu izolovaného dáta setu / registra alebo sa žiada posúdenie v širšom kontexte celého ekosystému zdrojových systémov dát, ich spracovania integrácie a transformácie
6	Procesy dátovej kvality	Návrh metodiky pre riadenie a meranie dátovej kvality má byť orientovaný na procesy
7	Procesy dátovej kvality	Metodika by mala zachytiť komplexný popis procesov pre potreby riadenia, správy a merania dátovej kvality
8	Procesy dátovej kvality	Metodika má vychádzať z „best practice“ riadenia dátovej kvality vo svete

© 2019 Firm name. All rights reserved.

P.č.	Oblasť	Popis biznis požiadavky
9	Nástroj merania dátovej kvality	Je potrebné špecifikovať požiadavky na nástroj pre meranie dátovej kvality
10	Nástroj merania dátovej kvality	Je nutné špecifikovať súčasnú aj cieľovú platformu a paradigmu riadenia (či pôjde o on premise architektúru alebo Cloud koncept (a keď Cloud, tak ktorý : privátny, verejný, hybridný)
11	Parametre a ukazovatele dátovej kvality	Ak existujú cieľové, prahové alebo benchmarkingové hodnoty a KPIs (kľúčové výkonnostné ukazovatele), treba ich uviesť a konkretizovať
12	Parametre a ukazovatele dátovej kvality	Je potrebné definovať minimálne nasledovné parametre, ktoré reflektujú dátovú kvalitu: - Presnosť (čistota) - Kompletnosť - Aktuálnosť - Unikátnosť - Referenčná integrita - Strojová spracovateľnosť - Konzistentnosť - Správnosť
13	Parametre a ukazovatele dátovej kvality	Každý parameter musí mať definované ukazovatele dátovej kvality, ktoré sa budú pri vyhodnocovaní používať
14	Parametre a ukazovatele dátovej kvality	Je potrebné definovať cieľové hodnoty pre ukazovatele dátovej kvality, ktoré určujú ambíciu pre dátovú kvalitu vo verejnej správe v nasledovnom období
15	Parametre a ukazovatele dátovej kvality	Je potrebné definovať prioritizáciu pre dimenzie a parametre dátovej kvality. Či sa treba primárne zamerať na úplnosť, presnosť, aktuálnosť alebo na integritu dát
16	Dátový kurátor	Metodika merania dátovej kvality vo verejnej správe má podporiť a usmerniť jednotlivých Dátových kurátorov v ich zodpovednosti za správu dát a ich kvalitu na ich rezorte. Dátoví kurátori by mali dostať jednotnú metodiku merania dátovej kvality a jej trendov tak, aby ju vedeli použiť na reálne dáta v ich kompetencii

Tabuľka 1: Biznis požiadavky na riadenie, správu a meranie dátovej kvality

© 2019 Firm name. All rights reserved.

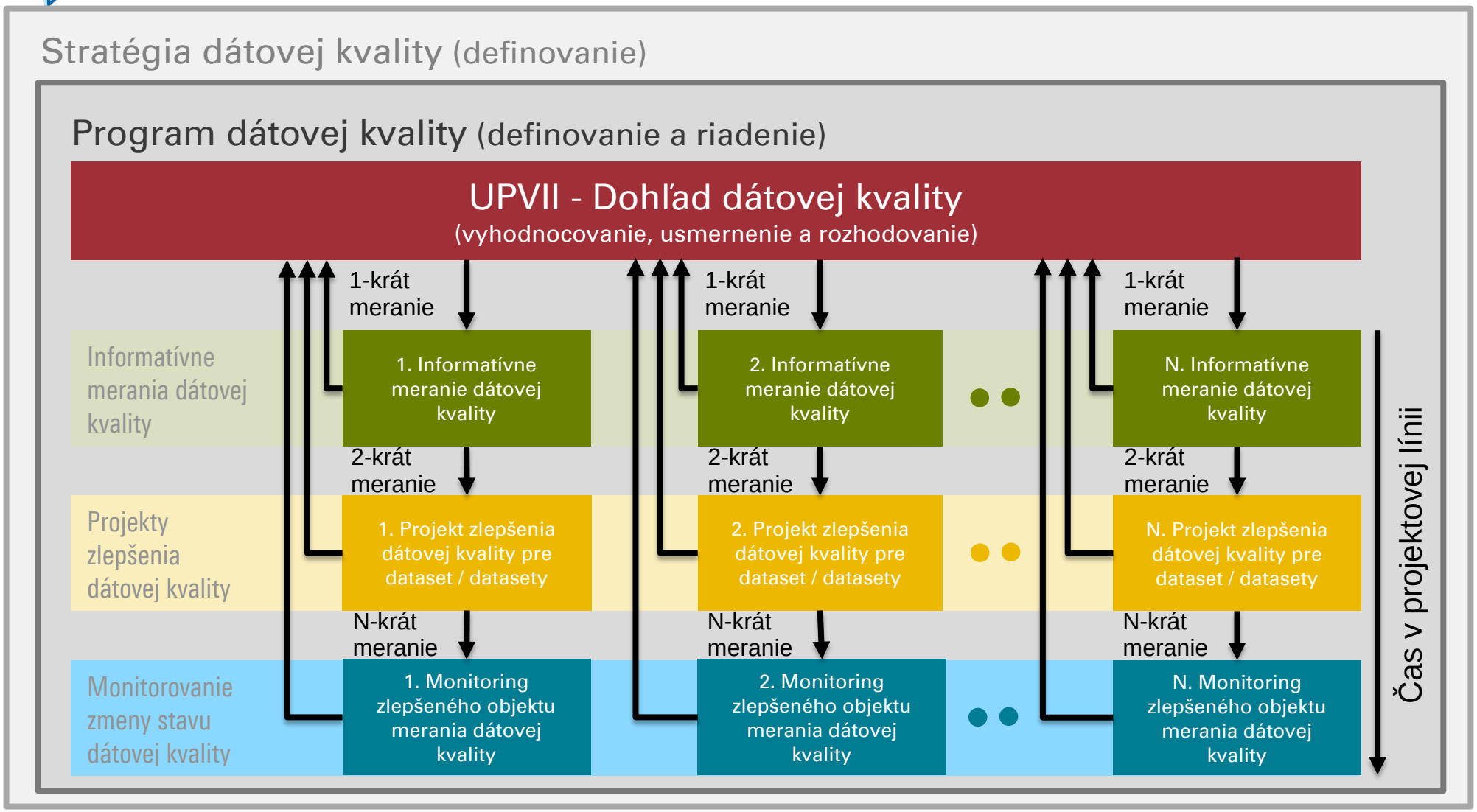
2.3 Návrh metodiky pre riadenie, správu a meranie dátovej kvality

2.3.1 Komplexný popis procesov pre potreby riadenia a správy dátovej kvality

2.3.1.1 Koncepcia riešenia - Zavádzanie dátovej kvality vo verejnej správe

Koncepcia riešenia pre zavádzanie dátovej kvality vo verejnej správe je zobrazená na obrázku (Obrázok 1) a vychádza z dvoch hlavných stavebných blokov. Prvým stavebným blokom je Stratégia dátovej kvality a druhým je definovanie a následné riadenie Programu dátovej kvality.

© 2019 Firm name. All rights reserved.



Obrázok 1: Koncepcia riešenia dátovej kvality vo verejnej správe

Stratégiu dátovej kvality je potrebné zadefinovať hneď na začiatku celej iniciatívy. Jej definícia má byť v súlade s časovým plánom ďalších aktivít súvisiacich s dátovou kvalitou vo verejnej správe. Obvykle sa stratégia definuje na dlhšie časové obdobie troch alebo piatich rokov. Pri definovaní stratégie je tento časový údaj smerodajný a vo svojej podstate charakterizuje aj odhad zmien v prostredí verejnej správy. Nemá zmysel definovať stratégiu na príliš dlhé časové obdobie pri prostredí, ktorého dynamika môže narušiť kontinuitu a zmysel definovaných cieľov v stratégií. Tieto ciele môžu byť potom neaktuálne a v horšom prípade nesprávne. Stratégiu dátovej kvality je potrebné nasledovať a po uplynutí jej časového obdobia následne aktualizovať. Nová stratégia by mala vychádzať zo zhodnotenia predošlej stratégie.

Program dátovej kvality je v tomto prípade súbor všetkých aktivít spojených s dátovou kvalitou vo verejnej správe. Riadenie programu vychádza zo stratégie dátovej kvality a obsahuje aktivity, ktoré by sa dali zhrnúť do štyroch hlavných oblastí:

- **Dohľad dátovej kvality**: Sú v ňom zahrnuté aktivity dohľadu, ktorých hlavná náplň práce sú vyhodnocovanie, usmernenie a rozhodovanie v spojení s dátovou kvalitou. Tieto aktivity nezahŕňajú samotné meranie dátovej kvality. Meranie však jednoznačne iniciujú a zbierajú aj výsledky týchto meraní.

- **Informatívne merania dátovej kvality:** Tento druh merania je vybratý ako hlavný nastroj zistenia úrovne dátovej kvality na ľubovoľnom objekte merania (dataset(y) s vybranými atribútmi) s určením ľubovoľnej kombinácie parametrov a ukazovateľov dátovej kvality pre potreby merania. Táto sonda dátovej kvality môže byť použitá v ľubovoľnom čase a na ľubovoľnom mieste podľa potreby.

- **Projekty zlepšenia dátovej kvality:** Tieto projekty sú určené ako hlavný nástroj zlepšenia dátovej kvality pre konkrétny objekt merania (dataset(y) s vybranými atribútmi). Ide o komplexnejší súbor aktivít, ktorých hlavným cieľom je zlepšiť dátovú kvalitu na základe presne definovaných cieľových hodnôt.

- **Monitorovanie zmeny stavu dátovej kvality:** Toto sledovanie zmien na zlepšenom objekte merania dátovej kvality je prirodzeným dôsledkom ukončenia projektu zlepšenia dátovej kvality pre definovaný dataset(y) s vybranými atribútmi. Merania sú pravidelné a súbor parametrov a ukazovateľov dátovej kvality nemá prekračovať maximálnu množinu definovaných parametrov a ukazovateľov z predošlého prislúchajúceho projektu zlepšenia. Každé monitorovacie merania je teda súčasťou prislúchajúcej skupiny monitorovacích meraní k ukončenému projektu zlepšenia dátovej kvality.

© 2019 Firm name. All rights reserved.

Lokalizácia merania dátovej kvality v hlavných oblastiach programu dátovej kvality

Pre lokalizáciu samotného merania dátovej kvality, ktoré je chápané ako najjednoduchšia forma merania (dáta sa vložia do nástroja a ten vygeneruje výsledok merania) existujú 4 miesta, kde ho je možné vykonať. V programe dátovej kvality ide o 3 oblasti:

- **Informatívne merania dátovej kvality (1- krát meranie):** Toto meranie svojou podstatou je veľmi flexibilne. Táto flexibilita je zabezpečená cez variabilné možnosti definovania objektu merania a parametrov s jednotlivými ukazovateľmi dátovej kvality. Meranie dátovej kvality sa vykonáva jeden krát za celé informatívne meranie.

- **Projekty zlepšenia dátovej kvality (2- krát meranie):** V projekte zlepšenia dátovej kvality je meranie dátovej kvality lokalizované na dvoch miestach. Tým prvým je meranie pre komplexnejšie zistenie aktuálneho stavu dátovej kvality (Zhodnotenie dátovej kvality) pre definovaný objekt merania. Druhým miestom je meranie stavu dátovej kvality po implementácií zlepšenia (Vykonávanie kontrol).

- **Monitorovanie zmeny stavu dátovej kvality (1- krát meranie):** Na rozdiel od informatívneho merania dátovej kvality, v monitorovacom meraní neexistuje taká príznačná flexibilita. Objekt merania je prebratý z relevantného projektu zlepšenia dátovej kvality a súbor parametrov a ukazovateľov musí byť z množiny prebratej z prislúchajúceho projektu zlepšenia. Meranie dátovej kvality sa vykonáva jeden krát za jedno monitorovacie meranie.

2.3.1.2 Projekt zlepšenia dátovej kvality pre dataset / datasety - 10-krokový postup zlepšenia dátovej kvality

Tento proces reprezentuje prístup pre zhodnotenie, zlepšovanie a vytvorenie dátovej kvality. Jednotlivé kroky sú ilustrované v obrázku (Obrázok 2) a popísané nižšie. Týchto 10 krokov popisuje konkrétne inštrukcie ako plánovať a implementovať projekty týkajúce sa zlepšenia dátovej kvality. Kroky 1 až 4 obsahujú základné techniky pre definovanie požiadaviek a zhodnotenie súčasného stavu dát, ako aj komunikáciu dôležitých aktivít zúčastneným stranám. Krok 5 je vstupným krokom, kde začínajú aktivity - riešenie príčiny problémov súčasného stavu a opravu chýb, ktoré vedú k zlepšeniu samotných dát a procesov. Slúži pre:

— Projekty zamerané na dátovú kvalitu – zhodnotenie dátovej kvality, identifikácia príčiny problémov a implementácia zlepšení
— Integráciu do iných projektov a metodík – napr. ERP migrácia, dátový sklad, atď.
— Použitie pri každodennej práci

© 2019 Firm name. All rights reserved.

Proces zlepšovania dátovej kvality má opakujúci sa charakter – projektové tímy sa môžu vrátiť k predchádzajúcemu kroku a tým obohatiť ich prácu, môžu si vybrať tie kroky, ktoré spĺňajú na začiatku definované požiadavky a môžu dokonca opakovať celý 10 krokový proces v zmysle podporiť neustále zlepšovanie dátovej kvality. Komunikácia aktivít a výsledkov je krok, ktorý beží súčasne pod všetkými fázami a je kritický pre trvalú podporu a úspech akéhokoľvek projektu.

Popis 10-krokového procesu

1. Definovanie potrieb a prístupu – definovanie a dohodnutie sa na hlavných problémoch, príležitostiach a cieľoch ako riadiť a vykonávať prácu počas celého projektu. Odkazovanie sa na tento krok aj počas ďalších krokov s cieľom udržať cieľ pri všetkých aktivitách.

2. Analýza informačného prostredia – zhromažďovanie, vytváranie a analýza informácií o súčasnom stave informačného prostredia. Dokumentovanie a overovanie životného cyklu informácií, ktoré poskytujú základ pre budúce kroky. Zabezpečuje, že relevantné prostredie je zhodnotené a pomáha objavovať príčiny problémov. Navrhuje sa plán zberu a vyhodnotenia dát.

3. Zhodnotenie dátovej kvality – vyhodnotenie dátovej kvality. Výsledky hodnotenia poskytujú základ pre budúce kroky ako napríklad identifikácia príčiny problémov, potrebné vylepšenia a oprava dát.

4. Zhodnotenie dopadu – použitím rôznych techník sa zhodnotí vplyv nízkej dátovej kvality. Tento krok poskytuje dôvod a argumenty pre zlepšenie dátovej kvality, získanie podpory pre dátovú kvalitu a určenie vhodných investícií do informačných zdrojov.

5. Identifikácia príčiny problémov – identifikácia a prioritizácia skutočných príčin problémov dátovej kvality a vypracovanie odporúčaní na ich riešenie.

6. Vypracovanie plánov na zlepšenie – finalizácia konkrétnych odporúčaní zlepšenia. Vypracovanie a realizovanie plánov zlepšenia na základe odporúčaní.

7. Prevencia budúcich chýb – implementácia riešení, ktoré riešia príčiny problémov dátovej kvality.

8. Oprava súčasných chýb – implementácia krokov na vykonanie vhodných opráv dát.

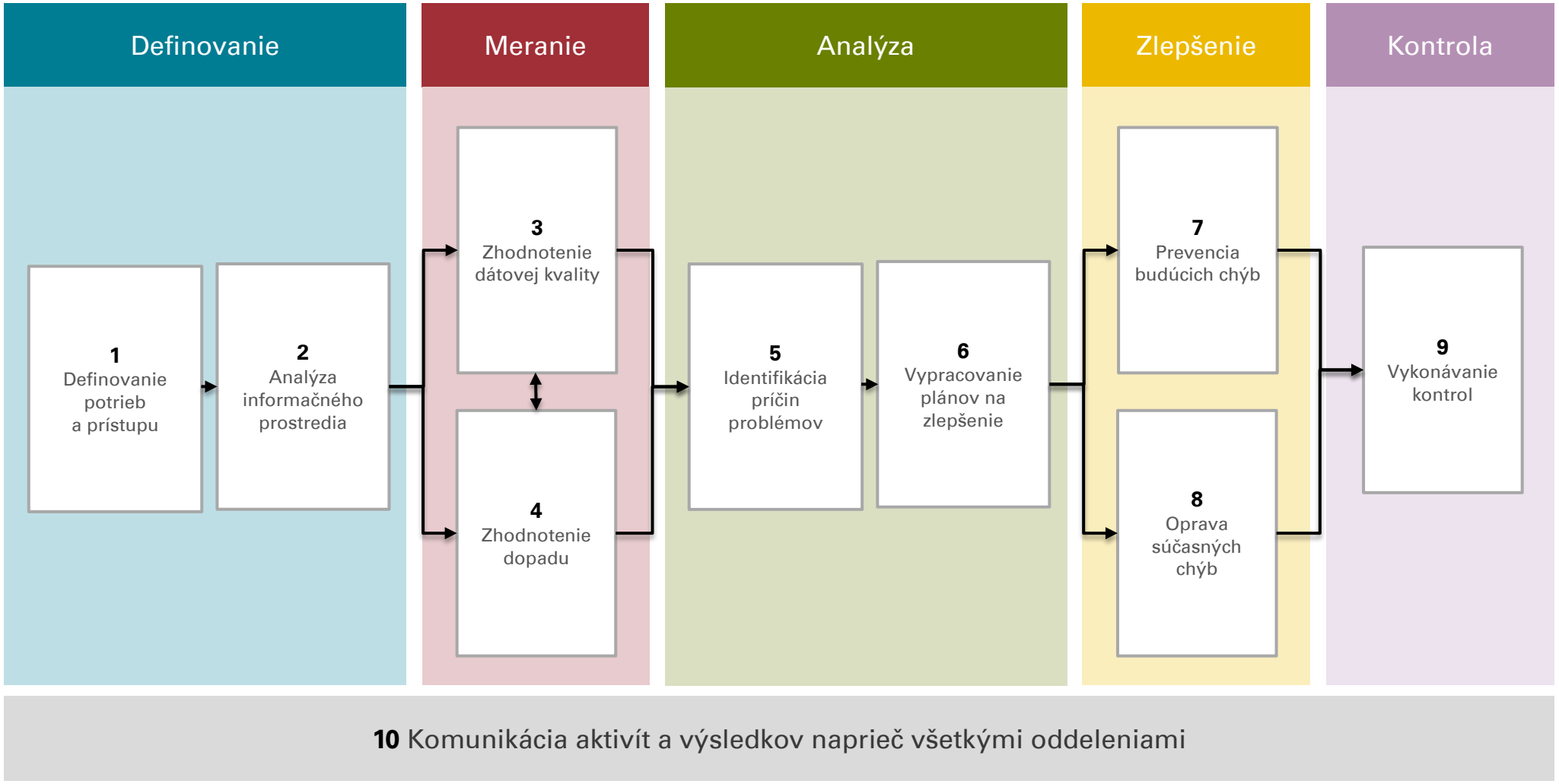
9. Vykonávanie kontrol – monitorovanie a overovanie zlepšení, ktoré boli vykonané. Udržanie zlepšených výsledkov pomocou štandardizácie, dokumentácie a nepretržitého monitoringu.

10. Komunikácia aktivít a výsledkov – dokumentácia a komunikácia výsledkov kvality testov a vykonaných zlepšení. Komunikácia je tak veľmi dôležitá, že je súčasťou každého kroku.

© 2019 Firm name. All rights reserved.

Po najbližšom obrázku nasleduje časť, v ktorej sa detailnejšie rozoberajú úrovne jednotlivých krokov 10-krokového procesu. Pomocou parametrov dátovej kvality sa meria súčasný stav dátovej kvality vo verejnej správe. Vďaka nim sa meria a riadi dátová kvalita, ktorej podstatou je neustále zlepšovanie sa.

© 2019 Firm name. All rights reserved.



Obrázok 2: 10-krokový proces pre zhodnotenie, zlepšenie a vytvorenie dátovej kvality

Krok 1 – Definovanie potrieb a prístupu

Určuje základ pre budúce aktivity dátovej kvality. Ak je tento krok spravený dobre, tak je odrazovým mostíkom pre úspešný projekt prinášajúci hodnotu verejnej správe. Sem patrí:

1) Prioritizácia problémov a príležitostí – výber správnej techniky na prioritizáciu problémov a príležitostí, ktoré sa budú riešiť

2) Plán projektu – efektívne plánovanie je základ pre úspešne zlepšenie dátovej kvality

Krok 2 – Analýza informačného prostredia

Dáva možnosť lepšie pochopiť a spojiť existujúce znalosti o procesoch, ľuďoch/organizácií a technológiách. Práve vďaka nim je možné vykonávať lepšie rozhodnutia o dátovej kvalite. Sem patria:

1) Požiadavky –
    a. Interné na – procesy, bezpečnosť, technológie
    b. Externé na – súkromie, zákony, regulácie
2) Dáta a špecifikácie – dátové štandardy, dátové modely, pravidlá, metadáta, referenčné dáta
3) Technológie –
    a. Pokrokové – databáza, aplikácie, programy
    b. Zastaralé - papier
4) Procesy – sústredenie sa na procesy naprieč životný cyklus informácie, ktoré ovplyvňujú dátovú kvalitu
5) Ľudia a organizácia (ministerstvá) – role a ich zodpovednosti
6) Definovanie životného cyklu informácie – cieľom je spojiť znalosti o procesoch, ľuďoch/organizáciách a technológiách naprieč všetkými fázami životného cyklu informácie: plánovanie, získavanie, ukladanie a zdieľanie, údržba, používanie a odstránenie.
7) Návrh plánu merania dátovej kvality -
    a. získanie dát:
        i. súbor (csv, xlsx, xml, json)
        ii. priamy prístup do databázy

© 2019 Firm name. All rights reserved.

Krok 3 – Meranie dátovej kvality

Najväčším prínosom merania dátovej kvality sú konkrétne dôkazy o problémoch a príležitostiach, ktoré boli identifikované. Výsledky merania poskytujú informácie potrebné na skúmanie príčin problémov, prevenciu budúcich chýb a opravu súčasných chýb. V tomto kroku sú predstavené **parametre dátovej kvality**, ktoré sú detailne popísané v kapitole (3.2) **:**

Prístup, akým sa vykonáva meranie dátovej kvality je nasledovný:

1) Výber parametrov pre dátovú kvalitu

2) Vykonanie merania dátovej kvality pre vybrané parametre

3) Spojenie všetkých výsledkov merania dátovej kvality a ich analýza

Krok 4 – Zhodnotenie dopadu

Definovanie dôvodov „prečo“ zlepšiť dátovú kvalitu. Použitím kvalitatívnych a kvantitatívnych techník ako napr. metóda „5 Whys“, stanovenie priorít, analýza nákladov a prínosov.

Krok 5 – Identifikácia príčiny problémov

Zabezpečuje smer, v ktorom odporúčania a budúce plány na zlepšenia sú sústredené na skutočné príčiny problémov dátovej kvality.

Krok 6 – Vypracovanie plánov na zlepšenie

Zakomponuje výsledky merania dátovej kvality, zhodnotenie dopadu a odporúčania do akčného plánu.

Krok 7 – Prevencia budúcich chýb

Zavedenie procesov, pomocou ktorých sa skvalitňuje čistota dát. Je to dlhodobý opakujúci sa proces.

Krok 8 – Oprava súčasných chýb

Spätná oprava chýb, ktoré spôsobujú problémy

Krok 9 – Vykonávanie kontroly

Určuje, či zlepšovanie činností dosiahlo požadovaný výsledok. Udržuje zlepšenie pomocou štandardizácie, dokumentácie a priebežného monitorovania.

© 2019 Firm name. All rights reserved.

Krok 10 – Komunikácia aktivít a výsledkov naprieč všetkými oddeleniami

Komunikuje výsledky a zlepšenia od začiatku až do konca projektu. Vzdeláva a zvyšuje povedomie o nutnosti dátovej kvality, zabezpečuje viditeľnosť a ukazuje reálne výsledky všetkým ovplyvneným v projekte.

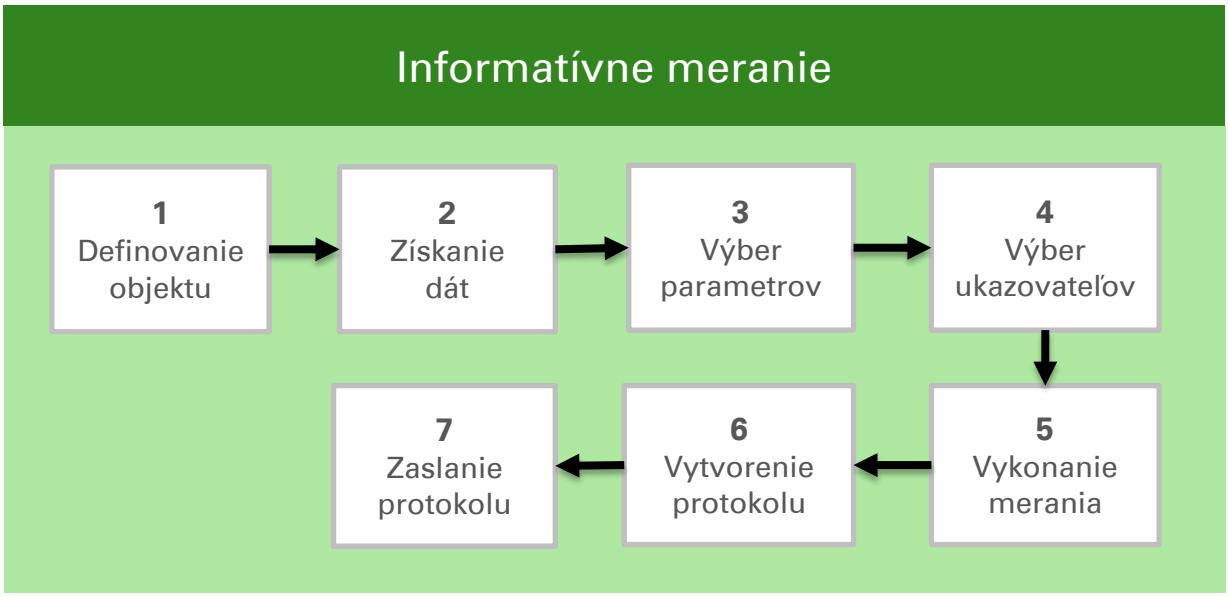
2.3.2 Komplexný popis procesov pre potreby merania dátovej kvality

2.3.2.1 Popis procesu pre potreby informatívneho merania dátovej kvality

Informatívne meranie reprezentuje jednu zo štyroch lokalizácií vykonávania merania dátovej kvality. Pomocou tohto merania sa dá zistiť aktuálny stav dátovej kvality pre ľubovoľne zvolený objekt merania (dataset(y) s vybranými atribútmi). Rovnako flexibilne je možne vybrať aj parametre a ukazovatele dátovej kvality. V nasledujúcej časti je predstavený 7-krokový proces pre potreby informatívneho merania dátovej kvality.

7-krokový proces pre informatívne meranie dátovej kvality

Tento proces reprezentuje najviac flexibilný prístup pre meranie dátovej kvality. Jednotlivé kroky sú ilustrované v nasledujúcom obrázku (Obrázok 3) a popísané nižšie.



Obrázok 3: 7-krokový proces pre informatívne meranie dátovej kvality

© 2019 Firm name. All rights reserved.

Popis 7-krokového procesu pre informatívne meranie dátovej kvality

Krok 1 – **Definovanie objektu**

V prvom kroku je potrebné definovať, aký objekt chceme merať. Zároveň sa definuje konkrétny zdroj objektu. Týmto objektom je myslený dataset s presne určenými atribútmi pre potreby merania. Objekt merania doplňujú dodatočné informácie, ktoré sú potrebné pre účely merania. Medzi tieto informácie patria:

— *Dátové štandardy*: sú pravidlá a smernice, ktoré určujú ako pomenovať dáta, ako ich definovať, ako vytvoriť platné hodnoty a ako špecifikovať biznis pravidlá. Príklady štandardov sú nasledovné:
  - Konvencia pomenovania tabuliek a polí: Ak dáta v poli obsahujú mená, tak názov stĺpca by mal obsahovať štandardnú skratku ako napríklad „NM“ spolu s popisnými slovami pre typ mena ako „NM_first“ alebo „NM_last“,
  - Dátová definícia a konvencia pre písanie biznis pravidiel: Môže existovať dokument s normami, ktoré popisujú minimálny súbor informácií, ktoré sa majú definovať pre každé pole – napríklad každé pole musí byť zdokumentované v dátovom slovníku, dokumentácia musí obsahovať názov súboru, popis, príklad dátového obsahu, či je pole povinné alebo nepovinné a nakoniec predvolenú hodnotu (ak existuje),
  - Zostavovanie, dokumentovanie a aktualizovanie zoznamov platných hodnôt: Je dôležité sa dohodnúť na platných hodnotách pre dané pole. Niekedy je platný zoznam hodnôt vyvinutý interne a niekedy môže byť použitý externý štandardný zoznam. V každom prípade by mal existovať proces, ktorý načrtne, ako sa zmeny v zozname vykonávajú a kto je do týchto rozhodnutí zapojený,
  - Výber metódy zápisu a modelovania pre modelovanie dát: výber by mal byť založený na cieli.

— *Dátový model*: je spôsob vizuálneho znázornenia štruktúry dát (ako sú dáta organizované) v rezorte. Je to zároveň tiež špecifikácia akou majú byť dáta reprezentované v databáze. V riadení dátovej kvality je dôležité rozumieť databáze obsahujúcej dáta a programy, ktoré ich zachytávajú, ukladajú, manipulujú, transformujú, mažú a zdieľajú. Termíny entita, entitná trieda a atribúty sú hlavnými konceptami pričom entitou sa chápe osoba, miesto, udalosť, vec s predmetom záujmu k biznisu, entitnou triedou sa chápe typ entity resp súbor tých vecí, ktorých inštancie sú jednoznačne identifikovateľné a atribútom sa rozumie definícia charakteristík, kvality a vlastností entitných tried. Príklad „Osoba“ je entitná trieda, „Jožko Mrkvička“ je entita a „Meno“ a „Priezvisko“ sú atribúty. Existuje niekoľko zápisov dátových modelov, všetky sú ale zložené z krabičiek reprezentujúce entitné triedy a spájajúce čiary reprezentujúce vzťah medzi 2 entitnými triedami.

— *Biznis pravidlá*: sú autoritatívne princípy alebo usmernenia, ktoré popisujú interakcie a stanovujú pravidlá pre akcie a výsledné správanie a integritu dát. Typy biznis pravidiel s príkladmi sú nasledovné:
  - Obmedzenie: zákazník nesmie zadať na svoj účet viac ako 3 spoplatnené objednávky.
  - Usmernenie: preferovaný zákazník by mal mať svoje objednávky prioritne vybavené.

© 2019 Firm name. All rights reserved.

- Výpočet: ročný objem objednávok zákazníka musí byť vypočítaný ako celkový predaj uzavretý počas fiškálneho roka spoločnosti.
- Odvodenie: zákazník musí byť považovaný za preferovaného, ak zadá viac ako 5 objednávok nad 1000 Eur.
- Načasovanie: objednávka musí byť pridelená expedíterovi, ak je expedovaná, ale nie ešte vyfakturovaná a to v rámci 72 hodín.
- Spúšťač: odoslať upozornenie, ktoré sa musí vykonať potom, ako bola objednávka odoslaná.

— *Metadáta*: sú doslova „dáta o dátach“. Metadáta označujú, popisujú alebo charakterizujú iné dáta pričom uľahčujú získavanie, interpretáciu alebo použitie informácií. Príklad“: Všetky plechovky v obchode na polici majú prázdny štítok. Ako sa zistí, čo sa nachádza v plechovkách? Meno produktu, distribútor, počet kalórií, nutričné hodnoty – všetko sú to metadáta, ktoré popisujú jedlo v plechovkách. Metadáta sú dôležité, pretože:

- poskytujú kontext a pomoc pri pochopení významu dát
- uľahčujú objavovanie relevantných informácií
- organizujú elektronické zdroje
- uľahčujú interoperabilitu medzi systémami
- uľahčujú integráciu informácií
- podporujú archiváciu a uchovávanie dát a informácií

— *Referenčné dáta*: sú množiny hodnôt alebo klasifikačné schémy, na ktoré sa vzťahujú systémy, aplikácie, dátové úložiská, procesy a reporty ako aj kmeňové a transakčné dáta. Je to zvyčajne stabilná informácia so známymi súbormi hodnôt, ktoré sa zriedkavo menia. Ako už názov napovedá, referenčné dáta sú navrhnuté tak, aby sa na ne mohla odvolávať ďalšie údaje. Príklad: zoznam platných hodnôt pre rod môže byť M, Ž, kde M = Muž a Ž = Žena.

Krok 2 – Získanie dát

Keď je objekt merania definovaný, nastáva krok, v ktorom sa získavajú samotné dáta. Vopred je nutné špecifikovať, v akom formáte majú byť tieto dáta získané. Existujú dva hlavné spôsoby získania:

1. Pomocou exportu dát v definovanom formáte súboru (csv, xlsx, xml, json)

2. Pomocou priameho prístupu do databázy

Niekedy môže nastať situácia, v ktorej získame väčšiu množinu atribútov, ak je potrebné. Celý dataset je potom potrebné redukovať o tie atribúty, ktoré nie sú špecifikované v objekte merania.

© 2019 Firm name. All rights reserved.

Krok 3 – Výber parametrov

Pri meraní sa odporúča vyberať len tie parametre, ktoré prioritne reflektujú biznis potreby. Výber tých správnych parametrov má úzko súvisieť s požiadavkami, ktoré sú naviazané na problémy a príležitostí. Najlepšia skúsenosť je začať vykonaním profilovania dát, ktoré obsahujú informácie o štruktúre, obsahu a kvalite dát. Aj keď môže byť hlavným cieľom zistiť presnosť, poprípade duplicitu dát, stále je dôležité vykonať profilovanie dát ako prvé. Po vykonaní profilovania dát sa môže pokračovať ďalej vo výbere parametrov, ktoré riešia spomínané problémy a príležitosti.

Parametre dátovej kvality sú vyberané z definovaného zoznamu v kapitole (3.2) , v ktorej sa nachádzajú aj ich detailné popisy.

Je odporúčané vykonať najprv merania pomocou profilovania dát, pretože väčšina ďalších parametrov vychádza práve z toho. Profilovanie dát je vysvetlené nižšie.

Príklad: v prípade merania duplicity dát je nutné rozumieť, ktoré dátové polia indikujú unikátnosť. Ak nie je známy základ unikátnosti alebo obsahu dát, tak vytvorený algoritmus pre počítanie duplicít, založený na chýbajúcich dátach, môže byť nesprávny. Profilovanie dát je veľmi dôležité na lepšie pochopenie dát a ich kontext.

Krok 4 – Výber ukazovateľov

K vybraným parametrom z kroku 3 prislúchajú konkrétne ukazovatele. Výber ukazovateľov je flexibilný a rovnako úzko súvisí s požiadavkami, ktoré sú naviazané na problémy a príležitostí. Jedná sa o kritéria, ktorých výsledné hodnoty budú vyhodnocované po vykonania merania dátovej kvality. Určujú sa dva typy ukazovateľov:

— Kvalitatívne
— Kvantitatívne

Pre výber ukazovateľov je nutná doménová znalosť. Ich detailné definície sú rozpísané v kapitole (3.3), v ktorej sú ukazovatele zároveň pridelené k jednotlivým parametrom dátovej kvality.

Krok 5 – Vykonanie merania

Meranie konkrétneho data-setu prebehne vo zvolenom nástroji. Detailný popis a postup ako merať je v kapitole (4), kde je popísaný aj technický návod na meranie dátovej kvality prostredníctvom zvolenej skupiny parametrov a ukazovateľov. Piaty krok informatívneho merania je rozdelený do štyroch hlavných etáp:

1. **Vloženie dát objektu merania** do vybratého nastroja na meranie dátovej kvality

2. **Vytvorenie skriptov -** jeden skript bude spúšťať vykonanie merania pre ľubovoľnú skupinu parametrov a ukazovateľov

3. **Prvé spustenie skriptu - profilovanie dát**

© 2019 Firm name. All rights reserved.

Ide o použitie analytických techník na objavenie štruktúry, obsahu a kvality dát. Je vykonávané pomocou SQL alebo špecifických nástrojov. Použitia profilovania dát sú nasledovné:

a. vytvorenie alebo overenie dátového modelu
b. skontrolovanie údajov pochádzajúcich z externých dátových zdrojov
c. lepšie mapovanie dát
d. odkrytie špecifických problémov dátovej kvality
e. potvrdzovanie kritérií selekcie
f. určovanie systémových záznamov
g. porovnávanie a analyzovanie zdrojových a cieľových dátových skladov
h. identifikovanie transformačných pravidiel
i. podporovanie monitorovania dátovej kvality

4. **Ďalšie spúšťanie skriptov –** merania podľa ďalších vybratých parametrov a ukazovateľov dátovej kvality

Krok 6 – Vytvorenie protokolu

Vygeneruje sa report z nástroja (protokol z merania), v ktorom sa vykonalo informatívne meranie dátovej kvality. Šablóna reportu bude dodaná neskôr po vykonaní merania 3 registrov.

Krok 7 – Zaslanie protokolu

Vygenerovaný report (protokol z merania) sa odosiela dohľadu, ktorý je v tomto prípade UPVII – dátová kancelária. Ten rozhoduje o budúcich krokoch, možných inicializáciách projektov na zlepšenie dátovej kvality a podobne.

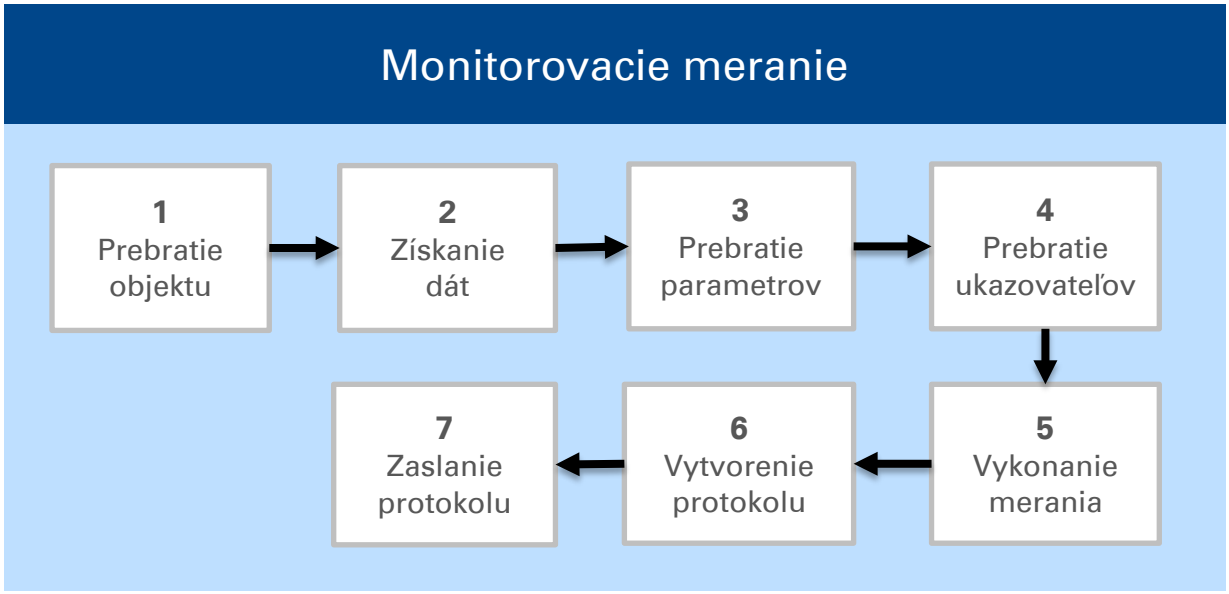
2.3.2.2 Popis procesu pre potreby monitoringu (monitorovacie meranie) zlepšeného objektu merania dátovej kvality

Monitorovacie meranie reprezentuje poslednú zo štyroch lokalizácií vykonávania merania dátovej kvality. Pomocou tohto merania sa dá zistiť aktuálny stav dátovej kvality pre pôvodne zvolený objekt merania (dataset(y) s vybranými atribútmi) v relevantnom projekte zlepšenia dátovej kvality. Rovnako pevne definované sú aj parametre a ukazovatele dátovej kvality, ktoré sú prebraté z pôvodne meraných parametrov a ukazovateľov v predchádzajúceho projektu zlepšenia dátovej kvality. Každé monitorovacie meranie teda vychádza z konkrétneho projektu zlepšenia a je teda prirodzenou a neoddeliteľnou súčasťou aktivít po jeho dokončení. Aktivity monitorovacích meraní sledujú presne stanovený harmonogram, ktorého hlavným príznakom je striktná pravidelnosť. V nasledujúcej časti je predstavený 7 krokový proces pre potreby monitorovacieho merania dátovej kvality.

© 2019 Firm name. All rights reserved.

7-krokový proces pre monitorovacie meranie dátovej kvality

Tento proces reprezentuje najmenej flexibilný prístup pre meranie dátovej kvality a vychádza z kontrolného merania v projekte zlepšenia dátovej kvality. Jednotlivé kroky sú ilustrované v nasledujúcom obrázku (Obrázok 4) a popísané nižšie.



Obrázok 4: 7-krokový proces pre monitorovacie meranie dátovej kvality

Popis 7-krokového procesu pre monitorovacie meranie dátovej kvality

Krok 1 – Prebratie objektu

Na rozdiel od informatívneho merania, v tomto kroku už nie je potrebné definovať, aký objekt chceme merať. Objekt merania sa striktne preberá z relevantného projektu zlepšenia dátovej kvality. Rovnako zdroj objektu je už dopredu definovaný. Objekt merania tiež doplňujú dodatočné informácie, ktoré sú potrebné pre účely merania a tiež sa preberajú z projektu zlepšenia. Medzi tieto informácie patria: dátové štandardy, dátový model, biznis pravidlá, metadáta, referenčné dáta. Ich detailný popis je v kapitole (Popis 7-krokového procesu pre informatívne meranie dátovej kvality) v kroku 1.

Krok 2 – Získanie dát

Získanie dát z prebratého objektu merania by malo byť podstatne jednoduchšie, keďže tento krok je len zopakovanie rovnakých aktivít, ktoré sa vykonali v projekte zlepšenia dátovej kvality. Tento projekt teda slúži ako inšpirácia zopakovania rovnakého alebo veľmi podobného postupu. Špecifikácia, v akom formáte majú byť tieto dáta získané, je rovnako prebratá z relevantného projektu zlepšenia. Existujú dva hlavné spôsoby získania:

© 2019 Firm name. All rights reserved.

1. Pomocou exportu dát v definovanom formáte súboru (csv, xlsx, xml, json)

2. Pomocou priameho prístupu do databázy

Niekedy môže nastať situácia, v ktorej získame väčšiu množinu atribútov. Celý dataset je potom potrebné redokovať o tie atribúty, ktoré nie sú špecifikované v objekte merania.

Krok 3 – Prebratie parametrov

Tak ako je objekt merania prebratý z relevantného projektu zlepšenia dátovej kvality, tak aj súbor parametrov nemá presiahnuť maximálnu množinu, ktorá je prebratá z prislúchajúceho projektu zlepšenia. Inak povedané, v monitorovacom meraní sa nemajú merať iné parametre, ako tie, ktoré boli merané v relevantnom projekte zlepšenia dátovej kvality.

Krok 4 – Prebratie ukazovateľov

Tak ako sú parametre merania prebraté z relevantného projektu zlepšenia dátovej kvality, tak aj súbor ukazovateľov nemá presiahnuť maximálnu množinu, ktorá je prebratá z prislúchajúceho projektu zlepšenia. Inak povedané, v monitorovacom meraní sa nemajú merať iné ukazovatele, ako tie, ktoré boli merané v relevantnom projekte zlepšenia dátovej kvality.

Krok 5 – Vykonanie merania

Meranie konkrétneho data-setu prebehne vo zvolenom nástroji. Detailný popis a postup ako merať je v kapitole (4), kde je popísaný aj technický návod na meranie dátovej kvality prostredníctvom zvolenej skupiny parametrov a ukazovateľov. Piaty krok monitorovacieho merania je v porovnaní s informatívnym meraním rozdelený do troch zjednodušených hlavných etáp:

1. **Vloženie dát z prebratého objektu merania** do vybratého nastroja na meranie dátovej kvality

2. **Prebratie skriptov z relevantného projektu zvýšenia dátovej kvality -** jeden skript bude spúšťať vykonanie merania pre ľubovoľnú skupinu parametrov a ukazovateľov

3. **Spúšťanie skriptov podľa plánu monitorovacieho merania –** merania podľa vybratých parametrov a ukazovateľov dátovej kvality

Krok 6 – Vytvorenie protokolu

Vygeneruje sa report z nástroja (protokol z merania), v ktorom sa vykonalo monitorovacie meranie dátovej kvality. Šablóna reportu bude dodaná neskôr po vykonaní merania 3 vybraných registrov.

© 2019 Firm name. All rights reserved.

Krok 7 – Zaslanie protokolu

Vygenerovaný report (protokol z merania) sa odosiela dohľadu, ktorý je v tomto prípade UPVII – dátová kancelária. Ten rozhoduje o budúcich krokoch, možných analýzach pri zmene stavu kvality a podobne.

2.3.2.3 Popis procesu pre potreby merania dátovej kvality v projekte zlepšenie dátovej kvality pre dataset / datasety - 10-krokový postup zlepšenia dátovej kvality

Projekt zlepšenia dátovej kvality definuje dve lokalizácie merania dátovej kvality. Prvá lokalizácia sa nachádza v kroku 3 v 10-krokovom postupe zlepšenia dátovej kvality a popisuje komplexné meranie súčasného stavu pre objekt merania. Ta druhá je identifikovaná v kroku 9 v 10-krokovom postupe a popisuje kontrolne meranie po implementácii zlepšení pre objekt merania. V nasledujúcej časti je vysvetlené aplikovanie 7-krokového procesu, ktorý vychádza z postupu informatívneho a monitorovacieho merania.

7- krokový proces pre komplexné meranie súčasného stavu v projekte zlepšenia dátovej kvality

Tento proces reprezentuje flexibilnejší prístup pre meranie dátovej kvality, ktorý ale prioritne vychádza z potrieb projektu zlepšenia dátovej kvality. Flexibilita je teda ohraničená rozsahom projektu v závislosti na jeho cieľoch. Jednotlivé kroky sú ilustrované v nasledujúcom obrázku (Obrázok 5) a popísané nižšie.

Obrázok 5: 7-krokový proces pre komplexné meranie súčasného stavu v projekte zlepšenia dátovej kvality

© 2019 Firm name. All rights reserved.

Popis 7-krokového procesu pre komplexné meranie súčasného stavu v projekte zlepšenia dátovej kvality

Krok 1 – **Upresnenie objektu**

Na rozdiel od informatívneho merania je v prvom kroku komplexného merania potrebné vychádzať z krokov 1 a 2 projektu na zlepšenie dátovej kvality. Objekt merania teda upresňujeme a nedefinujeme ho úplne nanovo. Konkrétny zdroj objektu je rovnako upresnený podľa indikátorov z predošlých krokov. Objektom merania je myslený dataset s presne určenými atribútmi pre potreby projektu zlepšenia. Objekt merania doplňujú dodatočné informácie, ktoré sú potrebné pre účely merania aj projektu. Medzi tieto informácie patria: dátové štandardy, dátový model, biznis pravidlá, metadáta, referenčné dáta. Ich detailný popis je v kapitole (Popis 7-krokového procesu pre informatívne meranie dátovej kvality) v kroku 1.

Krok 2 – Získanie dát

Keď je objekt merania upresnený, nastáva krok, v ktorom sa získavajú samotné dáta. Špecifikácia, v akom formáte majú byť tieto dáta získané, rovnako vychádza z predošlých krokov projektu. Existujú dva hlavné spôsoby získania:

1. Pomocou exportu dát v definovanom formáte súboru (csv, xlsx, xml, json)

2. Pomocou priameho prístupu do databázy

Niekedy môže nastať situácia, v ktorej získame väčšiu množinu atribútov. Celý dataset je potom potrebné redukovať o tie atribúty, ktoré nie sú špecifikované v objekte merania.

Krok 3 – Upresnenie parametrov

Pri komplexnom meraní sa odporúča vyberať len tie parametre, ktoré prioritne reflektujú potreby projektu zlepšenia. Výber tých správnych parametrov má úzko súvisieť s požiadavkami, ktoré sú naviazané na problémy a príležitostí. Najlepšia skúsenosť je začať s vykonaním profilovania dát, ktoré obsahujú informácie o štruktúre, obsahu a kvalite dát. Aj keď môže byť hlavným cieľom zistiť presnosť, poprípade duplicitu dát, stále je dôležité vykonať profilovanie dát ako prvé. Po vykonaní prvého profilovania dát sa môže pokračovať ďalej vo výbere parametrov, ktoré riešia spomínané problémy a príležitosti.

Parametre dátovej kvality sú vyberané z definovaného zoznamu v kapitole (3.2), v ktorej sa nachádzajú aj ich detailné popisy.

Je odporúčané vykonať najprv merania pomocou profilovania dát, pretože väčšina ďalších parametrov vychádza práve z neho. Profilovanie dát je vysvetlené nižšie.

© 2019 Firm name. All rights reserved.

Krok 4 – Upresnenie ukazovateľov

K vybraným parametrom z kroku 3 prislúchajú konkrétne ukazovatele. Výber ukazovateľov je úzko naviazaný na potreby projektu zlepšenia a rovnako úzko súvisí s požiadavkami, ktoré sú naviazané na problémy a príležitostí. Ide teda rovnako o upresnenie ukazovateľov. Dôležité je zohľadniť kritéria, ktorých výsledné hodnoty budú vyhodnocované a porovnávané v kroku 9 - kontrola. Určujú sa dva typy ukazovateľov:

— Kvalitatívne
— Kvantitatívne

Pre výber ukazovateľov je nutná doménová znalosť. Ich detailné definície sú rozpísané v kapitole (3.3), v ktorej sú ukazovatele zároveň pridelené k jednotlivých parametrov dátovej kvality.

Krok 5 – Vykonanie merania

Meranie konkrétneho datasetu prebehne vo zvolenom nástroji. Detailný popis a postup ako merať je v kapitole (4), kde je popísaný aj technický návod na meranie dátovej kvality prostredníctvom zvolenej skupiny parametrov a ukazovateľov. Piaty krok komplexného merania je rozdelený do štyroch hlavných etáp:

1. **Vloženie dát objektu merania** do vybratého nastroja na meranie

2. **Vytvorenie skriptov -** jeden skript bude spúšťať vykonanie merania pre ľubovoľnú skupinu parametrov a ukazovateľov

3. **Prvé spustenie skriptu - profilovanie dát**

   Ide o použitie analytických techník na objavenie štruktúry, obsahu a kvality dát. Je vykonávané pomocou SQL alebo špecifických nástrojov. Použitia profilovania dát sú nasledovné:

   a. vytvorenie alebo overenie dátového modelu
   b. skontrolovanie údajov pochádzajúcich z externých dátových zdrojov
   c. lepšie mapovanie dát
   d. odkrytie špecifických problémov dátovej kvality
   e. potvrdzovanie kritérií selekcie
   f. určovanie systém záznamov
   g. porovnávanie a analyzovanie zdrojových a cieľových dátových skladov
   h. identifikovanie transformačných pravidiel
   i. podporovanie monitorovania dátovej kvality

4. **Ďalšie spúšťanie skriptov –** merania podľa ďalších vybratých parametrov a ukazovateľov dátovej kvality

© 2019 Firm name. All rights reserved.

Krok 6 – Vytvorenie protokolu

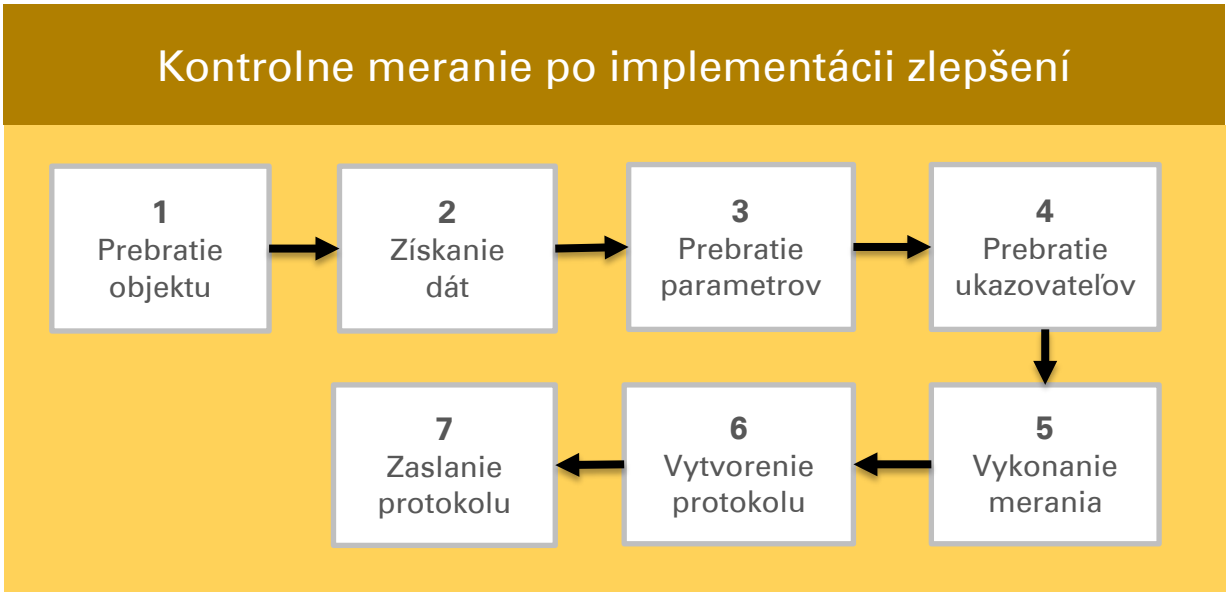
Vygeneruje sa report z nástroja (protokol z merania), v ktorom sa vykonalo komplexné meranie dátovej kvality. Šablóna reportu bude dodaná neskôr po vykonaní merania 3 vybraných registrov.

Krok 7 – Zaslanie protokolu

Vygenerovaný report (protokol z merania) sa odosiela dohľadu, ktorý je v tomto prípade UPVII – dátová kancelária. Rozhodnutie o budúcich krokoch ale zasahuje do agendy samotného projektu zlepšenia dátovej kvality a nie dohľadu dátovej kvality.

7-krokový proces pre kontrolné meranie po implementácii zlepšení v projekte zlepšenia dátovej kvality

Tento proces reprezentuje menej flexibilný prístup pre meranie dátovej kvality, ktorý vychádza z potrieb projektu zlepšenia dátovej kvality a zameriava sa prioritne na potvrdenie a splnenie cieľov projektu. Jednotlivé kroky sú ilustrované v nasledujúcom obrázku (Obrázok 6) a popísané nižšie.



Obrázok 6: 7-krokový proces pre kontrolné meranie po implementácii zlepšení v projekte zlepšenia dátovej kvality

© 2019 Firm name. All rights reserved.

Popis 7-krokového procesu pre kontrolné meranie po implementácii zlepšení v projekte zlepšenia dátovej kvality

Krok 1 – **Prebratie objektu**

Na rozdiel od komplexného merania, v tomto kroku už nie je potrebné definovať ani upresňovať, aký objekt chceme merať. Objekt merania sa preberá z komplexného merania v identickom projekte zlepšenia dátovej kvality. Rovnako zdroj objektu je už dopredu definovaný. Objekt merania tiež doplňujú dodatočné informácie, ktoré sú potrebné pre účely merania a tiež sa preberajú z počiatočných krokov projektu zlepšenia. Medzi tieto informácie patria: dátové štandardy, dátový model, biznis pravidlá, metadáta, referenčné dáta. Ich detailný popis je v kapitole (Popis 7-krokového procesu pre informatívne meranie dátovej kvality) v kroku 1.

Krok 2 – Získanie dát

Získanie dát z prebratého objektu merania by malo byť podstatne jednoduchšie, keďže tento krok je len zopakovanie rovnakých aktivít, ktoré sa vykonali v komplexnom meraní projektu zlepšenia dátovej kvality. Komplexné meranie teda slúži ako inšpirácia zopakovania rovnakého alebo veľmi podobného postupu. Špecifikácia, v akom formáte majú byť tieto dáta získané, je rovnako prebratá z relevantného komplexného merania projektu zlepšenia. Existujú dva hlavné spôsoby získania:

3. Pomocou exportu dát v definovanom formáte súboru (csv, xlsx, xml, json)

4. Pomocou priameho prístupu do databázy

Niekedy môže nastať situácia, v ktorej získame väčšiu množinu atribútov. Celý dataset je potom potrebné redukovať o tie atribúty, ktoré nie sú špecifikované v objekte merania.

Krok 3 – Prebratie parametrov

Tak ako je objekt merania je prebratý komplexného merania toho istého projektu zlepšenia dátovej kvality, tak aj súbor parametrov nesmie presiahnuť maximálnu množinu, ktorá je prebratá z prislúchajúceho komplexného merania. Inak povedané, v kontrolnom meraní nie je možné merať iné parametre, ako tie, ktoré boli merané v relevantnom komplexnom meraní toho istého projektu zlepšenia dátovej kvality.

Krok 4 – Prebratie ukazovateľov

Tak, ako sú parametre merania prebraté z relevantného komplexného merania v projekte zlepšenia dátovej kvality, tak aj súbor ukazovateľov nesmie presiahnuť maximálnu množinu, ktorá je prebratá z prislúchajúceho komplexného merania. Inak povedané, v kontrolnom meraní sa nemajú merať iné ukazovatele, ako tie, ktoré boli merané v relevantnom komplexnom meraní toho istého projektu zlepšenia dátovej kvality.

© 2019 Firm name. All rights reserved.

Krok 5 – Vykonanie merania

Meranie konkrétneho datasetu prebehne vo zvolenom nástroji. Detailný popis a postup ako merať je v kapitole (4), kde je popísaný aj technický návod na meranie dátovej kvality prostredníctvom zvolenej skupiny parametrov a ukazovateľov. Piaty krok kontrolného merania je v porovnaní s komplexným meraním rozdelený do troch zjednodušených hlavných etáp:

4. **Vloženie dát z prebratého objektu merania** do vybratého nastroja na meranie dátovej kvality

5. **Prebratie skriptov z relevantného komplexného merania -** jeden skript bude spúšťať vykonanie merania pre ľubovoľnú skupinu parametrov a ukazovateľov

6. **Spúšťanie skriptov podľa plánu monitorovacieho merania –** merania podľa vybratých parametrov a ukazovateľov dátovej kvality

Krok 6 – Vytvorenie protokolu

Vygeneruje sa report z nástroja (protokol z merania), v ktorom sa vykonalo kontrolné meranie dátovej kvality. Šablóna reportu bude dodaná neskôr po vykonaní merania 3 vybraných registrov.

Krok 7 – Zaslanie protokolu

Vygenerovaný report (protokol z merania) sa odosiela dohľadu, ktorý je v tomto prípade UPVII – dátová kancelária. Rozhodnutie o budúcich krokoch ale zasahuje do agendy samotného projektu zlepšenia dátovej kvality a nie dohľadu dátovej kvality.

2.3.2.4 Rola dátového kurátora v procesoch merania dátovej kvality

Dátový kurátor je kľúčová rola v celom riadení aj meraní dátovej kvality. Význam a vplyv tejto pozície bude v budúcnosti ešte viac rásť. Dátový kurátor má mať nielen zodpovednosti, ale aj príslušné kompetencie vyjadriť sa ku všetkým plánovaným zmenám a projektom v svojom rezorte, ktoré môžu mať vplyv na kvalitu dát a informácií v jeho zodpovednosti.

Meranie dátovej kvality má dátovým kurátorom a ich rezortným nadriadeným pomôcť externe prezentovať a „predať“ pozitívne trendy, ktoré boli vďaka nim spustené a dodané - a to na základe objektívnych metrík a presných overiteľných čísiel. V počiatočnej fáze merania dátovej kvality je kľúčové vyjadrenie skutočného východiskového stavu dátovej kvality – bez zbytočného prikľášovania alebo zametania pod koberec. Pritom tu paradoxne platí princíp, že čím horší je počiatočný AS-IS východiskový stav v rezortnej dátovej kvalite, tým väčší je potom priestor pre postupné zlepšenie a pre preukázateľnosť pridanej hodnoty a dôležitosti tejto novej role - ako aj konkrétnych osôb, ktoré ju vykonávajú.

© 2019 Firm name. All rights reserved.

V procese samotného merania dátovej kvality boli identifikované štyri základné typy zodpovednosti. Nižšie sú popísané zodpovednosti a kompetencie v spojení s rolou dátového kurátora. Tieto zodpovednosti sú definované cez štyri (RACI) úrovne:

— Zodpovedný vykonávateľ – Responsible (R),
— Zodpovedný vlastník – Accountable (A),
— Konzultuje – Consulted (C),
— Je informovaný – Informed (I).

Druh merania dátovej kvality	Názov kroku v meraní dátovej kvality	R	A	C	I
Informatívne meranie	1 Definovanie objektu		x	x	
	2 Získanie dát		x	x	
	3 Výber parametrov		x		
	4 Výber ukazovateľov		x		
	5 Vykonanie merania		x		
	6 Vytvorenie protokolu		x		
	7 Zaslanie protokolu	x	x		
Monitorovacie meranie	1 Prebratie objektu		x	x	
	2 Získanie dát		x	x	
	3 Prebratie parametrov		x		
	4 Prebratie ukazovateľov		x		
	5 Vykonanie merania		x		
	6 Vytvorenie protokolu		x		
	7 Zaslanie protokolu	x	x		
Komplexné meranie	1 Upresnenie objektu		x	x	
	2 Získanie dát		x	x	
	3 Upresnenie parametrov		x		
	4 Upresnenie ukazovateľov		x		
	5 Vykonanie merania		x		
	6 Vytvorenie protokolu		x		
	7 Zaslanie protokolu	x	x		

© 2019 Firm name. All rights reserved.

Druh merania dátovej kvality	Názov kroku v meraní dátovej kvality	R	A	C	I
Kontrolne meranie	1 Prebratie objektu		x	x	
	2 Získanie dát		x	x	
	3 Prebratie parametrov		x		
	4 Prebratie ukazovateľov		x		
	5 Vykonanie merania		x		
	6 Vytvorenie protokolu		x		
	7 Zaslanie protokolu	x	x		

Tabuľka 2: Zodpovednosti dátového kurátora v procesoch merania dátovej kvality

Zodpovednosti dátového kurátora v procese merania a riadenia dátovej kvality

Dátový kurátor:

- **Orchestruje a koordinuje** činnosti a komunikáciu súvisiace s meraním a riadením dátovej kvality vo svojom rezorte
- **Monitoruje procesy a vykonáva dozor** nad dodržiavaním povinností a smerníc, ktoré súvisia s dátovou kvalitou v ich oblasti
- **Komunikuje** so všetkými zúčastnenými relevantnými subjektami na meraní a riadení kvality dát v rámci svojich kompetencií
- **Monitoruje a sleduje** kvalitu dát a jej trendov vo svojom rezorte
- **Prijíma a vyhodnocuje** návrhy a podnety týkajúce sa dátovej kvality a tieto návrhy môže v prípade potreby postúpiť hlavnému dátovému kurátorovi alebo dátovej kancelárií
- **Kontroluje** dodržiavanie zákonov, smerníc a procesov, ktoré súvisia s dátovou kvalitou
- **Odporúča** opatrenia na zabezpečenie požadovanej dátovej kvality
- V rozsahu svojej pôsobnosti **vydáva** záväzné stanoviská, metodicky usmerňuje povinné osoby pri meraní a riadení kvality dát, podieľa sa aktívne na príprave všeobecne záväzných právnych predpisov v oblasti riadenia dátovej kvality
- Podieľa sa na **„evanjelizácií“** významu dátovej kvality aj na príprave s tým súvisiacich prezentácií, dokumentov a školení
- Aktívne sa **vzdeláva** v oblasti dátovej kvality a v spôsoboch jej riadenia a merania
- **Zodpovedá** za riadenie dátovej kvality **referenčných údajov** v rámci svojich kompetencií
- **Manažuje** procesy súvisiace s riadením kvality údajov v rámci svojej inštitúcie
- **Rieši strategický rozvoj** inštitúcie v oblasti využitia, prezentácie a analýzy dát
- **Zodpovedá** za **súlad údajov so štandardmi** dátovej kvality

© 2019 Firm name. All rights reserved.

— Funguje ako **komunikátor a centrálny kontaktný bod** (SPOC) s používateľmi údajov pre dátovú kvalitu a ich zrozumiteľnosť

Právomoci a kompetencie dátového kurátora pre meranie a riadenie dátovej kvality

Dátový kurátor:

— Má **pridelené všetky potrebné kompetencie a právomoci**, ktoré potrebuje k výkonu svojich zodpovedností v oblasti riadenia a merania dátovej kvality.
— Má **právomoc a možnosť vyjadriť sa** ku všetkým projektom a aktivitám, ktoré majú potenciálny dopad na dátovú kvalitu v jeho rezorte.
— Ma **pridelené všetky prístupové práva** k databázam a datasetom v jeho zodpovednosti za dátovú kvalitu v jeho rezorte – v súlade so zákonom, so záväznými smernicami a so schválenými procesmi.
— Má **koordinačné právomoci** pre komunikáciu a činnosti súvisiace s riadením dátovej kvality a jej vyhodnocovania.

© 2019 Firm name. All rights reserved.

3 Špecifikácie parametrov dátovej kvality vo verejnej správe

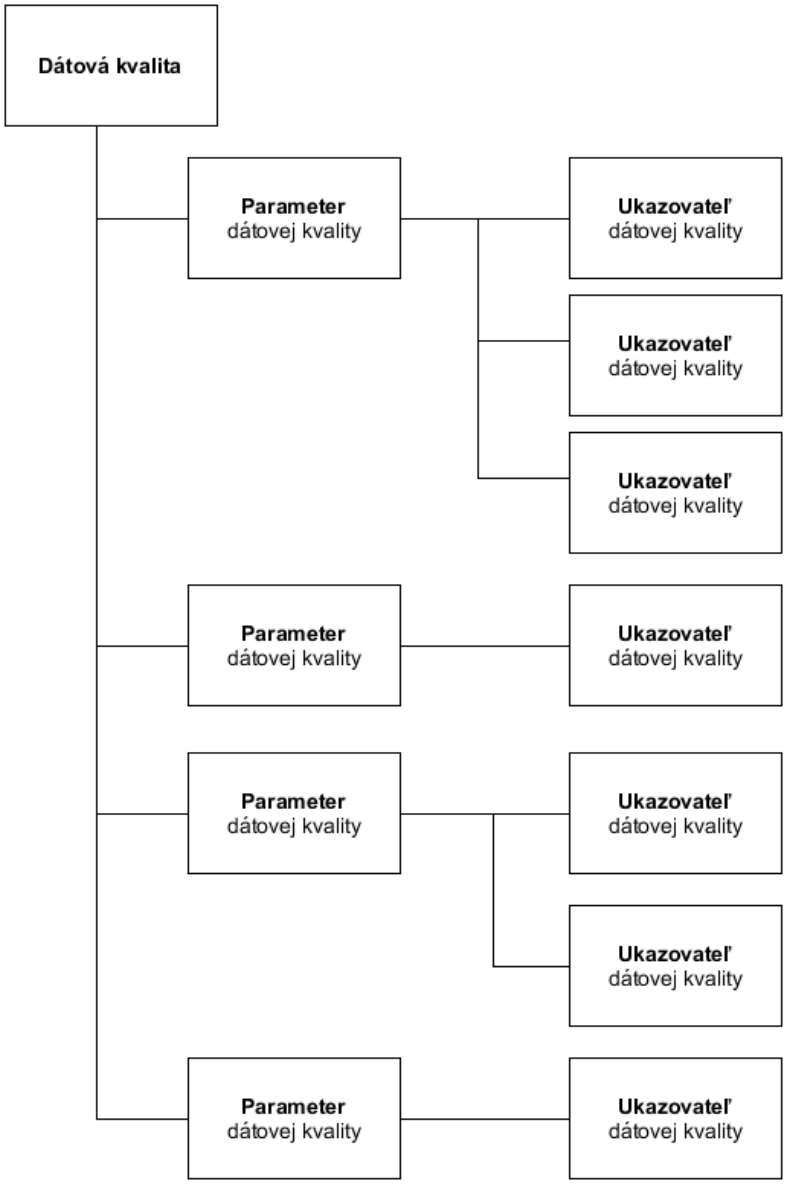
3.1 Úvod k špecifikácie parametrov dátovej kvality vo verejnej správe

V tejto kapitole sa budeme venovať definovaniu parametrov dátovej kvality, v anglickej literatúre často označované aj ako „data quality dimensions“. Vychádzať budeme z ôsmich parametrov definovaných v dokumente „Strategická priorita – Manažment údajov“ schváleného vládou SR, a doplníme niektoré ďalšia parametre, ktoré považujeme za užitočné.

- **Presnosť** (čistota) – miera, s akou objekt evidencie reprezentuje reálny svet, vyjadrená zhodou s referenčnými údajmi.
- **Kompletnosť** – Kompletnosť údajov znamená, že všetky údaje z objektu evidencie, považované alebo označené za povinné, sú prítomné v dátovom prvku.
- **Aktuálnosť** – Údaje sú časovo adekvátne a považované za aktuálne.
- **Unikátnosť** – Vyhodnotenie duplicity údajov vo vzťahu k jednoznačnému referencovateľnému identifikátoru.
- **Referenčná integrita** – Údaje v objekte evidencie sú referencované s referenčnými údajmi v podobe, v akej sú evidované v referenčných registroch. Referenčné údaje z objektu evidencie sú stotožnené so subjektom evidencie.
- **Strojová spracovateľnosť** – možnosti automatického spracovania údajov plynúce z formátu reprezentácie dát ako napr. spájanie rôznych dát v rôznych zdrojov, či spracovanie dát s ohľadom na ich význam.
- **Konzistentnosť** – vzájomné logické vzťahy v rámci objektu evidencie sú správne kompatibilné a v súlade so stanovenými biznis pravidlami.
- **Správnosť** – zhoda údajov s kritériami, ktoré stanovujú formát dát.

Každý tento parameter má viacero „odtieňov“ a dá sa vnímať v rôznych kontextoch. Parameter dátovej kvality predstavuje pohľad na isté spektrum problémov spojených s dátovou kvalitou a preto potrebujeme koncept (Obrázok 7), ktorý by hovoril o dátovej kvalite konkrétnejšie.

© 2019 Firm name. All rights reserved.



Obrázok 7: Špecifikácia dátovej kvality

Pre jednotlivé parametre dátovej kvality definujeme ukazovatele dátovej kvality. Zatiaľ čo parameter je komplexný, ukazovateľ je konkrétny pojem. Pre ukazovateľ existuje konkrétny vzorec alebo postup, ako ho môžeme vypočítať. Definovaniu výpočtov je venovaná kapitola (4). Dá sa povedať, že parameter dátovej kvality predstavuje skupinu viacerých ukazovateľov.

© 2019 Firm name. All rights reserved.

3.2 Definovanie parametrov dátovej kvality

3.2.1 Presnosť

Presnosť hovorí ako dobre dáta zodpovedajú skutočnostiam v reálnom svete. Prv než zadefinujeme niektoré ukazovatele, vysvetlime si tento pojem na príklade. Zoberme napríklad register fyzických osôb a v ňom záznam o osobe Dagmara Závodská. Rozoberieme rôzne stupne presnosti, podľa toho, ako vyzerá záznam o tejto osobe.

— V záznamoch máme meno Dagmara Závodská.
  To znamená, že meno máme zaznamenané úplne presne alebo presnosť je 100%.
— V záznamoch máme meno Dgmara Závodská.
  To znamená, že meno nie je zaznamenané úplne presne (chýba jedno „a“ v mene). Takúto chybu vieme pomerne ľahko identifikovať a aj automatizovane opraviť. Nemusí to byť to tak byť vždy, napríklad pri mene Sylvia, nie je jasné, či má ísť o meno Silvia alebo Sylvia. Aj pri mene Dagmara si vieme predstaviť viaceré prípustné hodnoty, napr. Dagmara, Dagmar.
— V záznamoch máme meno Zuzana Závodská (aj keď ide o Dagmar).
  Závažnejšia chyba, ktorá výrazne komplikuje identifikáciu osoby. Navyše nie je možné jednoducho takúto chybu odhaliť.

Z teoretického pohľadu presnosť delíme na dve základné kategórie:

1. **Syntaktická presnosť**. Syntaktická presnosť hovorí, či hodnota údaju je medzi prípustnými údajmi, prípadne ako veľmi je od nich vzdialená. V uvedených príkladoch meno Dgmara nie je v prípustných menách a jeho vzdialenosť od prípustnej hodnoty sa dá merať napríklad v počte písmen, ktoré je potrebné upraviť (1 písmeno).

   V prípade mien táto úloha nie je jednoduchá. Napríklad rozhodnutie, či meno má byť správne Dagmar alebo Dagmara, nie je zrejmé. Napriek tomu aj tu vieme opraviť viacero zrejmých prepisov.

   Ak by sme namiesto mien hovorili o názvoch miest, tu je jasné, čo sú a čo nie sú prípustné hodnoty. A má zmysel hovoriť o tom, ako ďaleko je uvádzaná hodnota od reálnej hodnoty.

   Syntaktická presnosť sa meria „vzdialenostnými“ funkciami, napr. počet písmen ktoré je potrebné upraviť. Táto vzdialenosť nám hovorí nie len o kvalite záznamu, ale môže dávať aj návod na jeho opravu. Napríklad môžeme prijať pravidlo, že ak stačí opraviť jedno písmeno, tak záznam automatizovane upravíme. Aj tu treba dať pozor, napr. záznam Ea má vzdialenosť 1 aj k menu Eva aj k menu Ema. Viac o možnosti zvyšovania kvality dát je v kapitole (2.3.1.2).

   Syntaktická presnosť má blízko k parametrom konzistentnosť a správnosť.

© 2019 Firm name. All rights reserved.

2. **Sémantická presnosť**. Syntaktická presnosť sa zameriava na porovnanie zadanej hodnoty s možnými prípustnými hodnotami bez ohľadu na skutočný objekt v reálnom svete. Sémantická presnosť hodnotí, či zaznamenaná hodnota zodpovedá skutočnej hodnote.

   To znamená, že z pohľadu syntaktickej správnosti je záznam Zuzana Závodská úplne v poriadku a syntaktická presnosť tohto záznamu je 100%. Sémantická presnosť tohto záznamu je 0, pretože nezodpovedá realite. Ináč povedané, meno Zuzana je syntakticky prípustné, ale sémanticky v tomto prípade nie, lebo osoba sa volá Dagmara.

   Zatiaľ čo syntaktickú presnosť má zmysel vyhodnocovať nejakou „vzdialenostnou“ funkciou (napr. počet písmen ktoré treba opraviť), sémantickú správnosť zväčša hodnotíme iba dvoma stavmi Správne/Nesprávne.

   Vyhodnocovať sémantickú správnosť je náročné, lebo k jej vyhodnoteniu potrebujeme vedieť skutočnú pravdu. Jednoduchý prípad je, ak zistíme, že záznam nemá 100% syntaktickú správnosť. Vtedy je zrejmé, že nemôže byť ani sémanticky správny. Ak je však syntaktická správnosť 100%, pre rozhodnutie, či je záznam aj sémanticky správny treba overenie voči nejakému zdroju pravdy.

Vyhodnocovanie ukazovateľov pre parameter Presnosť vyžaduje doménovú znalosť. Oplatí sa však zamerať na:

— Definovanie vzdialenostnej funkcie pre hodnotenie syntaktickej presnosti údajov.
— Definovanie kategórií pre hodnoty syntaktickej presnosti:
  - syntakticky presné
  - s malou syntaktickou chybou, ktorú je možné automatizovane opraviť
  - s väčšou syntaktickou chybou
— Definovanie postupu pre hodnotenie sémantickej presnosti:
  - správne – ak vieme, že záznam je presný
  - nesprávne – ak vieme, že záznam je nepresný
  - neznáme – ak nevieme, či je záznam správny alebo nie

3.2.2 Správnosť

Správnosť hovorí, či formát údajov zodpovedá definovaným pravidlám. Typickým príkladom je formát dátumu, času, formát čísel. Formát sa niekedy definuje aj pre komplikovanejšie kompozitné objekty, napríklad adresa, osoba.

Delíme ich na dve skupiny:

— Spoločné kritériá definuje výnos MFSR 55/2014 o štandardoch pre ISVS. Príkladom je používanie centrálnych číselníkov (napr. číselník obcí), použitie definovaných dátových prvkov v správnom formáte a podobne.

© 2019 Firm name. All rights reserved.

— Kritériá špecifické pre daný systém, napríklad počet desatinných miest, formát dátumu, maximálna dĺžka reťazca a podobne.

Správnosť údajov a dodržiavanie predpísaných formátov úzko súvisí s parametrami strojová spracovateľnosť a zrozumiteľnosť.

Tento parameter môžeme vyhodnocovať kvalitatívne, ale často je dôležitejšie kvantitatívne hodnotenie. Ak je proces spracovania dát nastavený tak, že dodržiava daný formát, je nepravdepodobné, že by sme našli hodnoty, ktoré danému formátu nevyhovujú. Preto ide skôr o kvalitatívne posúdenie, či dátový model zodpovedá definovaným štandardom a či sa dodržiavajú predpísané formáty.

3.2.3 Kompletnosť

Kompletnosť hovorí o tom, či dáta obsahujú dostatok údajov. V prvom rade je dôležité, či sú vyplnené všetky povinné údaje. O kvalite dát však hovorí aj kompletnosť nepovinných dát. Ak napríklad firma zbiera údaje o svojich zákazníkoch, bude kontakt na zákazníka pravdepodobne povinným atribútom, bez ktorého je záznam takmer bezcenný. Avšak aj atribút záujmy zákazníka, história nákupov a podobne majú veľkú pridanú hodnotu a vyplnenosť týchto údajov hovorí tiež o kvalite databázy.

Pri hodnotení kompletnosti využijeme kvalitatívny ukazovateľ, ktorého hodnoty sú:

1. Údaj nie je vyplnený a vieme, že nemá byť vyplnený. Z pohľadu dátovej kvality ide o kvalitný údaj. Príkladom môže byť, že vieme, že zákazník nevlastní motorové vozidlo.
2. Údaj nie je vyplnený a vieme, že má byť. Ide o nekvalitný údaj.
3. Údaj nie je vyplnený a nevieme, či má alebo nemá byť vyplnený. Aj toto je nekvalitný údaj.
4. Údaj je vyplnený. V takomto prípade sme z pohľadu kompletnosti spokojný, treba použiť iné parametre – správnosť, konzistentnosť a podobne – na hodnotenie jeho ďalšej kvality.

V súvislosti s kompletnosťou môžeme vyhodnocovať aj kvalitatívny ukazovateľ. Ten hovorí o kvalite návrhu dátovej schémy a hodnotí, či vieme rozlíšiť prvý, druhý a tretí stav, to znamená či vieme pri nevyplnenom údaji povedať, či je nevyplnený úmyselne alebo nie. Zväčša to znamená, že pri návrhu dátovej schémy rozlišujeme medzi prázdnou hodnotou a hodnotou null. Hodnota null v takomto prípade znamená, že vieme, že záznam má byť nevyplnený a môžeme ho považovať z pohľadu kompletnosti za správny.

Kompletnosť má zmysel vyhodnocovať na jednotlivých atribútoch, napr. kompletnosť dátumu narodenia. A má zmysel ju vyhodnocovať aj na záznamoch, to znamená, že vieme identifikovať, či je záznam:

© 2019 Firm name. All rights reserved.

— Úplne vyplnený
— Má vyplnené všetky povinné atribúty, ale nemá vyplnené všetky nepovinné atribúty (null považujeme za vyplnený atribút)
— Percento vyplnených atribútov

To nám umožní zaradiť záznamy do viacerých kategórií podľa kvality ich kompletnosti a môžeme použiť ukazovateľ, ktorý hovorí o počte vyplnených, čiastočne vyplnených a málo vyplnených záznamov. Takýto ukazovateľ dáva dobrý prehľad o použiteľnosti jednotlivých záznamov a môže slúžiť pre identifikovanie problémov a definovanie oblastí kvalitných dát.

3.2.4 Unikátnosť

Unikátnosť hovorí o duplicitách v dátach. Zvlášť bude diskutovaná pri referencovateľných identifikátoroch, kde sa požaduje aby každý identifikátor bol jedinečný. Požiadavka na unikátnosť má zmysel aj pre iné atribúty, napr. v registri daňových subjektov nemajú dva rôzne subjekty odkazovať na tú istú právnickú či fyzickú osobu. Alebo v obchodnom registri by mal byť názov obchodnej spoločnosti jedinečný, napriek tomu, že nie je referenčným identifikátorom (lebo sa môže v čase meniť).

Ide o kvantitatívny/dátový ukazovateľ. Vyhodnocujeme ho podobne ako jedinečnosť referenčného identifikátora, ako podiel počtu objektov s opakujúcim sa identifikátorom, ku počtu všetkých objektov.

Unikátnosť môžeme chápať aj o niečo všeobecnejšie, napríklad, že ak dve obchodné spoločnosti majú nie úplne zhodný názov, ale názov, ktorá je príliš podobný, aj takáto podobnosť je nežiadúca. Môžeme definovať funkciu, ktorá meria podobnosť a povedať, že ak je podobnosť menšia ako definovaná hranica, tak tieto dva záznamy nepovažujeme za unikátne. Touto definíciou unikátnosť približujeme parametru konzistentnosť a správnosť.

3.2.5 Aktuálnosť

Aktuálnosť hovorí, či sú dáta aktuálne, alebo hrozí, že sú tam aj staršie údaje. Problémy s aktuálnosťou údajov môže byť komplikované strojovo identifikovať. Často sú spôsobované komplikovanými procesmi, ktorými sú dáta zapisované do ISVS. Preto ide skôr o kvalitatívny parameter, posudzujúcu procesy súvisiace s vytváraním dát. Typické hodnoty tohto kritéria sú:

— Dáta sú aktualizované v systéme okamžite po prijatí rozhodnutia, v ideálnom prípade sa rozhodnutie robí priamo v registri alebo sa do neho automatizovane prenáša.
— Dáta sú aktualizované dávkovým automatizovaným spôsobom v pravidelných časových intervaloch. Napríklad prenos údajov z distribuovaných systémov do centra každú noc, raz týždenne a podobne.

© 2019 Firm name. All rights reserved.

— Údaje sú pravidelne manuálne prepisované do systému. Ide o systémový problém, chýbajúca automatizácia nezaručuje aktuálnosť dát.
— Údaje nie sú aktualizované. Údaje musia byť aktuálne, časovo príslušné. Nemalo by sa stať, že OVM vydá rozhodnutie, ale do ISVS príslušné údaje zapisuje iba raz za týždeň a teda údaje v registri niekoľko dní nezodpovedajú skutočnosti. Ďalšie zdržanie môže byť napríklad medzi zaevidovaním údaju v zdrojovom registri a jeho následnom spracovaní v referenčnom registri. Pokiaľ údaj v ISVS nie je, je ťažké zistiť, že nie je aktuálny, nemusí byť jasné, s čím sa má porovnať.

Ak by sme chceli strojovo vyhodnocovať aktuálnosť údajov, môžeme si pomôcť pojmov volatilita (miera premenlivosti), ktorý hovorí ako často sa údaje menia. Ak napr. hovoríme o dátume narodenia, tak vieme, že sa nemení, jeho volatilita je nízka a teda aj menej časté aktualizácie údajov nepredstavujú veľký problém. Podobne je to napríklad pri priezvisku osoby, ktoré sa síce môže zmeniť, ale je to iba výnimočne. Častejšie sa mení bydlisko, ale tiež nie veľmi rýchlo. Typickým príkladom atribútu s veľkou volatilitou sú ceny, kde je nutné aby boli aktualizované online.

Ďalším rozmerom, ktorý treba brať do úvahy je, že potrebujeme, aby údaje boli nie len aktuálne v priamočiarom slova zmysle, to znamená, že vieme aké hodnoty sú teraz. Často potrebujeme vedieť, aké hodnoty budú platné zajtra a pokiaľ je táto informácia známa, tak by mala byť aj dostupná.

3.2.6 Strojová spracovateľnosť

Strojová spracovateľnosť meria ako jednoducho je možné údaje automatizovane spracovať.
Pre jej hodnotenie budeme používať ukazovateľ (viď Tabuľka 3) 5★ Open Data stupnica. Hovorí o štruktúrovanosti / formáte v ktorom sú dáta poskytované, od strojovo nespracovateľného (napr. pdf), cez proprietárne formáty (napr. excel), cez otvorené formáty (napr. xml, json, csv) až po rdf a lod formát. Ide kvalitatívne/architektonické kritérium. Kvantitatívne sa môže nepriamo vyjadriť cez skoring model.

Počet hviezdičiek	Popis
★	Publikovanie otvorených dát tak aby boli dostupné konzumentom. Formát dát nie je vhodný na strojové spracovanie, napr. pdf, scan dokumentu a podobne.
★★	Dáta sú publikované v štruktúrovanej podobe, ale je potrebné použiť platené nástroje. Typickým príkladom je excel tabuľka, keď konzument musí mať komerčný nástroj na čítanie dát.
★★★	Podobne ako dvojhviezdičková úroveň, ale dáta sú publikované v otvorenom formáte, napr. csv namiesto excel.

© 2019 Firm name. All rights reserved.

Počet hviezdičiek	Popis
★★★★	Dáta obsahujú URI (identifkátori), vďaka čomu sa na ne dá odkazovať, dajú sa uložiť odkazy a podobne. Na tejto úrovni sa očakáva existencia ontológie ktorá definuje štruktúru dát.
★★★★★	Dáta sú prepojené s inými dátovými zdrojmi, čo zväčšuje potenciál využitia. Vhodné je využívať štandardizované ontológie.

Tabuľka 3: Hodnotenie strojovej spracovateľnosti

Ďalšie ukazovatele súvisiace so strojovou spracovateľnosťou hovoria, čo všetko ešte musíme s dátami urobiť, aby sme ich mohli požiť pre konkrétny účel.

3.2.7 Referenčná integrita

Referenčná integrita sa zameriava na možnosť použitia údajov naprieč verejnou správou. Nehovorí o presnosti, konzistentnosti a podobne, ale o previazanosti a použiteľnosti údajov v rámci verejnej správy a aj mimo nej ďalšími konzumentmi údajov.

V tejto súvislosti poznamenajme, že keď hodnotíme kvalitu údajov (predchádzajúce parametre), tak sa sústredíme na dáta, ktoré sú doménou daného registra, nie na referencované údaje. Napríklad, ak hovoríme o registri vozidiel, tak nás zaujíma kvalita údajov ako typ vozidla, rok výroby. Aj pri vlastníkovi vozidla, ktorý by mal byť referencovaný na RFO/RPO nás zaujíma jeho kvalita. Ale nezaujímame sa o kvalitu mena, priezviska, adresy, ... vlastníka, pretože tieto údaje by sa mali preberať z referenčného registra a ich kvalita je jeho zodpovednosť. Podmienkou je preberanie týchto údajov z referenčného registra. Schopnosť poskytovať údaje a mieru preberania údajov hodnotíme samostatným parametrom referenčná integrita.

V tejto časti sa zameriame na registre vo verejnej správe. Zmyslom existencie registra je evidovať informácie o objektoch, napr. Register fyzických osôb registruje fyzické osoby, Register vozidiel registruje vozidlá, Obchodný register registruje obchodné spoločnosti. Referenčná integrita vyžaduje, aby mal každý takýto objekt priradený referencovateľný identifikátor, ktorý potom využijú konzumenti dát na referencovanie objektu. Jednoduchým príkladom je IČO právnickej osoby alebo rodné číslo fyzickej osoby.

Referenčná integrita hovorí o kvalite využívania údajov medzi referenčným registrom ktorý údaj poskytuje a konzumentom, ktorý údaj používa. Preto aj pri definovaní ukazovateľov tohto parametra, sa budeme pozerať najprv očami producenta údajov (referenčný register), a potom aj z pohľadu konzumenta údajov.

Začneme ukazovateľmi zameranými na producenta údajov. Treba povedať, že popisované ukazovatele pre posúdenie kvality referencovateľného identifikátora sa netýka iba už vyhlásených referenčných registrov. Každý register obsahuje aspoň jeden objekt, ktorý spravuje. Bez ohľadu na to, či tieto údaje sú alebo nie sú vyhlásené za

© 2019 Firm name. All rights reserved.

referenčné, cieľom je, aby v krátkej dobe za referenčné vyhlásené boli a preto treba dbať na hodnotenie kvality.

3.3 Definovanie ukazovateľov dátovej kvality pre jednotlivé parametre

V tejto časti sa definujú ukazovatele prislúchajúce k jednotlivým parametrom a ich detailné popisy v tabuľke nižšie a typy ukazovateľov dátovej kvality.Ukazovatele dátovej kvality môžeme rozdeliť do dvoch skupín podľa spôsobu ako ich vyhodnocujeme.

1. **Kvalitatívne ukazovatele.**

   Tieto ukazovatele vyjadrujú kvalitu dát z architektonického pohľadu. Posúdenie a vyhodnotenie týchto ukazovateľov sa nedá vždy automatizovať a často vyžaduje manuálne vstupy. Napríklad zrozumiteľnosť atribútu je na ľudskom posúdení a vyhodnocuje skôr kvalitu architektonického návrhu dát než samotný obsah dát. Príkladom môže byť aj strojová spracovateľnosť, integrovateľnosť, interpretovateľnosť alebo pripravenosť na dátovú analytiku. Strojová spracovateľnosť závisí od formátu akým sú dáta zaznamenávané a poskytované (pdf, excel, xml) a opäť je hodnotením architektúry nie jednotlivých záznamov.

   Kvalitatívne ukazovatele sú dlhodobejšie charakteru, nemá zmysel ich vyhodnocovať príliš často. Zmena je viazaná na rozhodnutie vlastníka dát zmeniť procesy, dátový model a podobne.

2. **Kvantitatívne ukazovatele.**

   Tieto ukazovatele vyjadrujú „tvrdý“ pohľad na dáta. Príkladom môže byť počet nevyplnených hodnôt.
   Tieto ukazovatele sa zväčša dajú vyhodnocovať automatizovane. Keďže sa menia s tým, ako pribúdajú dáta, predpokladá sa, že sa budú vyhodnocovať v pravidelných, nie príliš dlhých intervaloch.
   Ak systém obsahuje veľa dát, tak pri ich postupnom náraste nedochádza k výraznejším zmenám v hodnote týchto atribútov. Aj tu, podobne ako pri kvalitatívnych ukazovateľoch, môže dochádzať k významným skokom, ak vlastník dát zavedie nový proces.

Vyhodnocovanie kvantitatívnych ukazovateľov je jednoduchšie, zväčša existuje vzorec ktorý môžeme použiť. Vyhodnotiť kvalitatívne ukazovatele je náročnejšie. Ide o posúdenie kvality procesov, architektúry. Nie vždy existuje jednoznačný vzorec či návod ako ohodnotiť daný ukazovateľ. Napriek tomu, posudzovanie kvalitatívnych ukazovateľov je dôležité, pretože poukazuje na principiálne problémy. Navyše, odstránenie niektorých takýchto problémov nemusí byť náročné.

© 2019 Firm name. All rights reserved.

Parameter	Ukazovateľ	Typ ukazovateľa	Popis ukazovateľa	Poznámka
Presnosť	Syntaktická presnosť hodnoty	Kvalitatívny	Definujeme vzdialenostnú funkciu - ako je zapísaná hodnota vzdialená od prípustných hodnôt. Pre daný atribút, napr. meno konkrétnej fyzickej osoby.	Neuvedené
	Syntaktická presnosť atribútu	Kvantitatívny	Hodnotíme, koľko záznamov má úplne presnú hodnotu, koľko záznamov má malú syntaktickú chybu, koľko záznamov má veľkú syntaktickú chybu. Pre daný atribút, napr. meno fyzickej hodnoty	Neuvedené
	Sémantická presnosť atribútu	Kvantitatívny	Hodnotíme, koľko záznamov je zaručene správnych, koľko nesprávnych a pri koľkých nevieme rozhodnúť. Pre daný atribút, napr. meno fyzickej hodnoty	Neuvedené
Konzistentnosť	Sledovanie konzistentnosti	Kvalitatívny	Posudzuje proces sledovania kvality. Pre každý atribút hovorí, či existujú biznis pravidlá, ktoré sa využívajú na kontrolu jeho konzistentnosti, alebo takéto pravidlá neexistujú.	Špeciálnym prípadom sú atribúty, kde sa konzistentnosť nesleduje. Ak napríklad atribút preberáme z referenčného registra, tak za jeho konzistentnosť zodpovedá referenčný register.
	Dodržiavanie biznis pravidla	Kvantitatívny	Pre jeden atribút, či skupinu atribútov môže byť definovaných niekoľko biznis pravidiel. Napríklad, dátum je po roku 2010, dátum nie je viac ako 5 dní v budúcnosti, vzdialenosť dvoch dátumov je najviac 10 rokov a podobne. Preto tento ukazovateľ vyhodnocujeme pre každé biznis pravidlo.	Neuvedené
Správnosť	Dodržiavanie formátu atribútu	Kvalitatívny	Hodnotíme, koľko záznamov dodržiava definovaný formát.	Neuvedené

Parameter	Ukazovateľ	Typ ukazovateľa	Popis ukazovateľa	Poznámka
			Pre daný atribút, napr. dátum	
	Design v súlade so štandardom	Kvalitatívny	Hodnotenie, ktoré objekty a atribúty sú zaznamenávané v súlade s definovanými štandardami.	Nie pre každý atribút má zmysel tento ukazovateľ vyhodnocovať, napr. pri voľnom texte.
Kompletnosť	Rozlišovanie null a prázdnej hodnoty	Kvalitatívny	Hodnotenie, či dizajn dátového modelu rozlišuje medzi úmyselne prázdnou hodnotou a hodnotou ktorá nie je vyplnená.	Neuvedené
	Vyplnenosť povinného údaja	Kvantitatívny	Neuvedené	Neuvedené
	Vyplnenosť nepovinného údaja	Kvantitatívny	Je na posúdení dátového kurátora s doménovou znalosťou, či je nebezpečnejšie, ak o atribúte nevieme či má alebo nemá byť vyplnený, čo znamená, že je šanca, že je záznam OK, ale nie sme si istí. Alebo je nebezpečnejšie, keď vieme že nám hodnota chýba, v tomto prípade zas vieme presne na čom sme.	Neuvedené
Unikátnosť	Unikátnosť identifikátora	Kvantitatívny	Neuvedené	Poznamenajme, že Unikátnosť referenčného identifikátora je definovaná inde.
	Unikátnosť atribútu	Kvantitatívny	Ukazovateľ hovoriaci o atribúte, ktorý nie je považovaný za identifikátor, ale často sa využíva ako alternatíva k identifikovaniu objektov a teda by bolo vhodné, aby bol unikátny.	Neuvedené
Aktuálnosť	Rýchlosť aktualizácie	Kvalitatívny	Ukazovateľ popisujúci procesy, ktorými sa dáta spracovávajú. Hodnotí sa, ako často sú údaje aktualizované. Cieľom je dosiahnuť online aktualizácie.	Treba však brať do úvahy volatilitu údajov a posudzovať potrebu aktualizácie údajov komplexnejšie.

Parameter	Ukazovateľ	Typ ukazovateľa	Popis ukazovateľa	Poznámka
	Aktuálnosť atribútu	Kvalitatívny	Aj keď pri parametri aktuálnosť je dôležitejšie kvalitatívne hodnotenie, ak pre atribút vieme hodnotiť ako dlho trvala jeho aktualizácia, je to užitočná informácia.	Neuvedené
Strojová spracovateľnosť	5* Open Data stupnica	Kvantitatívny	Základný ukazovateľ pre parameter strojová spracovateľnosť hodnotí, aké náročné je použiť dáta pre automatizované spracovanie v závislosti od použitého formátu. Základné informácie o stupnici boli popísané vyššie (viď Tabuľka 3).	Neuvedené
	Zrozumiteľnosť (Interpretovateľnosť)	Kvalitatívny	Hodnotíme aké komplikované je porozumieť modelu v ktorom sú dáta prezentované.	Neuvedené
	Transformovateľnosť	Kvalitatívny	Ukazovateľ hovorí ako ľahko sú dáta použiteľné pre transformáciu. Špeciálne sa dá hovoriť o pripravenosti na analytiku: koľko krokov musí používateľ urobiť, aby mohol dáta využiť pre vytváranie reportov a analýz.	Aj keď sú dáta dostupné napr. v json formáte, je rozdiel či je meno a priezvisko v jednom atribúte a musí ich prijímateľ dát oddeliť, alebo sú už predpripravené ako samostatné atribúty..
Referenčná integrita	Kompletnosť referenčného identifikátora	Kvantitatívny	Ukazovateľ hovoriaci, či majú všetky objekty evidencie priradený referenčný identifikátor. V prípade referenčného identifikátora nemá zmysel hovoriť o null hodnote, diskutovať, či môže existovať záznam, ktorý nemusí mať vyplnený tento atribút. Každý záznam musí byť identifikovateľný a teda referenčný identifikátor by mal byť vyplnený. Predpokladáme, že hodnota tohto ukazovateľa bude zväčša vysoká, chýbať môže napríklad pri niektorých historicky starších záznamoch.	Neuvedené

Parameter	Ukazovateľ	Typ ukazovateľa	Popis ukazovateľa	Poznámka
	Jedinečnosť referenčného identifikátora	Kvalitatívny	Ukazovateľ na posúdenie, či sa nepoužíva rovnaký identifikátor pre rôzne objekty. Predpokladáme, že hodnota tohto ukazovateľa bude zväčša veľmi nízka. Príkladom je napr. nejednoznačné rodné číslo, opakujúce sa ičo a podobne.	Neuvedené
	Použiteľnosť referencovateľného identifikátora	Kvalitatívny architektonický	Napr. použitie rodného čísla ako súčasti referencovateľného identifikátora zabraňuje jeho zverejňovaniu, čím sa stáva nepoužiteľným mimo chránených systémov VS.	Neuvedené
	Stabilita referencovateľného identifikátora	Kvalitatívny architektonický	Napr. ak je priezvisko súčasťou identifikátora, dá sa očakávať, že sa časom zmení v nemalom počte prípadov, čo komplikuje jeho použitie. Snahou je navrhnúť identifikátor tak, aby sa nikdy nemenil.	Neuvedené
	Formát referencovateľného identifikátora	Kvalitatívny architektonický	Očakáva sa, že referencovateľné identifikátory budú mať formát URI, pozri MetaIS.	Neuvedené
	Interoperabilita referencovateľného identifikátora	Kvalitatívny architektonický	Je vhodné, aby referencovateľný identifikátor bol navrhnutý v súlade s ontológiami, best practices zo zahraničia, aby sa zjednodušilo jeho použitie pri následnej integrácii.	Neuvedené

Parameter	Ukazovateľ	Typ ukazovateľa	Popis ukazovateľa	Poznámka
	Využívanie referencovaných objektov	Kvalitatívny/architektonický	Ukazovateľ, ktorý hovorí, koľko objektov sa pokúšame referencovať.	Napríklad register vecí eviduje informácie o vlastníkovi a používateľovi, pričom vlastníka referencuje na RFO/RPO, ale používateľa nereferencuje. Napriek tomu, že v takomto prípade by sa dalo povedať, že referencovanie je 50%, hodnotnejšia informácia je, čo sa vlastne referencuje a čo nie. Zároveň môže byť pravdou, že osoba sa referencuje voči RPO, ale nie voči RFO, čo je čiastočné referencovanie, preto hodnotenie číslom nie je výstižné. Cieľom je pre každý atribút, ktorý je evidovaný v inom registri povedať, či ho referencujeme alebo nie. Súčasťou vyhodnocovania kvality referencovania je aj informácia, či sa údaje automaticky upravujú podľa zmien v referenčnom registri, napr. či sa po zmene mena v RFO zmena prejaví aj v registri vecí. Možné prístupy sú po publikovaní informácie o zmene referenčným registrom, pravidelným zisťovaním zmien a podobne. Zákon o eGovernmente vyžaduje používanie referenčných údajov a preto prahová hodnota pre tento ukazovateľ je, že všetky objekty, pre ktoré existuje referenčný register, musia byť referencované.

Parameter	Ukazovateľ	Typ ukazovateľa	Popis ukazovateľa	Poznámka
	Podiel referencovaných objektov	Kvantitatívny	Pre každý atribút, ktorý referencujeme, sa vyhodnocuje podiel objektov, ktoré sa podarilo referencovať.	Pre nestotožnené údaje je dôležité doplniť upresňujúcu informáciu o tom, prečo sa nepodarilo údaj stotožniť: - Objekt sa v referenčnom registri nenašiel, a ani sa nájsť nemal (napr. zahraničná osoba, ktorá nie je evidovaná v RFO). Takýto údaj nepredstavuje zníženie kvality z pohľadu referenčnej integrity. - Objekt sme sa nepokúsili stotožniť s referenčným registrom, napr. historické údaje - Objekt sa v referenčnom registri nenašiel, ale mal by sa nájsť. Do veľkej miery to indikuje, že naše údaje o objekte nie sú správne. Môže sa stať, že ide o chybu referenčného registra. - Objekt sa nepodarilo jednoznačne identifikovať (napr. sme našli viacero osôb). Buď nemáme dostatok údajov a teda máme nekvalitné dáta, alebo ide o chybu v referenčnom registri.

Tabuľka 4: Definovanie ukazovateľov dátovej kvality pre jednotlivé parametre

3.4 Cieľové hodnoty pre ukazovatele dátovej kvality

3.4.1 Cieľové hodnoty parametrov

V dokumente „Strategická priorita - Manažment údajov“ sú pre jednotlivé parametre dátovej kvality definované hodnoty, ktoré je ambíciou dosiahnuť (viď Tabuľka 5).

Parameter	Ambícia	Vysvetlenie
Presnosť	1 %	Menej ako 1% objektov evidencie v ISVS, má zistené chyby
Kompletnosť	97 % 70 %	Aspoň 97% objektov evidencie v referenčných registrov, má prítomné všetky údaje vyžadované agendou. Aspoň 70% objektov evidencie ISVS, má prítomné všetky údaje vyžadované agendou.
Aktuálnosť	80 %	Aspoň 80 % objektov evidencie, má dátum aktualizácie rovnaký, ako dátum vzniku relevantnej skutočnosti.
Unikátnosť	0 %	Neexistujú duplicity
Referenčná integrita	100 % 100 %	Všetky objekty evidencie majú referenčný identifikátor Všetky údaje v objekte evidencie sú stotožnené s relevantnými referenčnými údajmi
Strojová spracovateľnosť	4 alebo 5	Aspoň 70 % objektov evidencie na úrovni 5 Aspoň 90 % objektov evidencie aspoň na úrovni 4
Konzistentnosť	100 %	Všetky údaje sú plne konzistentné
Správnosť	100 %	Všetky údaje sú úplne správne

Tabuľka 5: Ambícia cieľových hodnôt parametrov

Takto definované ambície sú naozaj ambiciózne. Niektoré z nich vyplývajú priamo zo zákona, napr. úplná referenčná integrita je vyžadovaná zákonom o eGovernmente.
Z praktického hľadiska je dosiahnutie 100% takmer nemožné. Aj pri najlepšej snahe sa chyby nedajú úplne vylúčiť. A v niektorých prípadoch je vhodné diskutovať, či náklady na dosiahnutie 100% kvality vyvážia benefity:

— Ak si predstavíme, že by sme pre posúdenie kvality niektorého atribútu museli otvoriť papierový spis, v ňom vyhľadať potrebnú informáciu a porovnať ju so záznamom, tak tento proces môže pri veľkom počte záznamov trvať roky

— Zároveň by sme museli platiť osoby, ktoré túto kontrolu manuálne vykonávajú a teda existujú aj priame finančné náklady na zvyšovanie tejto kvality

— A samozrejme, že aj v procese kontroly a opravy môže dochádzať k chybám

© 2019 Firm name. All rights reserved.

Nechceme povedať ani nijako relativizovať cieľ dosahovať 100% kvalitu dát. Pre praktické účely je však užitočné definovať nižšie čísla, ktoré hovoria nie o cieli, ktorý chceme dosiahnuť, ale o hranici, keď už z praktickéj výhody vyhlásenia dátového zdroja za kvalitný prevažujú nad rizikom problémov vyplývajúcich z nekvality dát.
V tomto dokumente preto definujeme prahové hodnoty ukazovateľov dátovej kvality (viď Tabuľka 6). Je nutné ich vnímať v tomto kontexte:

1. Ak systém dosiahol prahové hodnoty, neznamená to, že sa už nemá snažiť zvyšovať kvalitu svojich dát. Ambíciou vždy musí dosahovať 100% kvalitu.

2. Nekvalita dát môže mať rôzne dôsledky. Ak napríklad vďaka nesprávnemu záznamu o vlastníkovi vozidla príde pokuta nesprávnej osobe, stále zostáva priestor na vysvetlenie si problému a zjednanie nápravy. Nie je to príjemné, ale škoda nie je dramatická. Ak sa však napr. firma dostane na sankčný zoznam, môže to dlhodobo poškodiť jej imidž, prekaziť obchody, finančné dôsledky môžu byť veľké, aj keď sa chyba odstráni. Ešte závažnejšie je, ak by nesprávny záznam v zdravotnej dokumentácii ohrozil život osoby. Nie je preto možné brať definované prahové hodnoty ako absolútne. Osoby zodpovedné za dátovú kvalitu, znalé konkrétnej domény a dôsledkov prípadných chýb, musia posúdiť aká je prahová hodnota pre ich informačný systém v celom kontexte.

3. Je vhodné rozdeliť údaje na viacero skupín, napr. záznamy pred rokom 2010, záznamy po roku 2010. A vyhodnocovať kvalitu na týchto množinách, čo umožní lepšie posudzovať kvalitu a môže ukázať, že existuje užitočná podmnožina dát, ktorá spĺňa prahové hodnoty a dá sa jej teda dôverovať.

© 2019 Firm name. All rights reserved.

3.4.2 Prahové hodnoty ukazovateľov

Parameter	Ukazovateľ	Názov prahovej hodnoty	Limit prahovej hodnoty	Popis prahovej hodnoty
Presnosť	Syntaktická presnosť hodnoty	Neuvedené	NA	Neuvedené
	Syntaktická presnosť atribútu	Syntakticky presné záznamy	> 90%	Aspoň 90% záznamov je syntakticky bezchybných
		Záznamy s veľkou syntaktickou chybou	< 5%	Najviac 5% záznamov má veľkú syntaktickú chybu (a teda je ešte ďalších maximálne 5% údajov s malou syntaktickou chybou).
	Sémantická presnosť atribútu	Sémanticky presné záznamy	> 85%	Aspoň 85% je presných. Toto číslo nemôže byť väčšie ako syntaktická presnosť.
		Záznamy, ktoré sú nepresné	< 5%	Najviac 5% záznamov, pri ktorých vieme, že nie sú presné. Zvyšné záznamy majú neznámu sémantickú presnosť.
	Sledovanie konzistentnosti	Atribúty so sledovanou konzistentnosťou	> 80%	Aspoň 80% atribútov má definované biznis pravidlá pre ich konzistentnosť.
	Dodržiavanie biznis pravidla	Dodržané biznis pravidlo	> 90%	Aspoň 90% záznamov dodržiava definované biznis pravidlo.
Správnosť	Dodržiavanie formátu atribútu	Dodržaný formát	> 99%	Aspoň 99% dodržiava formát. Ak by bolo viac záznamov nedodržiavajúcich formát, pravdepodobne ide systémovú chybu, preto je tu prahová hodnota tak vysoká.
	Design v súlade so štandardom	Objekty dizajnované v súlade so štandardom	> 99%	Aspoň 99% objektov dodržiava štandard. Ak by objekt nebol v súlade so štandardom, je to výrazná prekážka

Parameter	Ukazovateľ	Názov prahovej hodnoty	Limit prahovej hodnoty	Popis prahovej hodnoty
				pre integráciu a ďalšiu použiteľnosť týchto údajov. Ide o dizajnové rozhodnutia.
		Atribúty spĺňajúce formát podľa štandardu	> 99%	Aspoň 99% atribútov musí dodržiavať štandard. Opäť, nedodržanie štandardu komplikuje integráciu a ďalšiu použiteľnosť týchto údajov.
Kompletnosť	Rozlišovanie null a prázdnej hodnoty	Neuvedené	NA	Neuvedené
	Vyplnenosť povinného údaja	Vyplnenosť povinného údaja	> 99%	Aspoň 99% záznamov má vyplnený povinný údaj. Keďže ide o povinný údaj, ak nie je vyplnený, tak samotný záznam má výrazne nižšiu hodnotu.
	Vyplnenosť nepovinného údaja	Vyplnenosť nepovinného údaja	> 90%	Aspoň 90% záznamov má vyplnený nepovinný údaj alebo vieme, že prázdna hodnota je tam úmyselne (čo sa dá považovať za druh vyplnenia).
		Nejasná vyplnenosť povinného údaju	< 10%	Nevyplnená hodnota, nevieme či má alebo nemá byť vyplnená.
		Chýbajúce údaje	< 10%	Nevyplnená hodnota, vieme že má byť vyplnená. K týmto prípadom pravdepodobne nedochádza často.
Unikátnosť	Unikátnosť identifikátora	Unikátne záznamy	> 90%	Aspoň 90% záznamov je zaručene unikátnych. To znamená, že je najviac 10% záznamov, ktoré sú duplicitné, alebo nie sme si istí, či sú, alebo nie sú unikátne
	Unikátnosť atribútu	Unikátne záznamy	> 80%	Aspoň 80% záznamov je zaručene unikátnych.

Parameter	Ukazovateľ	Názov prahovej hodnoty	Limit prahovej hodnoty	Popis prahovej hodnoty
Aktuálnosť	Rýchlosť aktualizácie	Neuvedené	NA	Neuvedené
	Aktuálnosť atribútu	Neuvedené	NA	Neuvedené
Strojová spracovateľnosť	5* Open Data stupnica	Neuvedené	NA	Minimálnou hodnotou je 3★ pre publikovanie datasetov. Ako prahovú hodnotu však označíme až 4★, pretože až tá umožňuje integráciu.
	Zrozumiteľnosť (Interpretovateľnosť)	Neuvedené	NA	Neuvedené
	Transformovateľnosť	Neuvedené	NA	Neuvedené
Referenčná integrita	Kompletnosť referenčného identifikátora	Kompletnosť	100%	Všetky záznamy musia mať vyplnený referenčný identifikátor, inak sú nepoužiteľné.
	Jedinečnosť referenčného identifikátora	Duplicity	>95%	Skoro všetky záznamy musia byť jednoznačné, inak sú nepoužiteľné. V niektorých prípadoch môže byť procesne náročné dodatočne zmeniť historicky duplicitné identifikátory. Bolo by škodou nepoužiť takmer úplne správne údaje.
	Použiteľnosť referencovateľného identifikátora	Neuvedené	NA	Neuvedené
	Stabilita referencovateľného identifikátora	Neuvedené	NA	Neuvedené

Parameter	Ukazovateľ	Názov prahovej hodnoty	Limit prahovej hodnoty	Popis prahovej hodnoty
	Formát referencovateľného identifikátora	Neuvedené	NA	Neuvedené
	Interoperabilita referencovateľného identifikátora	Neuvedené	NA	Neuvedené
	Využívanie referencovaných objektov	Neuvedené	NA	Neuvedené
	Podiel referencovaných objektov	Referencované objekty	> 90%	Aspoň 90% objektov je referencovaných. Objekty, ktoré nemá zmysel referencovať, neberieme do úvahy.
		Objekty, ktoré sme neskúšali stotožniť	< 1%	Všetky objekty by sme sa mali pokúsiť stotožniť, iba vo veľmi výnimočných prípadoch (staré historické záznamy) nemusíme.

Tabuľka 6: Prahové hodnoty ukazovateľov

4 Technický návod na výpočet parametrov dátovej kvality

4.1 Úvod k technickému návodu na výpočet parametrov dátovej kvality

Kvalita dát sa meria pomocou výsledných hodnôt ukazovateľov príslušných parametrov. V predchádzajúcej časti boli vysvetlené prahové hodnoty ukazovateľov ako aj ambicióznych hodnôt parametrov. Na zachovanie konzistentnosti by rezorty mali používať jednotné spôsoby a spoločnú metodiku výpočtu parametrov dátovej kvality. Momentálne je predstavený zoznam 25-ich ukazovateľov, pričom tento zoznam môže byť v budúcnosti doplnený o ďalšie parametre a metriky. Celý koncept je škálovateľný a parametrizovateľný.

4.2 Návod na počítanie ukazovateľov pre parametre dátovej kvality v praxi

Návod ma pomôcť ako prakticky vypočítať tieto konkrétne hodnoty. K jednotlivým parametrom dátovej kvality sú definície ukazovateľov aj s príkladmi a pomocnými výpočtovými vzorcami.

4.2.1 Biznis pravidlá pre potreby ukazovateľov dátovej kvality

Prečo sú biznis pravidlá dôležité?

Bez dobrej znalosti biznis pravidiel nie je možné efektívne merať ukazovatele dátovej kvality. Tieto princípy alebo usmernenia vysvetľujú interakcie a stanovené pravidlá pre akcie a výsledné správanie a integritu dát. Biznis pravidlá často popisujú biznis procesy, z ktorých sa dajú odvodiť ukazovatele dátovej kvality a pomáhajú ľahšie identifikovať formát dátových atribútov.

Kde sa uplatňujú biznis pravidlá?

V dátovej kvalite sa biznis pravidlá uplatňujú na jednotlivých atribútoch dátových entít datasetu. Často pomáhajú odpovedať na otázku, ako má dátový atribút vyzerať.

Kto definuje biznis pravidlá?

Osoby alebo organizácie, ktoré používajú dáta, definujú biznis pravidlá. V niektorých prípadoch biznis pravidlá vychádzajú primárne z existujúcej legislatívy a dodatočne sú rozširované so zameraním na ďalšie používanie dát. Rozširovania biznis pravidiel môžu spôsobiť aj potreby informačného systému, ktorý s dátami pracuje.

Kto riadi, zbiera a spravuje biznis pravidlá?

Vlastníci datasetov, napr. vlastníci referenčných registrov.

© 2019 Firm name. All rights reserved.

Prečo je potrebné centrálne riadenie biznis pravidiel?

V dynamickom prostredí neustálych legislatívnych zmien a komplexného využívania dát rôznymi organizáciami verejného aj súkromného sektoru, nie je jednoduché efektívne riadiť biznis pravidlá. Centrálne riadenie túto úlohu podstatne zjednodušuje. Jednotlivé pravidlá je potrebné párovať na konkrétne atribúty dátových entít centrálneho dátového modelu verejnej správy. Narastá aj potreba aplikovania rovnakých biznis pravidiel naprieč všetkými informačnými systémami, ktoré pracujú s identickými atribútmi dátových entít v celej verejnej správe. Biznis pravidlá, ktoré striktne vychádzajú z legislatívy, je potrebné pripojiť k relevantným zákonom. Napríklad pravidlá často potrebujú zdôrazniť existenciu verejných referenčných číselníkov, ktoré zjednocujú povolené hodnoty pre zvolené atribúty dátových entít. Tvorenie týchto číselníkov je vedľajší efekt centrálneho riadenia biznis pravidiel a priamo podporuje schopnosť merať dátovú kvalitu. Biznis pravidlá sa neustále vyvíjajú. Vznikajú, zanikajú a menia sa na rôznych miestach v rôznych inštitúciách. Vývoj týchto pravidiel je potrebné monitorovať a zdieľať medzi inštitúciami, ktoré si ich potrebujú osvojiť. Jedna z úloh centrálneho riadenia biznis pravidiel je zabezpečiť túto koordináciu.

4.2.2 Zoznam parametrov a ich ukazovatele

Jednoduchá tabuľka ôsmich parametrov a k nim prislúchajúcim 25 ukazovateľom (viď Tabuľka 7)

Parameter	Ukazovateľ
Presnosť	Syntaktická presnosť hodnoty (viď Tabuľka 8)
	Syntaktická presnosť atribútu (viď Tabuľka 9)
	Sémantická presnosť atribútu (viď Tabuľka 10)
Konzistentnosť	Sledovanie konzistentnosti (viď Tabuľka 11)
	Dodržiavanie biznis pravidla (viď Tabuľka 12)
Správnosť	Dodržiavanie formátu atribútu (viď Tabuľka 13)
	Design v súlade so štandardom (viď Tabuľka 14)
Kompletnosť	Rozlišovanie ‘null‘ a prázdnej hodnoty (viď Tabuľka 15)
	Vyplnenosť povinného údaja (viď Tabuľka 16)
	Vyplnenosť nepovinného údaja (viď Tabuľka 17)
Unikátnosť	Unikátnosť identifikátora (viď Tabuľka 18)
	Unikátnosť atribútu (viď Tabuľka 19)
Aktuálnosť	Rýchlosť aktualizácie (viď Tabuľka 20)

© 2019 Firm name. All rights reserved.

Parameter	Ukazovateľ
	Aktuálnosť atribútu (viď Tabuľka 21)
Strojová spracovateľnosť	5★ Open Data stupnica (viď Tabuľka 22)
	Zrozumiteľnosť (interpretovateľnosť) (viď Tabuľka 23)
	Transformovateľnosť (viď Tabuľka 24)
Referenčná integrita	Kompletnosť referenčného identifikátora (viď Tabuľka 25)
	Jedinečnosť referenčného identifikátora (viď Tabuľka 26)
	Použiteľnosť referencovateľného identifikátora (viď Tabuľka 27)
	Stabilita referencovateľného identifikátora (viď Tabuľka 28)
	Formát referencovateľného identifikátora (viď Tabuľka 29)
	Interoperabilita referencovateľného identifikátora (viď Tabuľka 30)
	Využívanie referencovaných objektov (viď Tabuľka 31)
	Podiel referencovaných objektov (viď Tabuľka 32)

Tabuľka 7: Zoznam parametrov a ich ukazovateľov

4.2.2.1 *Presnosť*

Syntaktická presnosť hodnoty

Názov ukazovateľa	**Syntaktická presnosť hodnoty**
Definícia	Miera rozdielnosti uložených hodnôt od reálnych (prípustných) hodnôt.
Metrika	Vzdialenosť hodnôt.
Rozsah	Akýkoľvek objekt, ktorý môže byť reprezentovaný záznamom v databáze alebo datasete.
Jednotka metriky	Počet odlišných znakov.
Príbuzné parametre	Správnosť, Kompletnosť, Konzistencia.
Voliteľnosť	Povinné – ak sú dáta nepresné, môžu byť nevhodné na používanie.
Príklad	Ak pre osobu s menom Dagmara Závodská existuje v databáze záznam ‘Dagmara Závodská’, ide o úplnú syntaktickú presnosť. Ak má záznam hodnotu ‘Dgmara Závodská‘, ide o nepresnosť v 1 znaku.

© 2019 Firm name. All rights reserved.

Pseudokód (výpočtový vzorec)	DIFF(‘db_value‘, ‘expected_value‘).

Tabuľka 8: Ukazovateľ Syntaktická presnosť hodnoty

Syntaktická presnosť atribútu

Názov ukazovateľa	**Syntaktická presnosť atribútu**
Definícia	Miera, do akej dáta správne reprezentujú skutočnosť.
Metrika	Pomer uložených hodnôt bez syntaktickej chyby, uložených hodnôt s malou syntaktickou chybou a hodnôt s veľkou syntaktickou chybou.
Rozsah	Akýkoľvek objekt, ktorý môže byť reprezentovaný záznamom v databáze alebo datasete.
Jednotka metriky	Percento.
Príbuzné parametre	Správnosť, Kompletnosť, Konzistencia.
Voliteľnosť	Povinné – ak dáta obsahujú veľké chyby, nie sú vhodné na používanie.
Príklad	Ak v databáze obsahujúcej 100 záznamov majú 2 osoby preklep v uloženom mene a 1 osoba má úplne nesprávne meno, ide o 97 % syntaktickú presnosť atribútu ‘meno‘ a 1 % zastúpenie záznamov s veľkou chybou.
Pseudokód (výpočtový vzorec)	100 * (COUNT (‘db_value‘ == ‘expected_value‘) / COUNT ‘db_value‘); 100 * (COUNT (‘db_value‘ != ANY (‘similar_values‘)) / COUNT ‘db_value‘).

Tabuľka 9: Ukazovateľ Syntaktická presnosť atribútu

Sémantická presnosť atribútu

Názov ukazovateľa	**Sémantická presnosť atribútu**
Definícia	Miera, do akej sú dáta zaručene správne, nesprávne, resp. nevieme rozhodnúť.
Metrika	Pomer uložených hodnôt bez akejkoľvek chyby, uložených hodnôt s chybou a hodnôt, pri ktorých nie je možné rozhodnúť.
Rozsah	Akýkoľvek objekt, ktorý môže byť reprezentovaný záznamom v databáze alebo datasete.
Jednotka metriky	Percento.
Príbuzné parametre	Správnosť, Kompletnosť, Konzistencia.
Voliteľnosť	Povinné – ak sú dáta nepresné alebo nejednoznačné, môžu byť nevhodné na používanie.

© 2019 Firm name. All rights reserved.

Príklad	Ak v databáze obsahujúcej 100 záznamov má 1 osoba meno uložené v tvare, kde nie je chyba jednoznačne opraviteľná (napr. ‘Ea‘ môže predstavovať chybu v mene ‘Eva‘, ‘Ema’ alebo ‘Ela’), ide o 99 % sémantickú presnosť atribútu ‘meno‘.
Pseudokód (výpočtový vzorec)	100 * (COUNT (‘db_value‘ == ‘expected_value‘) / COUNT ‘db_value‘).

Tabuľka 10: Ukazovateľ Sémantická presnosť atribútu

4.2.2.2 Konzistentnosť

Sledovanie konzistentnosti

Názov ukazovateľa	**Sledovanie konzistentnosti**
Definícia	Kvalita nastavenia dátového modelu z pohľadu konzistentnosti.
Metrika	Hodnotenie na definovanej stupnici.
Rozsah	Jednotlivé atribúty dátového modelu.
Jednotka metriky	— Definované pravidlá — Čiastočne definované pravidlá — Nedefinované pravidlá — Nie je potrebné kontrolovať
Príbuzné parametre	Presnosť.
Voliteľnosť	Odporúčané – hodnotenie týmto ukazovateľom je často aj návodom ako zlepšiť dátový model.
Príklad	Jedná sa o kvalitatívny ukazovateľ, kde hodnotíme jednotlivé parametre alebo skupiny parametrov, či k nim existujú pravidlá pre sledovanie/kontrolu konzistentnosti. Posúdenie je často subjektívne a pre jeho vykonanie je potrebná dobrá znalosť domény a skúsenosť. Predpokladajme, že vývoj ceny komponentov v databáze má atribúty minimálna a maximálna cena. — Ak neexistujú žiadne pravidlá pre minimálnu a maximálnu cenu, tak hodnota ukazovateľa je „Nedefinované pravidlá“ — Ak existuje pravidlo, pri ktorom platí, že cena_min ≤ cena_max. Hodnota ukazovateľa je „Čiastočne definované pravidlá“. V prípade, že nemá zmysel definovať ďalšie pravidlá, tak „Definované pravidlá“ — Ak existujú aj ďalšie pravidlá, napr. 0 ≤ cena_min alebo cena_max ≤ 1000*cena_min. Je výlučne na danej doméne, aké pravidlá majú ešte zmysel a teda na subjektívnom posúdení, čí je hodnota ukazovateľa „Čiastočne definované pravidlá“ alebo „Definované pravidlá“

© 2019 Firm name. All rights reserved.

	— Pre niektoré atribúty nemá zmysel definovať pravidlá, napr. poznámka je voľný text, ktorý nechceme nijak kontrolovať. V takomto prípade je hodnota ukazovateľa „Nie je potrebné kontrolovať“ Pravidlá môžu byť definované na rôznej úrovni, napr. ako constraints v databáze, v samostatnom rule engine, ako súčasť biznis logiky a podobne.
Pseudokód (výpočtový vzorec)	Nepočíta sa na základe údajov v databáze, ale určuje sa subjektívnym posúdením dátového modelu.

Tabuľka 11: Ukazovateľ Sledovanie konzistentnosti

Dodržiavanie biznis pravidla

Názov ukazovateľa	**Dodržiavanie biznis pravidla**
Definícia	Miera dodržiavania biznis pravidla (ak existuje) pre konkrétny atribút.
Metrika	Analýza opakujúceho vzoru a frekvencie hodnoty.
Rozsah	Hodnoty a formáty v dátach, záznamoch, súboroch a databázach .
Jednotka metriky	Percento.
Príbuzné parametre	Presnosť a unikátnosť.
Voliteľnosť	Konzistentnosť je možná aj bez presnosti.
Príklad	Dátum narodenia študenta má rovnakú hodnotu a formát v registri škôl ako aj v databáze študenta konkrétnej školy.
Pseudokód (výpočtový vzorec)	Select COUNT DISTINCT on `Datum narodenia`.

Tabuľka 12: Ukazovateľ Dodržiavanie biznis pravidla

4.2.2.3 Správnosť

Dodržiavanie formátu atribútu

Názov ukazovateľa	**Dodržiavanie formátu atribútu**
Definícia	Miera dodržiavania definovaného formátu atribútu.
Metrika	Miera, do akej dáta odrážajú reálne vlastnosti objektu.
Rozsah	Akýkoľvek objekt, ktorý môže byť reprezentovaný ako dátová položka, záznam, dataset alebo databáza.
Jednotka metriky	Percento údajov, ktoré prechádzajú pravidlami presnosti údajov.

© 2019 Firm name. All rights reserved.

Príbuzné parametre	Presnosť, Úplnosť, Konzistentnosť a Unikátnosť.
Voliteľnosť	-
Príklad	Zdravotná sestra zadávala informácie do systému 1. Mája 2018, nedávno emigrovala z USA a nebola veľmi oboznámená s Európskym systémom. Výsledkom bolo, že zadala všetky údaje o narodení vo formáte MM/DD/YYYY namiesto požadovaného formátu DD/MM/YYYY. Zadané dáta prešli kontrolou overenia systému, pretože hodnoty boli v rámci rozsahu. Hodnoty nie sú správne a systém prijal všetky deti, ktoré sa narodili 5. Januára 2018.
Pseudokód (výpočtový vzorec)	((Count of accurate objects)/(Count of accurate objects + Counts of inaccurate objects)*100.

Tabuľka 13: Ukazovateľ Dodržiavanie formátu atribútu

Design v súlade so štandardom

Názov ukazovateľa	**Design v súlade so štandardom**
Definícia	Posúdenie kvality dátového modelu a jeho súladu s definovanými štandardami.
Metrika	Áno / Nie / -.
Rozsah	Akýkoľvek objekt, ktorý môže byť reprezentovaný ako dátová položka, záznam, dataset alebo databáza, na ktorý sa vzťahujú štandardy.
Jednotka metriky	Áno / Nie / -.
Príbuzné parametre	Presnosť, Úplnosť, Konzistentnosť.
Voliteľnosť	Povinné – hodnotenie týmto ukazovateľom je často aj návodom ako zlepšiť dátový model a dodržiavanie štandardov je vynucované aj legislatívou. Navyše nedodržanie štandardu znemožňuje integráciu a výmenu dát.
Príklad	Ako príklad uvedieme súlad objektu Fyzická osoba s výnosom o štandardoch pre ISVS. V prílohe 2 tohto štandardu je definovaný katalóg dátových prvkov a v ňom je pre fyzickú osobu definované, aké informácie musí objekt fyzická osoba obsahovať, ako majú byť jednotlivé atribúty nazvané. Napríklad: PhysicalPerson obsahuje PersonName, AlternativeName, MaritalStatus, ... Samotný atribút PersonName obsahuje GivenName, FamilyName, OtherName ... . Nie je požadované, aby atribúty v databáze boli pomenované presne v súlade s výnosom, používanie týchto názvov sa požaduje až pri výmene informácií medzi systémami. Na druhú stranu, návrh databázového modelu musí umožňovať vytvorenie dátového prvku v súlade so štandardom, takže musíme mať napríklad oddelené meno, priezvisko a podobne.
Pseudokód (výpočtový vzorec)	Nepočíta sa na základe údajov v databáze, ale určuje sa subjektívnym posúdením dátového modelu.

Tabuľka 14: Ukazovateľ Design v súlade so štandardom

© 2019 Firm name. All rights reserved.

4.2.2.4 Kompletnosť

Rozlišovanie „null“ a prázdnej hodnoty

Názov ukazovateľa	**Rozlišovanie „null“ a prázdnej hodnoty**
Definícia	Posúdenie kvality dátového modelu z pohľadu zaznamenávania prázdnych hodnôt.
Metrika	Áno / Nie.
Rozsah	Akýkoľvek atribút.
Jednotka metriky	Áno / Nie.
Príbuzné parametre	Presnosť, Správnosť, Konzistentnosť.
Voliteľnosť	Odporúčané – hodnotenie týmto ukazovateľom je často aj návodom ako zlepšiť dátový model.
Príklad	Cieľom tohto ukazovateľa je posúdiť, či vieme oddeliť prázdnu hodnotu od nevyplnenej hodnoty. Ako príklad predpokladajme, že v databáze máme atribút počet detí, čo je celé číslo. Mali by sme vedieť rozlíšiť, či niekto má 0 detí (nemá žiadne deti) alebo nevieme či má alebo nemá deti. — Ak dátové pole umožňuje nevyplnenie hodnoty, tak nevyplnená hodnota atribútu znamená „nevieme, či má deti“, zatiaľ čo hodnota atribútu 0 znamená „nemá deti“. Pri takomto modeli je hodnota ukazovateľa „áno“, pretože vieme rozlíšiť prázdnu a nulovú hodnotu. — Ak dátové pole neumožňuje nevyplnenie hodnoty (tzn. že vždy musí byť vyplnené nejaké číslo), tak môžeme použiť napr. hodnotu -1 na reprezentáciu toho, že nevieme, či osoba má alebo nemá deti. Pri takomto modeli je hodnota ukazovateľa „áno“, pretože vieme rozlíšiť prázdnu a nulovú hodnotu. — Ak dátové pole neumožňuje nevyplnenie hodnoty (tzn. že vždy musí byť vyplnené nejaké číslo), a používame hodnoty 0, 1, 2, ..., tak pri čísle 0 nevieme rozlíšiť, či daná osoba nemá deti, alebo to nevieme. Pri takomto modeli je hodnota ukazovateľa „nie“, pretože nevieme rozlíšiť prázdnu a nulovú hodnotu.
Pseudokód (výpočtový vzorec)	Nepočíta sa na základe údajov v databáze, ale určuje sa subjektívnym posúdením dátového modelu.

Tabuľka 15: Ukazovateľ Rozlišovanie "null" a prázdnej hodnoty

Vyplnenosť povinného údaja

Názov	**Vyplnenosť povinného údaja**
Definícia	Miera, do akej sú v databáze alebo datasete vyplnené povinné záznamy.
Metrika	Pomer uložených dát v tabuľke voči tabuľke vyplnenej na 100 %.

© 2019 Firm name. All rights reserved.

Rozsah	Akýkoľvek údaj, ktorý je označený biznis pravidlami ako povinný.
Jednotka metriky	Percento.
Príbuzné parametre	Správnosť, Presnosť.
Voliteľnosť	Povinný údaj.
Príklad	Každý zamestnanec vypĺňa formulár s kontaktnými údajmi. Ten zahŕňa aj povinné vyplnenie telefónneho čísla. Z 387 zamestnancov nemal telefónne číslo vyplnený 1 zamestnanec (nevlastní osobný telefón), teda položka telefónne číslo v databáze zamestnancov dosahuje 100 * 386/387 = 99,7 % kompletnosť.
Pseudokód (výpočtový vzorec)	100 * (COUNT ‘tel_number‘ WHERE NOT blank) / (COUNT ‘employees‘)

Tabuľka 16: Ukazovateľ Vyplnenosť povinného údaja

Vyplnenosť nepovinného údaja

Názov ukazovateľa	**Vyplnenosť nepovinného údaja**
Definícia	Miera, do akej sú v databáze alebo datasete vyplnené nepovinné záznamy.
Metrika	Pomer uložených dát v tabuľke voči tabuľke vyplnenej na 100 %.
Rozsah	Akýkoľvek údaj, ktoré nie je označený biznis pravidlami ako povinný.
Jednotka metriky	Percento.
Príbuzné parametre	Správnosť, Presnosť.
Voliteľnosť	-
Príklad	Každý zamestnanec vypĺňa formulár s kontaktnými údajmi. Ten zahŕňa aj nepovinné vyplnenie ICE čísla (In Case of Emergency) pre prípad núdzovej situácie. Z 387 zamestnancov malo ICE číslo vyplnené 364 zamestnancov, teda položka ICE číslo v databáze zamestnancov dosahuje 100 * 364/387 = 94% kompletnosť.
Pseudokód (výpočtový vzorec)	100 * (COUNT ‘ICE_number‘ WHERE NOT blank) / (COUNT ‘employees‘).

Tabuľka 17: Ukazovateľ Vyplnenosť nepovinného údaja

© 2019 Firm name. All rights reserved.

4.2.2.5 Unikátnosť

Unikátnosť identifikátora

Názov ukazovateľa	**Unikátnosť identifikátora**
Definícia	Každý záznam musí mať unikátny primárny kľuč. Identifikátor musí byť vždy známy tzn. nikdy nesmie byť NULL.
Metrika	Pomer identifikátorov ktorých hodnota sa opakuje, ku počtu všetkých záznamov.
Rozsah	Iba pre identifikátory.
Jednotka metriky	Percento.
Príbuzné parametre	Konzistencia.
Voliteľnosť	Povinné – unikátnosť identifikátora je jedným z kľúčových ukazovateľov.
Príklad	Predpokladajme, že v databáze máme 10 000 záznamov, kde dva záznamy majú rovnaký identifikátor (napr. „12345“) a iné tri záznamy tiež rovnaký identifikátor (napr. „6789“), takže celkovo je 5 nejednoznačných záznamov s identifikátormi. Ak databáza obsahuje 10 000 záznamov, unikátnosť identifikátora je 9 995 / 10 000 = 0.9995, teda 99.95%.
Pseudokód (výpočtový vzorec)	100 * ((COUNT(all_values) - COUNT(duplicate_values)) / COUNT(all_values)).

Tabuľka 18: Ukazovateľ Unikátnosť identifikátora

Unikátnosť atribútu

Názov ukazovateľa	**Unikátnosť atribútu**
Definícia	Nič sa nezaznamená viac ako 1 krát na základe toho, ako sa vec identifikuje.
Metrika	Analýza počtu vecí hodnotených v reálnom svete v porovnávaní s počtom záznamov v datasete.
Rozsah	Meraný voči všetkým záznamom v rámci jedného datasetu.
Jednotka metriky	Percento.
Príbuzné parametre	Konzistencia.
Voliteľnosť	Záleží na okolnostiach.
Príklad	Škola ma 120 súčasných študentov a 380 bývalých študentov (t.j. celkovo 500). Avšak databáza študenta ukazuje 520 odlišných študentských záznamov. Môže obsahovať meno Jožo Mrkvička a Jožko Mrkvička ako

© 2019 Firm name. All rights reserved.

	samostatný záznam aj napriek skutočnosti, že v škole existuje len 1 študent menom Jožko Mrkvička. To znamená, že unikátnosť je 500/520 = 96,2%.
Pseudokód (výpočtový vzorec)	(Number of things in real world) / (Number of records describing different things).

Tabuľka 19: Ukazovateľ Unikátnosť atribútu

4.2.2.6 *Aktuálnosť*

Rýchlosť aktualizácie

Názov ukazovateľa	**Rýchlosť aktualizácie**
Definícia	Hodnotenie ako často sú údaje aktualizované.
Metrika	Dĺžka časového intervalu medzi jednotlivými aktualizáciami .
Rozsah	Akýkoľvek údaj, ktorý môže byť reprezentovaný záznamom v databáze alebo datasete.
Jednotka metriky	Časové jednotky (sekundy, minúty, hodiny, dni, ..) podľa dĺžky intervalu.
Príbuzné parametre	Presnosť.
Voliteľnosť	Povinné – ak dáta nie sú pravidelne aktualizované, nie je možné zabezpečiť ich kvalitu.
Príklad	24 hodín – Typickým príkladom je spúšťanie pravidelných jobov počas noci, ktoré zbierajú údaje z distribuovaných databáz do centrálnej databázy a teda údaje sú aktualizované raz za 24 hodín. 5 min – Údaje sú synchronizované medzi viacerými inštanciami, pričom synchronizačné procesy si vymieňajú zmenené údaje každých 5 minút. 0 s – údaje sú „online“, neustále aktualizované po ich zadaní. V prípade, že ide o distribuovaný systém, aj pri okamžitej aktualizácii je treba počítať s istým technologickým zdržaním, ktoré je vhodné transparentne deklarovať.
Pseudokód (výpočtový vzorec)	Nepočíta sa na základe údajov v databáze, ale určuje sa na základe hodnotenia procesov aktualizácie dát.

Tabuľka 20: Ukazovateľ Rýchlosť aktualizácie

© 2019 Firm name. All rights reserved.

Aktuálnosť atribútu

Názov ukazovateľa	**Aktuálnosť atribútu**
Definícia	Priemerná doba medzi zmenou a aktualizáciou atribútu.
Metrika	Dĺžka časového intervalu medzi zmenou a aktualizáciou atribútu.
Rozsah	Akýkoľvek atribút, ktorý môže byť reprezentovaný záznamom v databáze alebo datasete.
Jednotka metriky	Časové jednotky (sekundy, minúty, hodiny, dni, ..) podľa dĺžky intervalu.
Príbuzné parametre	Presnosť.
Voliteľnosť	Nepovinné – dôležitejšou informáciou je ukazovateľ Rýchlosť aktualizácie, ktorý hovorí o systémových chybách v aktualizácii dát.
Príklad	Výpočet budeme ilustrovať na príklade atribútu, ktorý sa aktualizuje raz za 24 hodín, napríklad o 02:00 ráno. Pre každý atribút, ktorý sa za posledných 24 hodín zmenil, vypočítame dĺžku intervalu medzi jeho zmenou a aktualizáciou. Ak sa napríklad záznam v lokálnej databáze zmenil o 9:30, tak dĺžka intervalu je 02:00 – 09:30 = 26:00 – 9:30 = 16:30, takže 16 a pol hodiny. Ak sa záznam v lokálnej databáze zmenil o 21:00, tak dĺžka intervalu je 02:00 – 21:00 = 26:00 – 21:00 = 5:00, takže 5 hodín. Ak sa za posledných 24 hodín zmenili iba tieto dva atribúty, tak priemerná aktuálnosť atribútu je (16,5+5)/2 = 10,75 hodiny, takže 10 hodín a 45 minút.
Pseudokód (výpočtový vzorec)	SUM (‘Time_of_actualization‘ – ‘Time_of_change‘) / COUNT(‘changed_values‘).

Tabuľka 21: Ukazovateľ Aktuálnosť atribútu

4.2.2.7 *Strojová spracovateľnosť*

5★ Open Data Stupnica

Názov ukazovateľa	**5★ Open Data stupnica**
Definícia	Aké náročné je použiť dáta pre automatizované spracovanie v závislosti od použitého formátu.
Metrika	Hodnotenie na definovanej stupnici.
Rozsah	Hodnotíme celý dataset.
Jednotka metriky	Stupnica s definovanými hodnotami, podľa 5★ Open Data stupnica ★ Dáta sú dostupné, bez ohľadu na formát ★★ Dáta sú publikované v štruktúrovanej podobe

© 2019 Firm name. All rights reserved.

	★★★ Dáta publikované v štruktúrovanej podobe v niektorom otvorenom formáte ★★★★ Dáta obsahujú URI (identifikátory), vďaka čomu sa na ne dá odkazovať ★★★★★ Dáta sú prepojené s inými dátovými zdrojmi
Príbuzné parametre	Konzistentnosť, Správnosť.
Voliteľnosť	Povinné – aj keď nejde o počítané kritérium, hodnoty sú dobre definované a ukazovateľ dáva dobrú informáciu o použiteľnosti dát pre automatizáciu, integráciu.
Príklad	Príklady formátov pre jednotlivé úrovne strojovej spracovateľnosti ★ pdf, image, scan dokumentu ★★ excel ★★★ csv ★★★★ rdf – Resource Description Framework ★★★★★ lod – linked open data
Pseudokód (výpočtový vzorec)	Nepočíta sa na základe údajov v databáze, ale určuje sa subjektívnym hodnotením.

Tabuľka 22: Ukazovateľ 5★ Open Data Stupnica

Zrozumiteľnosť (interpretovateľnosť)

Názov ukazovateľa	**Zrozumiteľnosť (interpretovateľnosť)**
Definícia	Miera komplikovanosti dátového modelu.
Metrika	Hodnotenie na definovanej stupnici.
Rozsah	Hodnotíme celý dataset, prípadne jeho vybranú podmnožinu.
Jednotka metriky	Stupnica s definovanými hodnotami: — Výborná — Dobrá — Uspokojivá — Slabá — Zlá

© 2019 Firm name. All rights reserved.

Príbuzné parametre	Presnosť, Konzistentnosť, Správnosť.
Voliteľnosť	Odporúčané – aj keď ide o subjektívne hodnotenie zrozumiteľnosti, odporúča sa ho urobiť, pretože dôvody, pre ktoré hodnotiteľ znížil mieru zrozumiteľnosti poskytujú návod na zvýšenie kvality údajov.
Príklad	Do veľkej miery ide o subjektívne hodnotenie, ale môže sa podporiť všímaním si viacerých kritérií, napríklad: — Používanie skratiek v názvoch atribútov, skratky často komplikujú zrozumiteľnosť — Používanie výstižných a zaužívaných názvov, napr. generické názvy atribútov ako riadok 1, riadok 2, hodnota a podobne znižujú zrozumiteľnosť — Počet väzieb medzi objektami — Hĺbka vnorení jednotlivých objektov
Pseudokód (výpočtový vzorec)	Nepočíta sa na základe údajov v databáze, ale určuje sa subjektívnym hodnotením, ktoré v niektorých prípadoch môže byť aj komplexné (vtedy poskytuje návod na zlepšenie kvality dát).

Tabuľka 23: Ukazovateľ Zrozumiteľnosť (interpretovateľnosť)

Transformovateľnosť

Názov ukazovateľa	**Transformovateľnosť**
Definícia	Miera pripravenosti dát na ďalšie spracovanie.
Metrika	Hodnotenie na definovanej stupnici.
Rozsah	Hodnotí sa celý dataset, prípadne jeho vybraná podmnožina.
Jednotka metriky	Stupnica s definovanými hodnotami: — Výborná — Dobrá — Uspokojivá — Slabá — Zlá
Príbuzné parametre	Presnosť, Konzistentnosť, Správnosť.
Voliteľnosť	Odporúčané – aj keď ide o subjektívne hodnotenie zrozumiteľnosti, odporúča sa ho urobiť, pretože dôvody, pre ktoré hodnotiteľ znížil mieru zrozumiteľnosti poskytujú návod na zvýšenie kvality údajov.
Príklad	Do veľkej miery ide o subjektívne hodnotenie, ale môže sa podporiť všímaním si viacerých kritérií, napríklad:

© 2019 Firm name. All rights reserved.

	— Spájanie viacerých atribútov do jedného (napr. adresa). Treba povedať, že adresa v jednom atribúte môže v niektorých prípadoch zvyšovať zrozumiteľnosť, ale určite znižuje použiteľnosť pre štatistické spracovanie — Potreba škálovania parametrov aby boli v podobnom rozsahu ako ostatné — Odstraňovanie nadbytočných medzier z textov
Pseudokód (výpočtový vzorec)	Nepočíta sa na základe údajov v databáze, ale určuje sa subjektívnym hodnotením, ktoré v niektorých prípadoch môže byť aj komplexné (vtedy poskytuje návod na zlepšenie kvality dát).

Tabuľka 24: Ukazovateľ Transformovateľnosť

4.2.2.8 Referenčná integrita

Kompletnosť referenčného identifikátora

Názov ukazovateľa	**Kompletnosť referenčného identifikátora**
Definícia	Miera, do akej je v databáze alebo datasete vyplnený referenčný identifikátor.
Metrika	Pomer záznamov s vyplneným referenčným identifikátorom, ku počtu všetkých záznamov.
Rozsah	Iba pre referenčné identifikátory.
Jednotka metriky	Percento.
Príbuzné parametre	Kompletnosť.
Voliteľnosť	Povinné – záznam bez referenčného identifikátora nie je použiteľný, preto ide o kľúčový ukazovateľ.
Príklad	Ak referenčný register obsahuje 20 000 záznamov, pričom 5 záznamov nemá vyplnený referenčný identifikátor, tak kompletnosť referenčného identifikátora je (20 000 – 5) / 20 000 = 99,975%. Poznamenajme, že aj pri takejto vysokej kompletnosti ide o vážny problém, pretože z pohľadu okolia záznamy bez referenčného identifikátora akoby ani neexistovali.
Pseudokód (výpočtový vzorec)	100 * (COUNT ‘ID‘ WHERE NOT blank) / (COUNT ‘all_records‘).

Tabuľka 25: Ukazovateľ Kompletnosť referenčného identifikátora

Jedinečnosť referenčného identifikátora

Názov ukazovateľa	**Jedinečnosť referenčného identifikátora**
Definícia	Miera duplicít referenčného identifikátora.

© 2019 Firm name. All rights reserved.

Metrika	Pomer identifikátorov ktorých hodnota sa opakuje, ku počtu všetkých záznamov.
Rozsah	Iba pre referenčné identifikátory.
Jednotka metriky	Percento.
Príbuzné parametre	Unikátnosť.
Voliteľnosť	Povinné – jedinečnosť referenčného identifikátora je jedným z kľúčových ukazovateľov.
Príklad	Predpokladajme, že v databáze majú dva záznamy rovnaký identifikátor (napr. A) a iné tri záznamy tiež rovnaký identifikátor (napr. B), takže celkovo je 5 nejednoznačných záznamov. Ak databáza obsahuje 10 000 záznamov, jedinečnosť referenčného identifikátora je 9 995 / 10 000 = 0.9995, teda 99.95%.
Pseudokód (výpočtový vzorec)	100 * ((COUNT(all_values)-COUNT(duplicate_values))/ COUNT(all_values)).

Tabuľka 26: Ukazovateľ Jedinečnosť referenčného identifikátora

Použiteľnosť referencovateľného identifikátora

Názov ukazovateľa	**Použiteľnosť referencovateľného identifikátora**
Definícia	Miera ako je identifikátor použiteľný.
Metrika	Hodnotenie na definovanej stupnici.
Rozsah	Hodnotíme konkrétny identifikátor.
Jednotka metriky	Stupnica s definovanými hodnotami: — Výborná — Dobrá — Uspokojivá — Slabá — Zlá
Príbuzné parametre	Strojová spracovateľnosť.
Voliteľnosť	Povinné – aj keď ide o subjektívne hodnotenie použiteľnosti, ak je identifikátor nepoužiteľný, stráca význam.
Príklad	Do veľkej miery ide o subjektívne hodnotenie, ale môžeme ho podporiť všímaním si viacerých kritérií, napríklad: — Či identifikátor obsahuje informáciu, ktorá sa nemá zverejňovať (napr, rodné číslo) — Či je identifikátor možné použiť aj mimo systémov verejnej správy

© 2019 Firm name. All rights reserved.

	— Či je identifikátor možné použiť aj pri integrácii so zahraničnými systémami — Či sa identifikátor môže zverejňovať — Či je identifikátor v súlade s GDPR
Pseudokód (výpočtový vzorec)	Nepočíta sa na základe údajov v databáze, ale určuje sa subjektívnym hodnotením.

Tabuľka 27: Ukazovateľ Použiteľnosť referencovateľného identifikátora

Stabilita referencovateľného identifikátora

Názov ukazovateľa	**Stabilita referencovateľného identifikátora**
Definícia	Miera duplicít referenčného identifikátora.
Metrika	Hodnotenie na definovanej stupnici.
Rozsah	Iba pre referenčné identifikátory.
Jednotka metriky	Stupnica s definovanými hodnotami: — Stabilný — Nestabilný
Príbuzné parametre	Unikátnosť.
Voliteľnosť	Povinné – ak nie je zaručená stabilita referencovateľného identifikátora, komplikuje to jeho použitie.
Príklad	— Názov mesta, trojpísmenkový kód krajiny, rodné číslo – všetky tieto identifikátory sú stabilné. Aj keď sa názov mesta môže zmeniť (napr. po 1989 sa menili názvy niektorých miest v Československu), nepredpokladá sa, že sa to udeje. Pri rodnom čísle poznamenajme, že z pohľadu stability ide o dobrý identifikátor, (ale má iné problémy merané inými ukazovateľmi). — Priezvisko je typický príklad ukazovateľa so zriedkavými zmenami. Rozdiel oproti názvu mesta je v tom, že zatiaľ čo pri názve mesta predpokladáme, že sa nezmení, u priezviska vieme, že sa u polovice občanov môže zmeniť. Mení sa teda zriedka, možno zriedkavejšie ako názov niektorých miest, ale dá sa očakávať, že sa zmení. Preto sú identifikátory tohto typu nevhodné. — Adresa sa u väčšiny ľudí počas života mení zriedkavo, ale aj tu sa dá očakávať, že sa meniť bude, preto nemôže byť považovaná za stabilný údaj. — Za nestabilný považujeme teda každý údaj, u ktorého sa predpokladá, že sa zmení, aj keď iba zriedkavo. Údaj, ktorý očakávame, že sa meniť nebude (aj keď sa to môže stať) považujeme za stabilný.
Pseudokód (výpočtový vzorec)	Určiť stabilitu údaju je jednoduchšie zo skúsenosti, analýzou, ale je možné ju podložiť aj výpočtom. Zrátame počet objektov, ktorým sa za dané obdobie zmení tento údaj a vypočítame, aký je ich podiel zo všetkých objektov. Ako príklad použijeme zmenu mena, pričom vychádzať budeme iba z hrubých štatistických údajov. Presné výpočty by sa dali urobiť s dátami RFO. Predpokladajme, že na SR je priemerne ročne 30 000 sobášov a 8 000 rozvodov a predpokladajme, že pri 95% sobášov a 40% rozvodov dôjde k zmene mena, tak ročne to znamená

© 2019 Firm name. All rights reserved.

	približne 28 800 zmien mena. Ak predpokladáme, že Slovensko má 5 000 000 obyvateľov, tak ročne si zmení meno 0,58% obyvateľov. Za desať rokov je to skoro 6%, za 30 rokov viac ako 17%. Vhodné je pri analýze nakresliť krivku podielu zmien podľa dĺžky obdobia, ak má jasne stúpajúci charakter, tak ide o nestabilný údaj.

Tabuľka 28: Ukazovateľ Stabilita referencovateľného identifikátora

Formát referencovateľného identifikátora

Názov ukazovateľa	**Formát referencovateľného identifikátora**
Definícia	Ohodnotenie, či referencovateľný identifikátor má formát URI.
Metrika	áno / nie.
Rozsah	Iba pre referenčné identifikátory.
Jednotka metriky	áno / nie.
Príbuzné parametre	Správnosť.
Voliteľnosť	Povinné.
Príklad	Príklady schválených URI je možné nájsť v MetaIS na adrese https://metais.vicepremier.gov.sk/uri/acceptedlist . Definícia formátu referencovateľného identifikátora je v §46 Výnosu 55/2014 o štandardoch pre informačné systémy verejnej správy.
Pseudokód (výpočtový vzorec)	Nepočíta sa, ide o kvalitatívny ukazovateľ.

Tabuľka 29: Ukazovateľ Formát referencovateľného identifikátora

Interoperabilita referencovateľného identifikátora

Názov ukazovateľa	**Interoperabilita referencovateľného identifikátora**
Definícia	Hodnotenie, či referencovateľný identifikátor bol navrhnutý v súlade s ontológiami alebo best practices.
Metrika	Hodnotenie na definovanej stupnici.
Rozsah	Iba pre referenčné identifikátory.
Jednotka metriky	Stupnica s definovanými hodnotami: — Štandard — Štandard s modifikáciami — Vlastný formát

© 2019 Firm name. All rights reserved.

Príbuzné parametre	Správnosť
Voliteľnosť	Odporúčané – navádza na využívanie štandardov.
Príklad	Pre vytváranie URI existuje postup definovaný v MetaIS, https://metais.vicepremier.gov.sk/howto/URI.URI/URI_HOWTO .
Pseudokód (výpočtový vzorec)	Nepočíta sa, ide o kvalitatívny ukazovateľ.

Tabuľka 30: Ukazovateľ Interoperabilita referencovateľného identifikátora

Využívanie referencovaných objektov

Názov ukazovateľa	**Využívanie referencovaných objektov**
Definícia	Kvalitatívne hodnotenie, či referencujeme objekty voči referenčným registrom.
Metrika	áno / nie / čiastočne.
Rozsah	Iba pre objekty, ktoré je možné referencovať.
Jednotka metriky	áno / nie / čiastočne.
Príbuzné parametre	Presnosť.
Voliteľnosť	Povinné – zákon o eGovernmente vyžaduje využívanie referenčných registrov.
Príklad	Predpokladajme, že budujeme register vecí, pri každej veci evidujeme vlastníka, ktorým môže byť fyzická alebo právnická osoba. — Ak vlastníka vôbec nereferencujeme, tak je hodnota ukazovateľa nie — Ak vlastníka referencujeme, tak je hodnota áno — Ak vlastníka referencujeme na RFO ale nie na RPO, tak je hodnota ukazovateľa čiastočne
Pseudokód (výpočtový vzorec)	Nepočíta sa, ide o kvalitatívny ukazovateľ.

Tabuľka 31: Ukazovateľ Využívanie referencovaných objektov

Podiel referencovaných objektov

Názov ukazovateľa	**Podiel referencovaných objektov**
Definícia	Miera ako sa darí referencovať objekty na referenčný register.
Metrika	Pomer referencovaných záznamov ku počtu všetkých záznamov.

© 2019 Firm name. All rights reserved.

Rozsah	Iba pre objekty, ktoré je možné referencovať.
Jednotka metriky	Percento.
Príbuzné parametre	Kompletnosť, Presnosť.
Voliteľnosť	Povinné – tento ukazovateľ hovorí o kvalite prepojenia údajov vo verejnej správe.
Príklad	Predpokladajme, že v databáze je 10 000 záznamov. Z toho: — 9 200 máme referencovaných na objekty v referenčnom registri. — 20 záznamov sa nepokúšame referencovať, pretože ide o mimoriadne záznamy, ktoré sa nemajú nájsť v referenčnom registri. — 90 záznamov sa nepodarilo referencovať, pretože referenčný register vráti viac ako jeden záznam, to znamená, že záznamy nevieme jednoznačne referencovať. — 430 záznamov sa nepodarilo referencovať, pretože sa nenašli v referenčnom registri. — 260 záznamov sme sa nepokúsili referncovať, pretože nemáme dostatok údajov aby sme objekty stotožnili s referenčným registrom. Hodnota ukazovateľa je 9 200 / 10 000 = 92%. Doplnkovým údajom je, že 260 / 10 000 = 2,6% údajov sme sa nepokúsili stotožniť, (90+430) / 10 000 = 5,4% údajov nevieme stotožniť kvôli možným chybám v referenčnom registri.
Pseudokód (výpočtový vzorec)	100 * (COUNT(refID) / COUNT(all_values)).

Tabuľka 32: Ukazovateľ Podiel referencovaných objektov

© 2019 Firm name. All rights reserved.

4.3 Zoznam požiadaviek na informačný systém resp. Centrálnu informačnú platformu

Vzhľadom na to, že sa stále zvyšuje potreba dátovej kvality, existujú aj nástroje na zabezpečenie dátovej kvality. Aby výpočty parametrov mohli byť automatizované, sú dodané požiadavky resp. 10 požiadaviek v tabuľke (Tabuľka 33).

Poradové číslo	Požiadavka	Popis
1	Riadenie metadát	Systém by mal poskytovať archív na dokumentovanie metadát, vrátane názvov, typov, zdieľaných referenčných údajov s cieľom ich trvalo využívať
2	Profilovanie dát	Analýza dát s ohľadom na nájdenie vzorov, chýbajúcich hodnôt, a ďalších základných štatistických charakteristík poskytujúce prehľad a pomoc pri identifikácii problémov súvisiacich s dátovou kvalitou. Môže byť použité pri zhodnotení dátovej kvality
3	Monitoring	Zavedenie kontrol na zabezpečenie dodržiavania stanovených biznis pravidiel, ktoré definujú kvalitu dát
4	Dashboarding	Informačný panel môže vizualizovať agregovaný stav monitorovaných údajov, generovať upozornenia na oznámenie problémov
5	Verifikácia dát	Umožňuje verifikovať dáta na základe nadefinovaných pravidiel. Prejdú len tie dáta, ktoré spĺňajú požadované verifikačné kritériá
6	Párovanie (matching)	Identifikácia, spájanie alebo zlučovanie zhodných záznamov pre účely deduplikácie redundantných záznamov na základe zadefinovaných kritérií zhody medzi alebo v rámci nich
7	Obohatenie (enrichment)	Zvýšenie hodnoty interných dát pridaním atribútov z externých zdrojov (napríklad o demografické alebo geografické dáta)
8	Čistenie (cleansing)	Modifikácia hodnôt tak, aby spĺňali doménové obmedzenia, obmedzenia integrity alebo iných obchodných pravidiel, ktoré definujú dostatočnú kvalitu dát. Napríklad korekcia neplatných hodnôt
9	Štandardizácia a parsing	Rozklad textových polí na časti podľa štandardov, užívateľsky definovaných pravidlách, vzoroch a hodnotách

© 2019 Firm name. All rights reserved.

Poradové číslo	Požiadavka	Popis
10	Pripojenie k dátovým zdrojom	Dôležité je používať nástroj, ktorý sa dokáže pripojiť k rôznym typom dátových zdrojov s možnosťou flexibilnej transformácie a úpravy dát pre analytické potreby a dočisťovanie (ETL funkcionalita)

Tabuľka 33: 10 požiadaviek pre automatizované počítanie parametrov dátovej kvality

4.3.1 Pilotné overenie metodiky pomocou nástroja TALEND pre automatizáciu výpočtov parametrov

4.3.1.1 Talend

Celá koncepcia a metodika merania monitorovania a vyhodnocovania dátovej kvality je technologicky nezávislá a platformovo agnostická. Dá sa aplikovať pre všetky dostupné štandardné riešenia pre riadenie a meranie dátovej kvality.

Pre konkrétne dáta a meranie ich kvality už ale je potrebné použiť niektoré konkrétne technologické riešenie. Nástroj Talend bol vybraný pre pilotné overenie z 3 hlavných dôvodov :

1) Na niektorých rezortoch verejnej správy sa už teraz používa a sú tu naakumulované veľmi cenné interné skúsenosti a zručnosti.

2) Magic quadrant pre nástroj dátovej kvality 2019 od Gartner Group (viď Obrázok 8) ukazuje výbornú pozíciu tohto nástroja pri meraní a riadení dátovej kvality.

3) Riešenie kombinuje big data, integráciu dát, dátovú kvalitu, cloudovú integráciu, prípravu dát a integráciu aplikácií do jednej integračnej platformy so spoločným vývojom a riadiacim prostredím. Riešenie môže byť nasadené v on-premise alebo v cloud móde.

Riešenie pre dátovú kvalitu umožňuje používateľom profilovať, čistiť, maskovať a pripravovať dáta a zároveň monitorovať dátovú kvalitu v čase bez ohľadu na ich formát alebo veľkosť. Umožňuje opätovné použitie pravidiel dátovej kvality v rôznych integráciách. Pravidlá zahŕňajú de-duplikáciu, validáciu, štandardizáciu a obohatenie na základe strojového učenia. Použitý koncept podporuje vykonávanie úloh dátovej kvality aj priamo v Hadoope.

© 2019 Firm name. All rights reserved.

Obrázok 8: Magic Quadrant pre nástroje dátovej kvality 2019

© 2019 Firm name. All rights reserved.

4.3.1.2 Hardvérové a softvérové požiadavky pre inštaláciu Talend klienta

Kategória	Popis
HW - Procesor	64-bit procesor je vyžadovaný
HW – RAM	Minimálne 3 GB RAM (4 GB RAM odporúčané)
HW – Miesto na disku	Minimálne 3 GB
SW – Operačný systém	Minimálne Microsoft windows 7 alebo Linux Ubuntu 14.04
SW – JAVA JRE	Oracle 8 odporúčaný
SW – JAVA JDK	1.7

Tabuľka 34: Požiadavky pre Centrálnu informačnú platformu – Talend klient

4.3.1.3 Hardvérové požiadavky pre centrálnu informačnú platformu (Talend)

Pracovná stanica / Server	Operačný systém	CPU	RAM	SSD Disk
Client PC	Windows/Linux/Mac	4-jadrový i7 procesor	3 GB, 4 GB odporúčané	3 GB
Talend Administration Center	Windows/Linux	4 jadrá minimum,	2 GB minimum, 8 GB odporúčané	1GB minimum + projekt, 20GB odporúčané
Job Server	Windows/Linux	4 jadrá minimum,	2 GB minimum, 8 GB odporúčané (následne podľa počtu procesov)	2 GB minimum, 50 GB odporúčané (podľa potreby projektu)
Run time (ESB)	Windows/Linux	4 jadrá minimum,	1 GB minimum, 16 GB odporúčané (následne podľa počtu procesov)	2 GB minimum, 20 GB odporúčané (podľa potreby projektu)
Data Prep & Data stewardship Server	Windows/Linux	4 jadrá minimum	1 GB, 2 GB odporúčané	2 GB minimum, 20GB odporúčané

© 2019 Firm name. All rights reserved.

Pracovná stanica / Server	Operačný systém	CPU	RAM	SSD Disk
Centralized log server	Windows/Linux	4 jadrá minimum	8 GB RAM	2 GB minimum, 20GB odporúčané (podľa potreby projektu)
Shared Nexus Server	Windows/Linux	4 jadrá minimum	8 GB RAM	2 GB minimum, 20GB odporúčané (podľa potreby projektu)
Git Server (lepsie v SaaS mode)	Windows/Linux	4 jadrá minimum	8 GB RAM	2 GB minimum, 20GB odporúčané

Tabuľka 35: Hardvérové požiadavky pre centrálnu informačnú platformu - Talend

© 2019 Firm name. All rights reserved.

5 Zoznam

5.1 Zoznam tabuliek

5.2 Zoznam obrázkov

© 2019 Firm name. All rights reserved.

5.3 Prílohy

Nástroj pre výpočet dátovej kvality - zoznam vzorcov ukazovateľov.xlsx

© 2019 Firm name. All rights reserved.