

Výstup č. 2.1.1:

Popis štandardizácie dátových prvkov a entít pre referenčné údaje

Realizačná zmluva o poskytnutí služieb a o dielo č. 445/2022

Projekt:

**Zlepšenie využívania údajov vo verejnej
správe**

ITMS kód projektu:

314011S979

Document review and approval

Revision history

Version	Author	Date	Revision
1.0	Celuchová Bošanská Bárdy Janík	6.7.2023	

This document has been reviewed by

Reviewer	Date reviewed
1	
2	
3	
4	
5	

This document has been approved by

Subject matter experts

Name

Signature

Date reviewed

1

2

3

4

5

Tabuľka 1: Zoznam skratiek

Skratka	Význam
AdES	Zdokonalený elektronické podpisy (Advanced Electronic Signature)
CA	Certifikačná autorita (Certification Authority)
CDEI	Centrum pre dátovú etiku a inováciu (Centre for Data Ethics and Innovation)
CIP	Centrálne integračná platforma
CMÚ	Centrálne model údajov
CRL	Zoznam zrušených certifikátov (Certificate Revocation List)
DCMS	Misterstvo pre digitálnu transformáciu, kultúru, médiá a šport (Department for Digital, Culture, Media and Sport of the United Kingdom)
DSS	Služba digitálneho podpisu (Digital Signature Service)
EDPB	Európsky výbor pre ochranu osobných údajov (European Data Protection Board)
eID	Elektronická identita
eIDAS	Nariadenie Európskej únie č. 910/2014 o elektronickej identifikácii a dôveryhodných službách pre elektronické transakcie na vnútornom európskom trhu.
GDPR	Všeobecné nariadenie o ochrane osobných údajov (General Data Protection Regulation)
HSM	Hardware Security Module
ICO	Úrad komisára pre informácie (Information Commissioner's Office)
IS CSRÚ	Informačný systém centrálnej správy referenčných údajov

Skratka	Význam
IS VS	Informačný systém verejnej správy
JSON	JavaScript Object Notation
JSON-LD	JSON pre linkované údaje (JSON for Linking Data)
JWE	JSON Web Encryption
JWS	JSON Web Signature
JWT	JSON Web Token
mID	Mobilná identita
MIRRI SR	Ministerstvo investícií, regionálneho rozvoja a informatizácie
MOU	Manažmentu osobných údajov
MV SR	Ministerstvo vnútra SR
OCSP	Online Certificate Status Protocol
OVM	Orgán verejnej moci
PET	Technológie na zvýšenie súkromia (Privacy-enhancing Technologies)
PIMS	systemy na správu osobných informácií (Personal Information Management System)
PKI	Infraštruktúra verejných kľúčov (Public Key Infrastructure)
RA	Registračná autorita
RDF	Resource Description Framework

Skratka	Význam
SES	Jednoduchý elektronický podpisy (Simple Electronic Signature)
SvM	Slovensko v mobile
QES	Kvalifikovaný elektronický podpis (Qualified Electronic Signature)
TRUSTS	Trusted Secure Data Sharing Space
URI	Jednotný referencovateľný identifikátor
VC	Overiteľné poverenia (Verifiable Credentials)
W3C	World Wide Web Consortium
XML	Extensible Markup Language
ZKP	Zero-Knowledge Proofs

Obsah

1	Úvod a zhrnutie	1
1.1	Kontext	1
1.2	Metodika realizácie výstupu	1
1.3	Manažérske zhrnutie	2
2	Koncept referenčných údajov	5
2.1	Princíp „jeden-krát a dost“	5
2.2	Základné registre a rámec pre európsku interoperabilitu	7
2.3	Teoretické východiská pre previazanie dátových prvkov	11
2.3.1	Referenčné údaje	11
2.3.2	Prepojené sociálne údaje (Social Linked Data alebo SOLID)	12
2.3.3	Dátová integrácia	13
2.4	Súčasný stav v rámci Slovenskej republiky	16
2.5	Dobrá prax zo zahraničia	22
2.5.1	Prehľad zahraničných riešení	22
2.5.2	Otvorený prístup k registrom a referencovaniu (UK)	23
2.5.3	Základné registre s registrom práv a povinností (CZ)	26
2.5.4	Systémové katalógy (NL)	29
3	Definície a účel referenčných údajov	32
4	Návrh nového modelu	38
4.1	Strategické rozhodnutia	38
4.1.1	Právny rámec pre zdieľanie údajov	38
4.1.2	Mechanizmy pre princíp „jeden-krát a dost“	41
4.1.3	Manažment kmeňových údajov	44
4.2	Prepojenie s manažmentom údajov	51
4.2.1	Zdieľané údaje musia byť kvalitné	51
4.2.2	Je potrebné vedieť, kto a aké údaje spravuje	51
4.2.3	Funkčná platforma pre integráciu údajov	52
5	Odporúčania a usmernenia	54
5.1	Odporúčania pre Dátovú kanceláriu	54

5.1.1	Zosúladiť definíciu pojmov a procesov s technológiami a nástrojmi manažmentu údajov („data governance“)	54
5.1.2	Posilniť význam manažmentu metadát cez Centrálny model údajov	55
5.1.3	Vybudovať nástroje pre správu referenčných údajov ako súčasť celkového manažmentu údajov	57
5.1.4	Zvážiť ako alternatívu decentralizovaný kmeňový záznam pomocou MOU	57
5.1.5	Využívať viac platformu Talend aj pre oblasť správy kmeňových údajov (“MDM”) alebo prejsť inú platformu, viac kompatibilnú s CMÚ	58
5.2	Aktualizované metodické usmernenie	59
5.2.1	Postup vyhlasovania referenčných údajov (procesná legislatívna procesná),	59
6	Mapovanie na Centrálny model údajov	61
6.1	Dátové modely pre referenčné údaje	61
6.2	Prehľad údajov	61

1 Úvod a zhrnutie

1.1 Kontext

Dokument bol pripravený v rámci projektu „Zlepšenie využívania údajov vo verejnej správe“. Tento projekt má ambíciu transformovať fungovanie inštitúcií verejnej správy tak, aby dokázali maximálne efektívne spravovať a zdieľať údaje, využívať údaje pre lepšie rozhodovanie na základe faktov a dôkazov, pre zlepšenie efektivity a adresnosti služieb na základe lepšieho využívania dát.

Projekt Zlepšenie využívania údajov vo verejnej správe realizuje Dátová kancelária verejnej správy ako špeciálna jednotka Ministerstva investícií, regionálneho rozvoja a informatizácie (ďalej aj MIRRI SR). Projekt sa okrem iných oblastí venuje aj systému referenčných registrov a rozširovaniu Centrálnej integračnej platformy.

Cieľom zavedenia referenčných registrov a Centrálnej integračnej platformy je umožniť zdieľanie a využívanie údajov medzi informačnými systémami OVM a ich poskytovanie pre výkon verejnej moci elektronicky. Aby údaje mohli byť vyhlásené za referenčné alebo aby sa dali efektívne zdieľať medzi OVM, musí byť ich dátový model spolu so štruktúrou jasne daný. Musí existovať všeobecné porozumenie o sémantike údajov, o tom, ako vznikajú a ako ich možno použiť pri elektronickom výkone verejnej moci. Práve na takéto porozumenie slúži tento dokument, ktorý štandardizuje kľúčové dátové prvky a entity s ohľadom na Centrálny model údajov.

Dokument zohľadňuje nielen spomenutý Centrálny model údajov, ale aj koncept „Social Linked Data“, implementovaný v Manažmente osobných údajov (MOU). Okrem toho sa venuje:

- definícii dátových prvkov,
- mapovaniu na centrálny model údajov,
- výberu štandardov a konvencií,
- modelovaniu dátových prvkov,
- referenčným údajom vo vzťahu k centrálnemu modelu údajov,
- previazaniu dátových prvkov a ich popisu.

Výstup vznikol ako realizácia aktivity číslo 2 Riadenie vyhlasovania referenčných údajov a prepájania údajov. Zámerom je zvýšiť mieru zdieľania údajov v rámci verejnej správy a smerovať tak k dosiahnutiu potenciálu moderného využívania údajov, ktoré je popísané v dokumente 1.3.1 Konceptia moderného využívania údajov.

1.2 Metodika realizácie výstupu

Dokument sa venuje definíciám základných konceptov, ako sú referenčné údaje, kmeňové údaje a agendové údaje a na základe dobrej praxe aj zo zahraničia navrhujeme, akú úlohu majú zohrávať v modernej dátovej infraštruktúre. Inštitúcie verejnej správy si môžu zdieľať aj údaje, ktoré nie sú referenčné, avšak im možno dôverovať a môžu slúžiť na zefektívnenie výkonu agendy alebo na lepšie rozhodovanie.

Následne sa venujeme možnostiam pre zavedenie konceptu základných registrov a definujeme vybranú množinu referenčných údajov, ktoré by mali byť vyhlásené a zavedené a používané formou stotožnenia.

Jednotlivé témy zdieľania údajov sú podrobnejšie vysvetlené v kontexte slovenského prostredia e-Governmentu. Identifikované sú základné varianty, ako k problematike pristúpiť a výber vhodného riešenia. Stratégiu je možné chápať ako sériu rozhodnutí a výber prístupov pre jednotlivé témy:

- Právny rámec pre zdieľanie údajov.
- Mechanizmy pre zdieľanie údajov a princíp „jeden-krát a dost“.
- Manažment kmeňových údajov.

Dôležitým krokom je prepojiť zdieľanie údajov s ostatnými procesmi v rámci Manažmentu údajov vo verejnej správe (dátová integrácia a dátová kvalita).

V rámci analýzy týchto častí nového konceptu sa vygenerovali základné otázky a kľúčové rozhodnutia, ktoré je potrebné urobiť. Napríklad ako by mala vyzeráť celoštátna správa a riadenie (governance) základných registrov a ktorý orgán by mal byť centrálnym orgánom pre správu a riadenie celej digitálnej agendy, aké základné druhy základných registrov by malo Slovensko používať, akým spôsobom by s nimi sprístupňovať údaje, aby to bolo efektívne, aké nástroje by mali byť použité pre naplnenie princípu „jeden-krát a dost“, aké mechanizmy je potrebné zvoliť pre sémantickú interoperabilitu údajov, aký štýl manažmentu kmeňových údajov by mal byť použitý, alebo aký by mal byť celkový koncept integrácie údajov.

1.3 Manažérske zhrnutie

Zavádzanie princípu „jeden-krát a dost“

Princíp „jeden-krát a dost“ začal fungovať v praxi najmä vďaka zákonu proti byrokracii a systému over.si pre webový prístup k potrebným údajom počas konaní. Postupne sa v troch vlnách zrušila povinnosť dokladať pre úrady v konaniach pre výpisy z 20 informačných systémov verejnej správy (ide o kľúčové systémy, ako matriky, kataster, obchodný register, register trestov, informačné systémy Finančnej správy, Register fyzických osôb a podobne). Zákon proti byrokracii rieši aj jeden z hlavných problémov konceptu referenčných údajov a stotožňovania ako sú vnímané podľa zákona o e-Governmente, a to nesúlad medzi všeobecným zákonom a špeciálnymi zákonmi. Zákon proti byrokracii novelizoval všetky predpisy, ktoré predkladanie výpisov explicitne vyžadovali.

Zákon proti byrokracii zaviedol jeden-krát a dost efektívnym spôsobom (najmä z pohľadu rýchlosti), využil pri tom fungujúcu dátovú infraštruktúru (CSRÚ) a možnosť využitia webového prístupu cez over.si. Daňou za rýchlosť je nekonceptné riešenie z pohľadu dátovej interoperability a automatizovaného spracovania údajov. Údaje sa vymieňajú najmä ako elektronické verzie papierových dokumentov (výpisy) a pracuje sa s nimi rovnako ako s papierovými dokumentami. Keďže úradníci môžu pristupovať k širokej množine výpisov o občanovi, ochrana osobných údajov je problematická (tu pomôže služba notifikácie o prístupoch k údajom, ktorá sa zavádza v rámci MOU).

Úspech zákona proti byrokracii pri zavádzaní princípu jeden krát do praxe zároveň zastavil snahy používať referenčné údaje a stotožňovanie. Tento pôvodný prístup bol z pohľadu interoperability a možnosti automatizácie koncepcnejší, zároveň si vyžadoval vysoké bariéry: nutnosť meniť informačné prostredie na strane konzumentov, náročné vyhlasovanie referenčných údajov, obavy inštitúcií vyhlásiť referenčné údaje z dôvodu ich nízkej kvality. Nový koncept pre zdieľanie údajov musí stanoviť formu údajov pri ich zdieľaní a výmene a podporiť automatizáciu procesov.

Návrh nového konceptu pre zdieľanie údajov

Údaje ako verejné aktíva by mali byť prístupné štyrmi spôsobmi:

- ostatným verejným inštitúciám cez Centrálnu integračnú platformu na základe zákonného oprávnenia (ktoré sa bude definovať v Metainformačnom systéme),
- oprávneným osobám, ktorých sa údaje týkajú cez Manažment osobných údajov,
- verejnosti v otvorenom formáte,
- subjektom, ktoré získajú prístup k údajom na základe aktu o správe údajov.

Z pohľadu formy je pre všetky spôsoby sprístupnenia cieľovým stavom použitie prepojených údajov na základe ontológií z Centrálného modelu údajov (RDF v JSON-LD). Takéto údaje bude možné strojovo spracovať a zároveň sémanticky interpretovať ich význam.

Pre naplnenie princípu „jeden-krát a dost“ navrhujeme použiť nasledovné nástroje:

- Stotožňovanie referenčných údajov. Centrálna integračná platforma by mala poskytovať nástroje pre stotožnenie referenčných údajov so svojou evidenciou a upozorňovať na zmeny referenčných údajov.
- Zdieľanie agendových údajov a kmeňových údajov (ktoré nie sú referenčné) získaných cez Centrálnu integračnú platformu v konsolidovanej podobe. Ak sa takéto údaje použijú ako podklad pre ďalšie rozhodnutie, je dôležité, aby bola zaznamenaná informácia o hodnote v konkrétnom čase. Konsolidovaný dátový objekt spája údaje z viacerých objektov evidencie (ktoré sú vedené vo viacerých registroch) do jednotného dátového objektu (z definovanou štruktúrou) pre vybraný prípad použitia. Použitie konsolidovaných dátových objektov je vhodné pri podpore princípu jeden-krát a dost v správnom konaní respektíve pri riešení životných situácií.
- Zasielanie údajov na základe súhlasu, teda na základe vyjadrenia vôle subjektu respektíve osoby, o ktorej sú údaje spracúvané prostredníctvom MOU.

Navrhujeme v Slovenskej republike zriadiť nasledovné druhy základných registrov:

- Register fyzických osôb.
- Register právnických osôb.
- Register adries.
- Register priestorových informácií.

- Register práv a povinností (ako súčasť Meta-informačného systému verejnej správy).

Z pohľadu manažmentu kmeňových údajov navrhujeme implementovať konsolidačný štýl pre manažment osobných údajov a koexistenčný štýl pre správu kmeňových údajov a agendových údajov.

Odporúčania pre Dátovú kanceláriu verejnej správy

Aby bolo možné zaviesť nový koncept pre zdieľanie údajov do praxe, je potrebné v rámci Dátovej kancelárie realizovať sústredené aktivity:

- Zosúladiť definíciu pojmov a procesov s technológiami a nástrojmi manažmentu údajov („data governance“) pomocou predstaveného návodu.
- Posilniť praktický význam manažmentu metadát cez Centrálny model údajov.
- Vybudovať nástroje pre správu referenčných údajov ako súčasť celkového manažmentu údajov.
- Zvážiť ako alternatívu decentralizovaný kmeňový záznam pomocou Manažmentu osobných údajov (MOU).
- Využívať viac platformu Talend aj pre oblasť správy kmeňových údajov (“MDM“) alebo prejsť inú platformu, viac kompatibilnú s Centrálnym modelom údajov (CMÚ).

2 Koncept referenčných údajov

2.1 Princíp „jeden-krát a dost“

„Jeden-krát a dost“ a digitálna transformácia

Odomknutie sily digitálnej transformácie: Zjednodušenie verejných služieb pomocou princípu jeden-krát a dost

V dnešnom rýchlo sa meniacom svete sa digitálna transformácia stala kľúčovým faktorom pri modernizácii verejného sektora. V srdci tejto transformácie sa nachádza revolučný koncept známy ako princíp jeden-krát a dost (OOP).

Princíp jeden-krát a dost eliminuje potrebu, aby občania a podniky opakovane poskytovali rovnaké údaje vláde. Namiesto toho, ak sú požadované informácie už k dispozícii, môžu byť bezpečne znovu použité. Tým sa znižuje administratívne zaťaženie a zjednodušuje poskytovanie verejných služieb, čo vedie k úsporám nákladov a zlepšeniu efektívnosti.

Vlády si stanovujú za cieľ prostredníctvom tohto princípu urobiť svoje služby efektívnejšími, jednoduchšími a zameranými na občanov. Predstavte si svet, kde občania a podniky musia poskytnúť konkrétne (štandardné) informácie verejným orgánom len „jeden-krát a dost“. Tieto orgány sú následne povinné interné opätovné použitie týchto údajov, pričom dodržiavajú pravidlá ochrany údajov, aby sa občanom a podnikom nevznášalo žiadne ďalšie zaťaženie.

Avšak na úspech tejto transformačnej vízie je nevyhnutná široká akceptácia a spolupráca všetkých zainteresovaných strán. Zdieľanie a opätovné použitie citlivých a necitlivých údajov občanov a podnikov si vyžaduje dôveru. Tento revolučný koncept vyžaduje pripravenosť verejných inštitúcií prispôbiť sa a vyvinúť sa, aby zabezpečili jeho úspech.

Digitalizácia a digitálna transformácia boli kľúčovými faktormi v stratégiách modernizácie vlád vo svete. V minulosti sa to označovalo ako elektronická vláda alebo digitálna vláda, ale teraz sa vyvinula do komplexného konceptu digitálnej transformácie. Tento prístup využíva inovatívne informačné a komunikačné technológie (IKT) na zlepšenie poskytovania verejných služieb, čo vedie k zvýšenej efektívnosti, účinnosti, transparentnosti a kvalite pre občanov a podniky. Zároveň si klade za cieľ dosahovať verejnú hodnotu, reakčnosť vlády a otvorenosť.

Prístup zameraný na zákazníka leží v srdci tejto transformácie a má za cieľ vytvárať hodnoty pre spoločnosť a ekonomiku. Keďže princíp jeden-krát a dost vyžaduje zdieľanie a opätovné použitie citlivých a osobných údajov, je nevyhnutné, aby verejné orgány dodržiavali prísne pravidlá ochrany údajov. Dôvera je kľúčovým prvkom pri implementácii princípu jeden-krát a dost. Navyše, digitálna transformácia nie je jednoduchý proces a digitálne nástroje samé osebe neprispievajú k uvedeným úspešným faktorom. Keďže digitálna transformácia vo vláde je považovaná za veľmi komplexný úkon, koordinačné a zapájacie koncepty, ako aj celkové architektúry, ktoré umožňujú bezpečný a dôveryhodný prístup k údajom a informáciám pri zdieľaní informácií medzi úradmi, sú ďalšie úspešné kritériá.

Zhrnutím, digitálna transformácia, ktorú poháňa princíp jeden-krát a dosť, mení verejný sektor tým, že zjednodušuje služby, znižuje administratívne zaťaženie a buduje dôveru. Táto transformačná cesta vyžaduje spoluprácu, inovácie a hlboké záväzky vytvoríť prepojený a občanmi zameraný vládny zážitok.

Východiská a predpoklady

Zavedenie princípu jeden-krát a dosť vyžaduje určité východiská a predpoklady:

- Digitalizácia a modernizácia: Je potrebné, aby vlády mali vytvorené vhodné digitálne infraštruktúry a technologické riešenia, ktoré umožnia efektívne zhromažďovanie, spracovanie a zdieľanie údajov medzi rôznymi verejnými orgánmi.
- Zákonný rámec a regulácie: Je nevyhnutné mať jasne stanovené právne predpisy a pravidlá, ktoré zabezpečujú ochranu osobných údajov a zároveň umožňujú zdieľanie a opätovné použitie údajov v súlade s platnými predpismi.
- Dôvera a súhlas občanov: Zavádzanie princípu jeden-krát a dosť vyžaduje dôveru občanov a ich súhlas s poskytovaním a zdieľaním ich údajov. Musí existovať primeraný rámec pre informovaný súhlas a transparentnosť v súvislosti s používaním a ochranou údajov.
- Integrované informačné systémy: Je dôležité, aby rôzne verejné orgány mali schopnosť vzájomne sa prepojiť a zdieľať údaje prostredníctvom integrovaných informačných systémov. Toto umožňuje efektívne využitie údajov a minimalizuje potrebu opakovaného získavania rovnakých informácií od občanov a podnikov.
- Vzájomná spolupráca a koordinácia: Úspešné zavedenie princípu jeden-krát a dosť si vyžaduje spoluprácu a koordináciu medzi rôznymi verejnými orgánmi a inštitúciami. Je potrebné vytvoriť partnerstvá a vzájomne sa podporovať pri zdieľaní a využívaní údajov.
- Bezpečnosť a ochrana údajov: Je nevyhnutné zabezpečiť primeranú úroveň bezpečnosti a ochrany údajov pri zdieľaní a opätovnom použití. Musia sa dodržiavať prísne bezpečnostné opatrenia a technické štandardy na minimalizáciu rizika úniku údajov a porušenia súkromia.

Prístupy k zavádzaniu jeden-krát a dosť

Pri zavádzaní princípu jeden-krát a dosť existujú rôzne prístupy a stratégie, ktoré môžu byť použité:

- Centralizovaný prístup: V tomto prípade je zodpovednosť za zhromažďovanie a spravovanie údajov zverená jednej centrálnaj inštitúcii alebo orgánu. Táto inštitúcia má za úlohu zabezpečiť, že ostatné verejné orgány majú prístup k týmto údajom a môžu ich znovu použiť. Centralizovaný prístup umožňuje väčšiu kontrolu nad údajmi a lepšiu koordináciu medzi rôznymi orgánmi.
- Decentralizovaný prístup: V tomto prípade sa údaje nachádzajú a spravujú na viacerých miestach a úrovniach verejného sektora. Jednotlivé orgány majú svoje vlastné systémy a zodpovedajú za zhromažďovanie a spracovanie údajov vo svojej jurisdikcii. Tieto orgány potom medzi sebou zdieľajú údaje podľa potreby.
- Hybridný prístup: Tento prístup kombinuje prvky centralizácie a decentralizácie. Niektoré údaje sú zhromažďované a spravované centrálné, zatiaľ čo iné sú

© yyyy Copyright owned by one or more of the KPMG International entities. KPMG International entities provide no services to clients. All rights reserved.

ponechané na jednotlivé orgány. Ide o flexibilný model, ktorý umožňuje prispôsobenie sa potrebám konkrétnych situácií a oblastí.

- Systém jednej brány (Single Gateway System): Tento prístup vytvára jedno miesto, cez ktoré občania a podniky môžu poskytnúť svoje údaje. Tieto údaje sú potom zdieľané a využívané rôznymi verejnými orgánmi. Systém jednej brány zjednodušuje proces poskytovania údajov a minimalizuje opakované žiadosti o rovnaké informácie.
- Prístup založený na interoperabilite: Tento prístup sa zameriava na zabezpečenie interoperability medzi rôznymi informačnými systémami a úradmi. Údaje sú štandardizované a zdieľané medzi orgánmi prostredníctvom definovaných rozhraní a štruktúr. Tento prístup umožňuje efektívne a bezpečné zdieľanie údajov.

2.2 Základné registre a rámec pre európsku interoperabilitu

Európska Komisia v roku 2017 prijala Európsky rámec pre interoperabilitu (European Interoperability Framework – EIF), ktorý predstavuje zhodu na interoperabilnom prístupe k európskym verejným službám.

Pre účely tohto rámca sa za interoperabilitu považuje schopnosť organizácií spolupracovať v záujme dosiahnutia spoločných cieľov, vrátane zdieľania informácií alebo znalostí medzi týmito organizáciami, a to prostredníctvom procesov podporujúcich výmenu údajov medzi ich informačnými systémami. Európskymi verejnými službami sú myslené akékoľvek služby verejného sektora, ktoré majú cezhraničný charakter a sú poskytované verejnou správou, a to buď v rámci verejnej správy, ale ide aj o služby poskytované občanom alebo podnikateľom v rámci EÚ.

Účelom EIF je:

- inšpirovať európske orgány verejnej správy v ich úsilí navrhnuť a poskytovať plynulé európske verejné služby iným orgánom verejnej správy, občanom a podnikom
- poskytnúť usmernenia pre orgány verejnej správy pri navrhovaní a aktualizáciách vnútroštátnych rámcov interoperability alebo vnútroštátnych politík, stratégií na podporu interoperability
- prispieť k vytvoreniu jednotného digitálneho trhu podporou cezhraničnej a medzisektorovej interoperability pri poskytovaní európskych verejných služieb.

EIF je založený na princípoch subsidiarity, proporcionality, transparentnosti, opätovnej použiteľnosti, technologickej neutralite, prenosnosti údajov, orientácie na používateľa, začlenenia, dostupnosti, bezpečnosti, súkromia, viacjazyčnosti, administratívneho zjednodušovania, uchovávanía informácií a hodnotenia efektívnosti a účinnosti. Všetky tieto princípy majú svoje opodstatnenie a na základe doterajšieho vývoja a skúseností je namieste, aby boli aplikované také riešenia, ktoré reflektujú alebo zodpovedajú týmto princípom a EIF obsahuje súbor konkrétnych odporúčaní, ktoré majú prispieť k naplneniu týchto princíпов.

Okrem toho EIF stanovuje štyri oblasti, na ktoré je potrebné sa sústrediť pri vytváraní modelu pre **riadenie a dosahovanie interoperability**:

- **Právna interoperabilita** – v európskom kontexte ide najmä o zabezpečenie interoperability medzi organizáciami, ktoré fungujú v rámci rôznych resp. odlišných právnych rámcov alebo právnych poriadkov, na vnútroštátnej úrovni ide viac o zjednotenie alebo zosúladenie rôznych právnych režimov regulujúcich právne vzťahy v rôznych spoločenských, podnikateľských ale aj verejnoprávných oblastiach, pričom EIF odporúča vykonávať napríklad kontrolu právnej interoperability skríningom existujúcej legislatívy tak, aby sa zistili bariéry interoperability v obmedzeniach týkajúcich sa ukladania dát, odlišných licenčných modelov, príliš reštriktívnych pravidiel používania digitálnych technológií, protichodných požiadaviek spojených s rovnakými alebo podobnými procesmi alebo zastaraných bezpečnostných potrieb.
- **Organizačná interoperabilita** – sa týka najmä zladenia procesov, kompetencií a očakávaní verejnej správy v prospech dosahovania dohodnutých a všeobecne výhodných cieľov, v praxi organizačná interoperabilita znamená integráciu alebo zosúladenie procesov a informácií alebo údajov, ktoré majú byť zdieľané. Organizačná interoperabilita je zameraná na splnenie požiadaviek používateľov týkajúcich sa prístupnosti takých služieb, ktoré budú ľahko identifikovateľné, dostupné a zamerané na používateľa. Organizačná interoperabilita podľa EIF vyžaduje, aby boli vzťahy medzi poskytovateľmi medzi poskytovateľmi služieb navzájom, ako aj vzťahy s používateľmi jasne definované, čo zahŕňa nájdanie správnych nástrojov pre vzájomnú asistenciu, spoločné aktivity a prepojené procesy.
- **Sémantická interoperabilita** – znamená, že presná forma a význam zdieľaných údajov a informácií je zachovaný a zrozumiteľný počas celej doby ich zdieľania, aby sa predišlo nedorozumeniam, nesprávnym interpretáciám a všeobecne chybám, ktoré by zmazali účel zdieľania konkrétnych údajov a informácií. Sémantický aspekt sa vzťahuje na význam údajov a informácií, ako aj na vzťahy medzi nimi a tento aspekt vyžaduje vytvorenie slovníkov a schém, ktoré budú exaktne popisovať zdieľanie alebo výmenu údajov a zabezpečia, že tieto údaje budú všetkými zúčastnenými stranami pochopené správne a jednotne. Syntaktický aspekt vyžaduje presnú formu údajov pokiaľ ide o ich zloženie, štruktúru, ale aj gramatiku. EIF odporúča napríklad prístup „data-driven design“ spoločne s technológiami využívajúcimi „linked data“, teda dizajnové princípy pre zdieľanie strojovo čitateľných prepojených dát nachádzajúcich sa na webe.
- **Technická interoperabilita** – pokrýva aplikácie, ako aj systémy prepájajúce infraštruktúru a služby a venuje sa špecifikácii rozhraní, službám prepojovania, službám integrácie dát, prezentácii dát a ich výmene, ako aj protokolom pre bezpečnú komunikáciu. Pre technickú interoperabilitu je najväčšou výzvou fragmentácia informačných systémov, ktoré je následne ťažké prepojiť tak, aby efektívne spolupracovali. EIF odporúča pre zabezpečenie technickej interoperability používať pri každej príležitosti formálne technické špecifikácie, ktoré budú jednotné a zrozumiteľné a najmä záväzné pre všetky zúčastnené strany.

Napriec týmito oblasťami by malo navyše fungovať **integrované riadenie verejných služieb**, ktoré vychádza z modelu riadenia interoperability a v praxi by malo toto integrované riadenie služieb fungovať v medziach pravidiel stanovených týmto modelom interoperabilného riadenia. Interoperabilné riadenie je výsledkom rozhodnutí o interoperabilných rámcoch, inštitucionálneho usporiadania, organizačných štruktúr, rolí a kompetencií, politík, dohôd a iných aspektov zabezpečujúcich a monitorujúcich interoperabilitu na vnútroštátnej aj európskej úrovni.

© yyyy Copyright owned by one or more of the KPMG International entities. KPMG International entities provide no services to clients. All rights reserved.

Integrované riadenie verejných služieb by malo zahŕňať minimálne:

- Jasné definovanie organizačných štruktúr, úloh, zodpovedností a rozhodovacích procesov
- Požiadavky na aspekty interoperability vrátane kvality, škálovateľnosti a dostupnosti opakovane použiteľných stavebných blokov vrátane informačných zdrojov (základné registre, portály s otvorenými údajmi atď.) a iných vzájomne prepojených služieb
- Požiadavky na externé informácie/služby preložené do jasných dohôd o úrovni služieb (vrátane interoperability)
- Plán pre riadenie zmien, ktorý definuje postupy a procesy potrebné na riešenie a kontrolu zmien
- Plán kontinuity podnikania/obnovy po havárii, aby sa zabezpečilo, že digitálne verejné služby a ich stavebné bloky budú naďalej fungovať v rôznych situáciách, napr. kybernetické útoky alebo zlyhanie stavebných blokov.

Podľa EIF by mal optimálny model sledovať nasledovné ciele:

- Odstránenie zložitosti pre koncového používateľa – prostredníctvom integrovaného poskytovania služieb založeného na koordinácii, ktorá zabezpečí, že potreby sú identifikované a k nim sú vytvorené alebo priradené náležité služby.
- Zabezpečenie dostupnosti digitálnych kanálov – pomocou politiky poskytovania služieb ktorá poskytuje alternatívne možnosti a kanály na poskytovanie služieb a v praxi znamená, že je možné sa dostať k službe viacerými spôsobmi.
- Zníženie nákladov a zvýšenie kvality a interoperability služieb – pomocou opätovného použitia údajov a služieb.
- Zvýšenie vyhľadateľnosti a využiteľnosti údajov – prostredníctvom katalógov popisujúcich opätovne použiteľné služby. Existujú rôzne typy katalógov, napr. adresáre služieb, knižnice softvérových komponentov, portály s otvorenými údajmi, registre základných registrov, katalógy metadát, katalógy noriem, špecifikácií a smerníc. Na umožnenie interoperability medzi katalógmi sú potrebné spoločne dohodnuté opisy služieb, údajov, registrov a interoperabilných riešení publikovaných v katalógoch.
- Zabezpečenie integrovaného riadenia verejných služieb – prostredníctvom zladovania a integrácie služieb a informačných systémov.
- Zaistenie bezpečnosti a súkromia – prostredníctvom prístupu bezpečnosť založená na dizajne a súkromie založené na dizajne, ako aj zabezpečením dodržiavania všeobecne záväzných pravidiel pre ochranu údajov a súkromia, plánov pre manažment rizík, plánov pre zachovanie kontinuity činnosti, záložných plánov a plánov pre obnovenie činnosti, plánov pre autorizáciu prístupu k dátam, ako aj prostredníctvom používania kvalifikovaných a dôveryhodných služieb riadiacich sa reguláciou eIDAS. Výmena alebo zdieľanie musí byť overené, šifrované a musí využívať časové známky a logovanie.

EIF odporúča vyvinúť **zdieľanú infraštruktúru zdrojov** pre opätovné zdieľanie služieb a informácií.

Základné registre (Base Registries) považuje EIF za základný kameň pre budovanie interoperability. Základné registre majú byť dôveryhodným a autoritatívnym zdrojom informácií, ktoré môžu byť opätovne používané. „Autoritatívny“ v tomto prípade znamená, že základný register sa považuje za zdroj informácií, ktorý spoľahlivo vykazuje správny stav, je aktuálny a má najvyššiu možnú kvalitu a integritu.

V rámci doterajších skúseností by mala byť vždy stanovená jedna (resp. jediná) organizácia zodpovedná za zhromažďovanie, používanie, aktualizáciu a uchovávanie informácií. V prípade centralizovaných registrov by mala byť za kvalitu údajov a za zavedenie opatrení na zabezpečenie správnosti údajov zodpovedná jedna organizačná jednotka. Takéto registre musia byť pod právnou kontrolou verejnej správy, pričom prevádzka a údržba môžu byť zverené aj iným organizáciám. Pre efektívnu správu základných registrov je kľúčové mať dokonalý prehľad o prevádzke základných registrov a o údajoch, ktoré uchovávajú (register registrov).

Pre základné registre poskytuje EIF nasledovné odporúčania:

- Pri implementácii prístupových a kontrolných mechanizmov na zaistenie bezpečnosti a súkromia v súlade s príslušnou legislatívou sprístupniť autoritatívne zdroje informácií.
- Vyvinúť rozhrania so základnými registrami a autoritatívnymi zdrojmi informácií, publikovať sémantické a technické prostriedky a dokumentáciu potrebnú pre pripojenie a opätovné použitie dostupných informácií.
- Priradiť ku každému základnému registru vhodné metadáta vrátane popisu jeho obsahu, zabezpečenia služieb a zodpovedností, typu kmeňových dát, ktoré uchováva, podmienok prístupu a príslušných licencií, terminológie, slovníka a informácií o všetkých kmeňových dátach, ktoré používa z iných základných registrov.
- Vytvárať a dodržiavať plány zabezpečenia kvality údajov pre základné registre a súvisiace kmeňové údaje.

EIF tiež rieši **princíp otvorených dát**, pri ktorom sa odporúča sústrediť na uvoľnenie strojovo čitateľných údajov s cieľom stimulovať transparentnosť, spravodlivú hospodársku súťaž, inovácie a ekonomiku založenú na údajoch. Na zabezpečenie rovnakých podmienok musí byť otváranie údajov a ich opätovné použitie nediskriminačné, čo znamená, že údaje musia byť interoperabilné, aby sa dali nájsť a spracovať.

EIF ohľadom otvorených dát odporúča nasledovné kroky:

- Vytvoriť postupy a procesy na integráciu otvárania údajov do bežných procesov, pracovných rutín a do vývoja nových informačných systémov.
- Zverejňovať otvorené dáta v strojovo čitateľných, nechránených formátoch. Zabezpečiť, aby boli otvorené údaje sprevádzané vysokokvalitnými, strojovo čitateľnými metaúdajmi v nechránených formátoch vrátane popisu ich obsahu, spôsobu zhromažďovania údajov a úrovne ich kvality a licenčných podmienok, za ktorých sú sprístupnené. Na vyjadrenie metadát sa odporúča používať bežné slovníky.

- Jasne komunikovať o práve na prístup a opätovné použitie otvorených údajov. Právne režimy na uľahčenie prístupu a opätovného použitia, ako sú licencie, by sa mali čo najviac štandardizovať.

2.3 Teoretické východiská pre previazanie dátových prvkov

2.3.1 Referenčné údaje

Referenčné údaje sú podmnožinou údajov pozostávajúcou z kódov alebo iných symbolov, ktoré sa používajú na reprezentáciu dôležitých informácií. Referenčné údaje sa zvyčajne používajú na klasifikáciu a popis rôznych prvkov pre aplikácie, formuláre a množiny údajov. Často sa používajú tiež na pomoc pri štandardizácii údajov a na zabezpečenie toho, aby sa všetky odkazy alebo časti údajov používali a interpretovali presne a konzistentne.

Referenčné údaje sa tiež označujú ako kmeňové údaje a vymenované údaje, pretože obsahujú hodnoty, ktoré pomáhajú definovať konkrétne objekty. Príkladom toho je definícia produktu z katalógu produktov. Referenčné údaje môžu zahŕňať položky, ako sú čísla produktov, popisy, kódy krajín, kódy mien a ďalšie. Referenčné údaje nemusia byť nevyhnutne statické; môže sa časom aktualizovať, keď sa pridajú nové informácie alebo sa zmenia staré informácie.

Referenčné údaje sa často ukladajú do centralizovaného úložiska údajov, aby sa umožnilo zdieľanie údajov medzi aplikáciami. Pretože referenčné údaje sú zdieľané, všetci používatelia rovnako chápu kódy a symboly používané v rôznych súboroch údajov. Pomáha to zabezpečiť, aby boli údaje konzistentné a presné, čím sa zabráni nejasnostiam, ktoré môžu nastať, keď sa referenčné údaje nepoužívajú správne.

Porovnanie referenčných a kmeňových údajov

Referenčné údaje sú definované ako prvky údajov, ktoré sa používajú na definovanie charakteristík organizácie alebo jej prostredia. Tento typ údajov má zvyčajne statický charakter, čo znamená, že sa často nemenia. Príklady referenčných údajov zahŕňajú identifikátory zákazníkov, kódy produktov, geografické identifikátory a kategórie produktov.

Kmeňové dáta sú na druhej strane dáta, ktoré sú spojené s podnikateľskými subjektmi. Tieto údaje sa často používajú na riadenie operácií v rámci organizácie. Kmeňové údaje sa zvyčajne považujú za jediný zdroj pravdy, čo znamená, že ide o definitívny zdroj údajov pre obchodné procesy. Príklady kmeňových údajov zahŕňajú záznamy zákazníkov, záznamy dodávateľov, finančné údaje a záznamy zamestnancov.

Stručne povedané, kľúčový rozdiel medzi referenčnými údajmi a kmeňovými údajmi je ten, že referenčné údaje sú statické a používajú sa na definovanie prvkov organizácie, zatiaľ čo kmeňové údaje sú dynamické a používajú sa na riadenie operácií v rámci organizácie. Medzi príklady referenčných údajov radíme napríklad základné číselníky:

- Kódy produktov, ako sú SKU alebo UPC,
- Kódy krajín a mien,
- geografické údaje ako zemepisná šírka a dĺžka,

- Kódy pohlaví,
- Spôsob platieb,
- Telefónne a faxové čísla,
- Národné identifikačné čísla.

2.3.2 **Prepojené sociálne údaje (Social Linked Data alebo SOLID)**

Solid (Social Linked Data) je webový decentralizačný projekt. Projekt viedol Sir Tim Berners-Lee, vynálezca World Wide Web a pôvodne bol vyvinutý v spolupráci na Massachusetts Institute of Technology (MIT). Cieľom projektu je „radikálne zmeniť spôsob, akým dnes webové aplikácie fungujú, výsledkom čoho je skutočné vlastníctvo údajov, ako aj lepšie súkromie pre používateľ. Výsledkom projektu je návrh platformy pre aplikácie s prepojenými údajmi, ktoré sú úplne decentralizované a plne pod kontrolou používateľov. Základným cieľom Solid je umožniť používateľom mať plnú kontrolu nad svojimi vlastnými údajmi, vrátane riadenia prístupu a miesta uloženia.

Ústredným zameraním konceptu Solid je umožniť vyhľadávanie a zdieľanie informácií spôsobom, ktorý zachováva súkromie. Používateľ ukladá osobné údaje do „podov“ (osobné online úložiská údajov), ktoré sú hostované kdekoľvek si to používateľ želá. Aplikáciám je následne povolené požadovať údaje, ak používateľ dal aplikácii povolenie. Používateľ sa potom môže pripojiť k overenej aplikácii tým, že jej udelí povolenie na prístup k príslušným informáciám v konkrétnom podu. Používateľ si ponecháva úplné vlastníctvo a kontrolu nad údajmi v používateľských moduloch aplikácie: aké údaje obsahuje každý modul, kde je každý modul uložený a ktoré aplikácie majú povolenie na používanie údajov.

Solid pozostáva z nasledujúcich komponentov:

- Organizovaná zbierka štandardov a dátových formátov/, ako je identita, overovanie, prihlasovanie, zoznamy povolení, správa kontaktov, posielanie správ, odbery služieb a podobne (zámerom je previazať z CMÚ).
- Špecifikácie a poznámky k návrhu, ktoré popisujú REST API na rozšírenie existujúcich štandardov, na usmernenie vývojárov pri vytváraní serverov alebo aplikácií.
- Servery, ktoré implementujú špecifikáciu Solid.
- Testovacia sada na testovanie a overovanie implementácií Solid.
- Ekosystém aplikácií, poskytovateľov identity a pomocných knižníc, ktoré bežia na platforme Solid.
- Komunita poskytujúca dokumentáciu, diskusie, návody a prezentácie.

Osobné úložisko (POD)

Osobné úložisko alebo pod je osobný online priestor pre úložisko dát hostený na Solid kompatibilnom servri. Je to miesto, kde používateľ môže ukladať dáta bezpečne a súkromne, s využitím protokolov a štandardov platformy Solid. Solid Pod je obvykle spojený s webID používateľa, čo mu poskytuje osobnú online identitu.

Solid server

Solid server je server, ktorý beží na platforme Solid a poskytuje služby, ako je spravovanie úložísk, autentifikácia, autorizácia, publikovanie a iné. Solid server poskytuje podporu pre aplikácie a webové stránky, ktoré majú prístup k dátam v úložisku používateľa. Solid server môže byť súkromný alebo verejný, čo znamená, že môže hostiť úložiská pre verejnosť alebo pre konkrétneho používateľa.

WebID

WebID je unikátny identifikátor používateľa webu, ktorý je používaný na identifikáciu používateľa. WebID je tiež známy ako identita URI (Uniform Resource Identifier) a je spravovaná prostredníctvom služby Solid. WebID umožňuje používateľom prístup k ich osobným údajom a obsahu na webe bez toho, aby museli zadávať svoje prihlasovacie údaje.

WebID je spôsob, ako môžu internetové služby a členovia vedieť, s kým komunikujú. Špecifikácie WebID definujú súbor návrhov editora na prípravu procesu štandardizácie identity, identifikácie a autentifikácie v sieťach založených na HTTP. Protokoly založené na WebID (Solid OIDC, WebID-TLS, WebID-TLS+Delegation) ponúkajú nový spôsob prihlasovania sa do internetových služieb. Namiesto použitia hesla sa napríklad člen odkáže na inú webovú adresu, ktorá sa za ňu môže zaručiť. WebID nie je konkrétna služba alebo produkt.

Technicky povedané, WebID je HTTP URI, ktoré označuje ("odkazuje" alebo "mená") agenta v sieti založenej na HTTP, ako je napríklad web alebo podnikový intranet. V súlade s princípmi prepojených údajov, keď je WebID dereferencované ("vyhľadané"), zmení sa na profilový dokument (WebID-Profile), ktorý popisuje jeho referenta (to, čo označuje). Tento profilový dokument pozostáva zo štruktúrovaných údajov založených na modeli RDF, pôvodne vytvorených primárne pomocou výrazov zo slovníka FOAF, ale teraz často zahŕňa výrazy z iných slovníkov.

Profilové dokumenty môžu byť uložené na vlastnom webovom serveri agenta a prístup k nim môže byť čiastočne alebo úplne obmedzený na špecifické identity agentov pomocou riadenia prístupu, aby sa zachovalo súkromie subjektu profilového dokumentu.

2.3.3 Dátová integrácia

Dátová integrácia je proces zlúčenia dát z rôznych zdrojov a ich transformácie do jednotného, súvislého a konsolidovaného formátu. Cieľom dátovej integrácie je umožniť efektívne a spoľahlivé spracovanie a analýzu dát bez ohľadu na ich pôvodný zdroj.

Integračné návrhové vzory (Enterprise integration patterns, EIP) sú súborom konceptov a postupov, ako najlepšie nakonfigurovať integrácie medzi systémami, aplikáciami alebo údajmi. Vzory dátovej integrácie tvoria dôležitý základ pre definovanie architektúry dátovej integrácie a budovanie integrácií aplikácií, ktoré pracujú s komplexným ekosystémom údajov a systémov. Vzor integrácie údajov je štandardizovaná metóda na integráciu údajov. Integrácia údajov zahŕňa presun, transformáciu a konsolidáciu údajov vo všetkých ich formách. Architektom tieto vzory pomáhajú k najlepšej integrácii v scenároch, kde existuje množstvo systémov a procesov, ktoré sú vzájomne prepojené.

Integrácia údajov je dôležitým procesom, ktorý umožňuje odomknúť cenné štatistiky kombináciou údajov z viacerých zdrojov do jedného jednotného zobrazenia. Tento proces zahŕňa štandardizáciu údajov a aplikáciu vzorov integrácie údajov, ako je synchronizácia údajov medzi viacerými zdrojmi, migrovanie do cieľového systému alebo agregácia do jedného úložiska. V literatúre sa bežne hovorí o 65 rôznych vzoroch integrácie. Väčšina vzorov súvisí so scenárom zasielania správ, kde si systémy často prenášajú balíky informácií. Vo všeobecnosti možno integrácie rozdeliť do piatich vzorových scenárov:

- Migrácia dát,
- Obojsmerná synchronizácia dát,
- Agregácia,
- Vysielanie (Broadcasting),
- Korelácia.

2.3.3.1 Vzory dátovej integrácie

Migrácia dát

Migrácia údajov je vzor integrácie údajov, ktorý umožňuje presúvať údaje zo zdrojového systému do cieľového systému. Migrácia umožňuje zachovať podnikové údaje pri prechode na rôzne nástroje manažérskych informačných systémov, preto je kritickým komponentom v prípadových štúdiách správy údajov a správe podnikových údajov.

Proces zahŕňa výber údajov, ktoré je potrebné presunúť, ich prípravu na migráciu, extrahovanie zo zdrojového systému a ich transformáciu do štandardného formátu. Migrácia sa najčastejšie využíva, keď podniky prechádzajú z jedného systému do druhého. Zálohovanie údajov a výmena databázového hardvéru sú tiež situácie, pre ktoré je migrácia užitočná. Migrácia je navyše užitočná pri pridávaní uzlov do databázových klastrov alebo pri spustení nového systému rozširujúceho súčasnú infraštruktúru.

Počas procesu migrácie môžu vzniknúť mnohé problémy, ktoré môžu ovplyvniť kvalitu, výkon a bezpečnosť integrácie údajov. Niektoré z hlavných problémov zahŕňajú problémy s kvalitou údajov, ako sú neúplné, nepresné, nekonzistentné alebo zastarané údaje. Na zabránenie alebo vyriešenie týchto problémov môže byť potrebné vykonať profilovanie údajov, čistenie, validáciu a obohatenie. Okrem toho môže byť potrebné údaje konvertovať, zmapovať alebo obohatiť tak, aby zodpovedali schéme a formátu cieľového systému, čo môže spomaliť rýchlosť migrácie údajov alebo ju skomplikovať. Na vyriešenie týchto problémov s transformáciou môže byť potrebné navrhnuť, otestovať a optimalizovať logiku transformácie údajov. Medzi riziká migrácie riadime tiež riziko porušenia súkromia - aby sa tomu zabránilo, šifrovanie, maskovanie, anonymizácia a audit by sa mali implementovať pred alebo po migrácii údajov.

Obojsmerná synchronizácia dát

Obojsmerná synchronizácia je kombináciou dvoch súborov údajov z rôznych systémov, ktoré sa vykonávajú nezávisle, pričom existujú ako samostatný súbor údajov. To sa líši od migrácie alebo vysielania, pretože údaje sa presúvajú viac ako jedným smerom.

Obojsmerná synchronizácia je vzor integrácie údajov, ktorý umožňuje dvom rôznym systémom fungovať ako jeden, pričom údaje stále uchovávajú oddelene. Tento proces pomáha organizáciám, ktoré potrebujú používať viacero systémov a obchodných procesov súčasne, pretože eliminuje potrebu manuálnej údržby údajov. Obojsmerná synchronizácia tak zabezpečuje vysokú kvalitu údajov a dostupnosť údajov v reálnom čase a optimalizuje obchodné procesy. Tento typ synchronizácie je komplikovanejší, keď dva systémy majú odlišné dátové modely, pretože môže byť potrebné implementovať ďalšie polia a transformácie dát.

Príkladom obojsmernej synchronizácie môže byť spoločnosť poskytujúca finančné služby, ktorá má rôzne systémy pre rôzne obchodné procesy. Obojsmerná synchronizácia by aktualizovala rôzne systémy s rovnakými zdieľanými údajmi v reálnom čase. Vďaka neustálej dostupnosti údajov v reálnom čase by spoločnosť poskytujúca finančné služby mohla stále udržiavať systémy špecifické pre alternatívne obchodné procesy. Hoci existujú rôzne súbory údajov, tieto dva systémy sa správajú ako jeden a zdieľajú pohľad na údaje v reálnom čase. Obojsmerná synchronizácia zachováva integritu údajov naprieč týmito systémami a zároveň umožňuje každému vykonávať svoje vlastné individuálne funkcie.

Agregácia

V scenári agregácie údajov sa údaje prenášajú z viacerých zdrojov do jedného cieľa. Pri agregácii údajov prostredníctvom integrácie sa akceptuje, že nie všetky systémy obsahujú rovnaké údaje, ale je potrebné zhromažďovať informácie na jednom mieste.

Agregácia je proces kombinovania alebo spájania údajov z viacerých zdrojov do jedného cieľového systému, ktorý poskytuje centralizovaný pohľad na údaje v reálnom čase z viacerých zdrojov a zabraňuje duplicitu údajov. Agregácia môže byť užitočná pri vytváraní rozhraní API na získanie údajov z viacerých systémov a ich doručenie v jednej odpovedi, alebo keď je potrebné údaje umiestniť do jedného systému, pričom pochádzajú z viacerých zdrojov. Okrem toho pomáha pri integrácii údajov tým, že udržiava kvalitu údajov, pričom sa vyhýba potrebe replikácie kapacít dátového skladu.

Dátová integrácia agregovaním údajov je opakom vysielania. Namiesto presúvania údajov z jedného zdroja do viacerých cieľov berie alebo prijíma údaje z viacerých zdrojov a presúva ich do jedného cieľa, takže sú vždy aktuálne a nebudú sa zbytočne duplikovať. Agregácia je užitočná pre scenáre ako:

- Modernizácia starších systémov pomocou API.
- Kombinácia údajov z viacerých systémov v zostavách alebo informačných paneloch.
- Udržiavanie súladu s údajmi z viacerých systémov.

Výhodou agregácie údajov je, že umožňuje udržať zdrojové systémy pomerne nedotknuté, zatiaľ čo cieľový systém musí byť schopný vyrovnávať sa s rôznymi údajmi získanými zo zdrojových systémov. Agregácia sa bežne vyskytuje v kontexte business intelligence (BI), kde sa informácie z viacerých zdrojov spájajú, aby sa zobrazili na paneloch a správach.

Vysielanie

Vysielanie je vzor integrácie údajov, ktorý sa zameriava na presun údajov z jedného zdroja do niekoľkých cieľových systémov nepretržite a v reálnom čase. Tento vzor umožňuje prístup k údajom v reálnom čase medzi viacerými systémami tým, že súčasne replikuje informácie prichádzajúce z jedného zdroja do každého z cieľov. Vysielanie je efektívnejšie ako migrácia údajov, pretože pri každom spustení potrebuje spracovať iba najnovšie prichádzajúce správy, čím sa zabezpečí rýchla a spoľahlivá aktualizácia údajov na viacerých miestach a zároveň sa minimalizuje riziko zlyhania transakcie.

Rovnako ako migrácia, integrácia údajov pomocou vysielaním má transakčný charakter - presúvanie údajov iba jedným smerom (zo zdroja do cieľa), čo robí tento vzor integrácie údajov jedinečným. Na rozdiel od migrácie však nepresúva celé množstvo údajov v jednom časovom bode. Namiesto presúva iba údaje, ktoré sa zmenili od poslednej synchronizácie. Týmto spôsobom sú vzory vysielania optimalizované pre rýchle a spoľahlivé spracovanie záznamu.

Korelácia

Korelácia je podobná obojsmernej synchronizácii, ale vykonáva sa iba s údajmi relevantnými pre oba systémy. To znamená, že pri vykonávaní synchronizácie berie do úvahy iba priesečník dvoch súborov údajov (položky, ktoré sa prirodzene vyskytujú v oboch). Kľúčový rozdiel je v tom, že korelácia odstraňuje irelevantné údaje z procesu zjednotenia, vďaka čomu je tento vzor jednoduchší a efektívnejší.

K tejto synchronizácii dochádza iba pri údajoch, ktoré sa nachádzajú v oboch množinách údajov, takže nepotrebné údaje sa neukladajú ani nesynchronizujú. Okrem toho nie je možné v rámci tejto integrácie vytvárať nové záznamy. Na rozdiel od obojsmernej synchronizácie korelácia jednoducho pretína dva súbory údajov a nepokúša sa ich zjednotiť.

Korelácia je ideálna, keď dva systémy chcú zdieľať záznamy, ktoré prirodzene existujú v oboch súboroch údajov. Nemocnica môže napríklad chcieť záznamy o pacientoch od iných lekárov, ale len pre pacientov, ktorí navštívili ich zariadenie. Agregácia všetkých pacientov by bola nadbytočná a mäťúca, takže na nájdenie priesečníka existujúcich pacientov a záznamov mimo nemocnice by sa použila korelácia.

2.4 Súčasný stav v rámci Slovenskej republiky

Predtým ako uvedieme príklady dobrej praxe zo zahraničia je vhodné zhrnúť stav využívania referenčných údajov, základných alebo referenčných registrov a interoperability na Slovensku.

V prvom rade je potrebné uviesť, že Slovenská republika nedisponuje národným rámcom pre interoperabilitu, teda koncepčným modelom ako je EIF (European Interoperability Framework) alebo BRAI (Base Registries Access and Interconnection Framework), ktoré poskytujú jednotné a zrozumiteľné princípy, podmienky a kritériá pre vytváranie základných registrov a ich využívanie pre zdieľanie údajov naprieč hranicami a sektormi. Tieto rámce sa sústreďujú na interoperabilitu základných registrov, pre ktorú definujú nasledovné základné atribúty:

- Právna interoperabilita
- Organizačná interoperabilita

— Sémantická a technická interoperabilita.

Právna interoperabilita

Z hľadiska právnej interoperability, teda vytvorenia právnych resp. zákonných podmienok pre interoperabilitu základných registrov, sa slovenský právny rámec skladá z piatich základných zákonných pilierov:

- a) zákon č. 211/2000 Z. z. o slobodnom prístupe k informáciám, ktorý garantuje prístup fyzických a právnických osôb k informáciám a údajom, ktorými disponuje verejná správa
- b) zákon č. 95/2019 Z. z. o informačných technológiách vo verejnej správe, ktorý reguluje základné podmienky pre zaistenie integrity a bezpečnosti informačných systémov verejnej správy, ako aj správu a prevádzku ústredného portálu verejnej správy
- c) vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 179/2020 Z. z., ktorou sa ustanovuje spôsob kategorizácie a obsah bezpečnostných opatrení informačných technológií verejnej správy, ktorá stanovuje technické štandardy, bezpečnostné štandardy a minimálne štandardy pre technickú podporu alebo dátové štandardy.
- d) zákon č. 305/2013 Z. z. o elektronickej podobe výkonu pôsobnosti orgánov verejnej moci (zákon o e-Governmente), ktorého cieľom je najmä zefektívniť a zjednodušiť existujúce právne procesy spojené so základnými registrami, napríklad zabezpečenie synchronizácie dát prostredníctvom referencovania údajov.
- e) Zákon č. 177/2018 Z. z. o niektorých opatreniach na znižovanie administratívnej záťaže využívaním informačných systémov verejnej správy (zákon proti byrokracii), ktorý vytvoril právny základ pre uplatňovanie princípu jedenkrát a dosť, v rámci ktorého sú orgány verejnej moci povinné a zároveň oprávnené získavať a používať údaje evidované v informačných systémoch verejnej správy a vyhotovovať si z nich výpisy a tieto údaje a výpisy si navzájom bezodplatne poskytovať.

Na základe týchto legislatívnych aktov je možné sledovať, že tu je dlhodobá snaha smerovať k efektívnemu zdieľaniu údajov, ktoré sa majú jednotlivé orgány verejnej moci k dispozícii, a to na základe nasledujúcich troch princípov:

- Princíp „jeden-krát a dosť“.
- Princíp referencovania a stotožňovania údajov.
- Princíp otvorených dát a údajov .

Princíp „jeden-krát a dosť“ bol na Slovensku uvedený do praxe na základe zákona proti byrokracii a prostredníctvom webu over.si, ktorý mal zabezpečiť prístup k potrebným údajom. Postupne sa v troch vlnách zrušila povinnosť predkladať v konaniach výpisy z dvadsiatich informačných systémov verejnej správy (ide o kľúčové systémy, ako napríklad matriky, kataster, obchodný register, register trestov, informačné systémy Finančnej správy alebo Register fyzických osôb).

Zákon proti byrokracii sa snažil riešiť aj jeden z hlavných problémov konceptu referenčných údajov a stotožňovania ako sú vnímané podľa zákona o e-Governmente, a to nesúlad medzi všeobecnou legislatívou a špeciálnymi zákonmi. Zákon proti byrokracii novelizoval a upravil všetky právne predpisy, ktoré predkladanie výpisov predtým explicitne vyžadovali.

Zákon proti byrokracii zaviedol do slovenského právneho poriadku princíp „jeden-krát a dost“ pomerne efektívnym spôsobom (najmä z pohľadu rýchlosti), využil pri tom fungujúcu dátovú infraštruktúru (CSRÚ) a možnosť využitia webového prístupu (cez over.si).

Nevýhodou takéhoto rýchleho riešenia je však v konečnom dôsledku implementácia nekoncepčného riešenia, najmä z pohľadu dátovej interoperability a automatizovaného spracovania údajov.

Údaje sa v praxi vymieňajú najmä ako elektronické verzie papierových dokumentov (výpisy) a pracuje sa s nimi rovnako ako s papierovými dokumentami. Keďže úradníci môžu pristupovať k širokej množine výpisov o občani, ochrana osobných údajov je problematická, pretože chýba napríklad systém notifikácií o prístupoch k údajom, aby mali dotknuté osoby prehľad o tom, ako sa s ich osobnými údajmi nakladá a tým aj kontrolu nad svojimi osobnými údajmi.

Takáto implementácia princípu „jeden-krát a dost“ do praxe zároveň zastavila legitímne a správne snahy používať referencovanie údajov a stotožňovanie. Tento prístup predchádzal prijatiu princípu „jeden-krát a dost“ a bol z pohľadu interoperability a možnosti automatizácie koncepcnejší.

Referencovanie údajov je však prístup, ktorý je oveľa náročnejší na zdolávanie vysokých bariér, ktoré sa v systémoch nachádzajú. Ide napríklad o bariéry, ktoré si vyžadujú významné zmeny informačného prostredia na strane konzumentov, komplikovaný proces vyhlasovania referenčných údajov, alebo obavy orgánov verejnej správy vyhlásiť referenčné údaje spojené s nízkou kvalitou údajov.

Z vyššie uvedeného je zrejmé, že pre budúcnosť sa javí ako kľúčové, aby boli tieto bariéry riešené novým konceptom, ktorý bude založený na všetkých troch princípoch, teda princípe „jeden-krát a dost“, princípe referencovania a princípe dostupnosti údajov, ktorý bol uvedený do praxe prostredníctvom ústredného portálu verejnej správy, a ktorý mal byť platformou pre efektívnu výmenu a zdieľanie údajov naprieč sektormi a bez ohľadu na hranice.

Organizačná interoperabilita

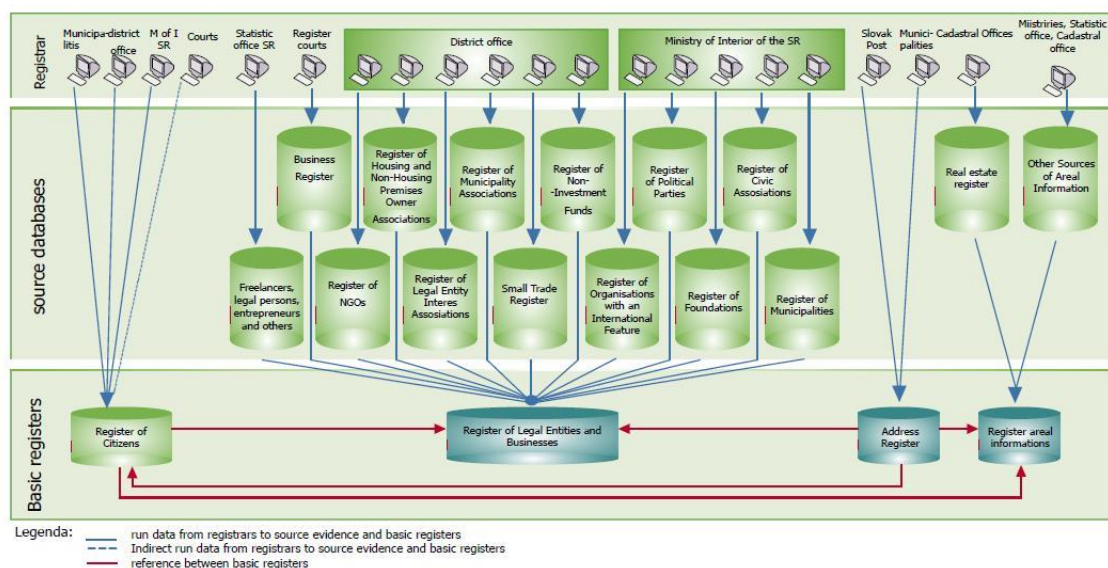
Referencovanie údajov a referenčné registre (základné registre) sú hlavným pilierom organizačnej interoperability. V súčasnosti fungujú nasledovné referenčné registre obsahujúce vyhlásené referenčné údaje:

- **Register fyzických osôb** (obsahuje 31 vyhlásených referenčných údajov, napríklad „meno“, „priezvisko“, „rodné priezvisko“, „rodné číslo“, „dátum narodenia“, „miesto narodenia“, alebo „trvalý pobyt“)
- **Register adries** (obsahuje 10 referenčných údajov, napríklad „názov kraja“, „názov okresu“, „skrátene názov okresu“, „názov obce“, „názov ulice“, alebo „súpisné číslo“)

- **Register právnických osôb, podnikateľov a orgánov verejnej moci** (obsahujúci 38 referenčných údajov, napríklad „obchodné meno alebo názov právnickej osoby“, adresa sídla právnickej osoby“, identifikačné číslo organizácie“, „vklady alebo podiely pre právnickú osobu“, alebo „právna forma právnickej osoby“)
- **Informačný systém služieb zamestnanosti** (obsahujúci referenčné údaje „dátum zaradenia uchádzača o zamestnanie do evidencie uchádzačov o zamestnanie“ a „dátum vyradenia uchádzača o zamestnanie z evidencie uchádzačov o zamestnanie“)
- **Register daňových subjektov** (obsahujúci referenčný údaj „identifikačné číslo pre daň z pridanej hodnoty“).

Nasledovný obrázok ilustruje štruktúru základných registrov, ich vzťahy s inými zdrojmi údajov a príslušných správcov registrov, zodpovedných za vedenie týchto registrov (Factsheet: Access to Base Registries in Slovakia, Európska Komisia, 2018)

Obrázok 1: Štruktúra základných registrov



Napríklad register právnických osôb, podnikateľov a orgánov verejnej moci je vedený Štatistickým úradom Slovenskej republiky, ako bolo stanovené prvou verziou Národnej koncepcie informatizácie verejnej správy (NKIVS, 2016). Tento referenčný register obsahuje údaje o entitách, ktorých zdrojom je viac ako 70 registrov a rôznych druhov informačných systémov.

Centrálny metainformačný systém verejnej správy (METAIS) slúži pre podporu riadenia informatizácie verejnej správy a systému pre tvorbu politík v oblasti informačnej spoločnosti, ako aj pre podporu cieľov efektívnej verejnej správy, ktorými sú vytvorenie inkluzívnej informačnej spoločnosti ako prostriedku pre rozvoj vysoko výkonnej vedomostnej ekonomiky. METAIS v praxi slúži ako katalóg základných registrov a poskytuje informácie potrebné pre zaradenie nových registrov do katalógu, spôsoby úpravy registrov a informácie správe referenčných údajov.

Sémantická a technická interoperabilita

Štandardom pre informačné systémy verejnej správy pre zabezpečenie technickej a sémantickej interoperability sú základné číselníky.

Vyplýva to vyhlášky Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 78/2020 Z. z. o štandardoch pre informačné technológie verejnej správy, ktorou sa stanovujú štandardy pre informačné technológie verejnej správy, ktorými sú štandardy vzťahujúce sa na technické prostriedky, sieťovú infraštruktúru a programové prostriedky, a to štandardy prepojenia, prístupu k elektronickým službám, webových služieb, integrácie dát, štandardy prístupnosti prístupnosti a funkčnosti webových sídiel a mobilných aplikácií vzťahujúce sa na aplikačné programové vybavenie podľa zákona, štandardy použitia súborov vzťahujúce sa na formáty výmeny údajov, štandardy názvoslovia elektronických služieb vzťahujúce sa na sieťovú infraštruktúru, dátové štandardy vzťahujúce sa na údaje, registre a číselníky, štandardy elektronických služieb verejnej správy vzťahujúce sa na údaje, registre, číselníky a aplikačné programové vybavenie podľa zákona, štandardy poskytovania údajov v elektronickom prostredí vzťahujúce sa na databázové prostredie, spoločné moduly, aplikačné programové vybavenie, údaje, registre, číselníky a formáty výmeny údajov, štandardy poskytovania cloud computingu a využívania cloudových služieb vzťahujúce sa na technické prostriedky a programové prostriedky, štandardy formátov elektronických dokumentov podpísateľných elektronickým podpisom, štandardy základných číselníkov.

Táto vyhláška definuje:

- **číselník** ako množinu údajov vo forme jednotlivých položiek číselníka, ktoré sú opísané najmenej dvojicou dátových prvkov „kód položky“ a „názov položky“
- **kód položky** ako textový reťazec, ktorý je v číselníku jedinečný
- **dataset** ako ucelená a samostatne použiteľná skupina súvisiacich údajov vytvorených a udržiavaných na určitý účel, uložených spoločne podľa rovnakej schémy a poskytovaných prostredníctvom súboru alebo aplikačného rozhrania
- **dátový zdroj** ako pôvodné miesto evidencie datasetu
- **dátový prvok** ako jednotka údajov, ktorá je jednoznačne a nedeliteľne špecifikovaná prostredníctvom súboru atribútov.

Táto vyhláška tiež definuje štandard otvoreného údaj, pričom **za otvorený údaj je možné považovať údaj ak** sú právne aspekty prístupu k údaju a jeho používaniu explicitne vysporiadané, je umožnené vytvorenie právnych vzťahov pre používanie údajaj prostredníctvom anonymného vzdialeného automatizovaného prístupu, prístup k údajej je umožnený všetkým osobám za rovnakých podmienok, pričom tieto podmienky sú explicitne uvedené, údaj je možné použiť na nekomerčný aj komerčný účel a je možné ho kombinovať s inými údajmi, dopĺňať, opravovať, modifikovať alebo použiť z datasetu bez povinnosti použitia ostatných údajov datasetu a poslednou podmienkou otvoreného údajaj, že jeho použitie je bezodplatné.

Štandardom základných číselníkov zverejňovaných alebo sprístupňovaných na použitie inými informačnými systémami verejnej správy je:

- používanie dátovej štruktúry základného číselníka podľa prílohy č. 9 vyhlášky o štandardoch pre informačné technológie verejnej správy

- vytváranie názvu základného číselníka tak, že stručne a zrozumiteľne opisuje obsah tohto číselníka
- vytváranie položiek základného číselníka najmenej v slovenskej jazykovej verzii
- vytváranie jazykových verzii položky základného číselníka prostredníctvom príslušných dátových prvkov ako súčasti príslušnej položky základného číselníka
- poskytovanie jazykovej verzie, len ak sú hodnoty dátového prvku „Názov položky“ všetkých položiek základného číselníka vyplnené v príslušnej jazykovej verzii
- vytváranie dátových prvkov položky základného číselníka typov „Lokalizovaný dátový prvok“ a „Lokalizovaný dátový prvok s históriou“ v jednom jazyku jedenkrát
- nevytváranie súčasne účinných dátových prvkov položky základného číselníka typov „Dátový prvok s históriou“ a „Lokalizovaný dátový prvok s históriou“
- upravovanie alebo pridávanie hodnôt dátových prvkov položky základného číselníka typov „Dátový prvok s históriou“ alebo „Lokalizovaný dátový prvok s históriou“, ktorá nadobudla účinnosť ukončením účinnosti príslušného pôvodného dátového prvku, ak existuje, a vznikom nového dátového prvku s upravenou alebo pridanou hodnotou a s účinnosťou nasledujúcou bezprostredne po ukončenej účinnosti pôvodného dátového prvku na úrovni sekúnd; ak ide o nový dátový prvok, ktorý položka dovtedy neupravovala alebo ktorý nemal hodnotu, účinnosť sa určuje na základe vlastného uváženia
- upravovanie alebo pridávanie hodnôt dátových prvkov položky základného číselníka iného typu ako uvedeného v predchádzajúcom bode priamym prepísaním existujúcich hodnôt alebo pridaním nových hodnôt
- nemožnosť vymazania položky základného číselníka, ktorá nadobudla účinnosť, a to ani po jej ukončení
- vytváranie položky základného číselníka s obnovenou účinnosťou zmenou dátového prvku „Účinný od“ tejto položky vyššie uvedeným postupom, pričom iné dátové prvky tejto položky sa vtedy neupravujú
- poskytovanie všetkých povinných dátových prvkov položky základného číselníka podľa prílohy č. 9 s vyplnenými hodnotami, a to pre všetky položky základného číselníka
- vytváranie identifikátorov položiek základného číselníka ako jednotných referencovateľných identifikátorov, a to podľa zoznamu referencovateľných identifikátorov zverejňovaného prostredníctvom centrálného metainformačného systému.

Štruktúru základného číselníka je možné na vnútorné účely použitia iným informačným systémom verejnej správy v tomto informačnom systéme rozšíriť o nové atribúty alebo položky, pričom takto upravený číselník sa označuje ako rozšírený číselník a tento číselník nie je základným číselníkom. Základné číselníky sa v elektronických formulároch používajú spravidla so všetkými položkami, ktoré sú v období účinnosti príslušného elektronického formulára legislatívne uzatelné. Pri prenose dátových prvkov alebo ich atribútov, ktorých hodnoty sú založené na základných číselníkoch, sa používa najmenej jednotný referencovateľný identifikátor položky základného číselníka.

Vyhlášku o štandardoch pre informačné technológie verejnej správy v roku 2021 zmenila a doplnila vyhláška ministerstva investícií, regionálneho rozvoja a informatizácie

Slovenskej republiky č. 546/2021 Z. z., pričom zmeny a doplnenia sa týkajú najmä štandardov pre grafické používateľské rozhranie elektronickej služby verejnej správy a vytvárania mobilných aplikácií verejnou správou.

2.5 Dobrá prax zo zahraničia

Analýza riešení problémov interoperability v rámci zahraničia (primárne EÚ) ukazuje viacero používaných prístupov na organizačnej, sémantickej a technologickej úrovni. Líši sa miera centralizácie, použité štandardy, štýly správy kmeňových údajov a podobne. Spôsob riešenia je do veľkej miery ovplyvnený historickými okolnosťami vývoja systému verejnej správy v jednotlivých krajinách. Vďaka snahe EÚ jednotlivé riešenia konvergujú v rámci iniciatívy technickému systému pre „jeden-krát a dost“ (Once-Only Technical System - OOTS¹).

2.5.1 Prehľad zahraničných riešení

Tabuľka 2: Porovnanie prístupov krajín EU k interoperabilite

Interoperabilita Krajina	Princíp jednou a dost' (<i>Once-Only Principle, OOP</i>)	Organizačná int. – Katalóg základných registrov (ZR)	Organizačná Správa základných registrov (ZR)	Medzinárodná interoperabilita
SK	Neuplatňuje OOP	Má katalóg ZR (MetaIS)	Centralizovaná správa hlavných ZR na národnej úrovni	Člen EUCARIS, ECRIS, EBR, a je čiastočne pripojeným členom EULIS
AT	Uplatňuje OOP	Nemá katalóg ZR	Centralizovaná správa hlavných ZR na národnej úrovni	Člen EUCARIS, ECRIS, EBR, a je plne pripojeným členom EULIS
CZ	Uplatňuje OOP	Má katalóg ZR (Služby-ISVS)	Decentralizovaná správa hlavných ZR na národnej a miestnej úrovni	Člen EUCARIS, ECRIS, EBR, ELF, a je čiastočne pripojeným členom EULIS

¹ <https://ec.europa.eu/digital-building-blocks/wikis/display/OOTS/About+OOTS>

Interoperabilita Krajina	Princíp jednou a dost' (<i>Once-Only Principle, OOP</i>)	Organizačná int. – Katalóg základných registrov (ZR)	Organizačná Správa základných registrov (ZR)	Medzinárodná interoperabilita
DE	Neuplatňuje OOP	Nemá katalóg ZR	Decentralizovaná správa hlavných ZR na národnej a miestnej úrovni	Člen EUCARIS, ECRIS, EBR
HU	Čiastočne uplatňuje OOP	Nemá katalóg ZR	Centralizovaná správa hlavných ZR na národnej úrovni	Člen EUCARIS, ECRIS
PL	Neuplatňuje OOP	Nemá katalóg ZR	Centralizovaná správa hlavných ZR na národnej úrovni	Člen EUCARIS, ECRIS, ELRA

2.5.2 Otvorený prístup k registrom a referencovaniu (UK)

Central Digital & Data Office, ktorý je súčasťou britského úradu vlády (Cabinet Office), vedie digitálnu, dátovú a technologickú agendu vlády a stanovuje stratégiu a štandardy pre technológie vo verejnej správe, vydalo usmernenie pre používanie referenčných údajov obsahujúce najlepšiu prax vyplývajúcu zo skúseností v tejto oblasti.

Toto usmernenie uvádza nasledovných päť krokov, ktoré sú kľúčové pre úspešnú implementáciu a používanie referenčných údajov.

1. Definovanie stratégie publikovania referenčných údajov
2. Vymenovanie vlastníka referenčných údajov a správcu referenčných údajov pre manažment stratégie publikovania
3. Vytvorenie a používanie jediného dôveryhodného zdroja referenčných údajov
4. Zverejnenie referenčné údaje tak, aby boli použiteľné a bezpečné
5. Poskytovanie užívateľskej podpory pre používanie referenčných údajov.

Definovanie stratégie publikovania referenčných údajov

Referenčné údaje kategorizujú ostatné údaje, pomáhajú im dať zmysel a robia ich spoľahlivými a dôveryhodnými. Referenčné údaje zvyčajne pozostávajú z kódov, popisov a definícií údajov. Napríklad kódy krajín ISO-3166 sú medzinárodne uznávaným súborom kódov, ktoré môžete použiť napríklad na označenie krajín.

Referenčné údaje sú cenným aktívom, ktoré môže informovať používateľov a pomôcť im pri rozhodovaní. Publikovanie referenčných údajov pre ostatných v rámci štátnej správy je vynikajúcim spôsobom, ako sa podeliť o svoje výhody a zároveň šetriť čas a náklady pre ostatných.

Referenčné údaje je nevyhnuté na zverejnenie správne pripraviť. Referenčné údaje by sa mali zverejňovať tak, aby boli:

- vždy dostupné a aktuálne
- overené a presné v akomkoľvek okamihu, s históriou zachovaných zmien a verziami, ktorých platnosť vypršala a stále dostupné na použitie
- vyhľadateľné a dostupné aj pre používateľov
- podporované organizačnou infraštruktúrou, ktorá dokáže zvládnuť procesy a nároky na vytváranie a údržbu samotných referenčných údajov.

Vymenovanie vlastníka referenčných údajov a správcu referenčných údajov pre manažment stratégie publikovania

Je kľúčové efektívne riadiť stratégiu publikovania referenčných údajov. Dobrým spôsobom, ako to urobiť, je vymenovať vlastníka referenčných údajov a správcu referenčných údajov, aby spolupracovali na správnej príprave referenčných údajov a podporovali ich používateľov po ich zverejnení.

Vlastník referenčných údajov „vlastní“ referenčné údaje a jeho hlavnou zodpovednosťou je:

- definovať referenčné údaje,
- vytvárať pravidlá a robiť rozhodnutia týkajúce sa referenčných údajov,
- rozhodnúť, kto môže pristupovať k referenčným údajom a meniť ich.

Správca referenčných údajov je vykonávateľom pravidiel stanovených vlastníkom referenčných údajov a mal by byť zodpovedný za:

- kvalitu údajov,
- súlad s regulačnými požiadavkami,
- súlad s dátovými princípmi a zásadami vašej organizácie,
- riešenie akýchkoľvek praktických problémov súvisiacich s referenčnými údajmi.

Vlastník referenčných údajov a správca nemusia byť jednotlivci. Zodpovednosť môže byť rovnomerne rozložená v rámci tímu, ktorý bude spoločne alebo jednotlivo plniť úlohy správcu a vlastníka. Je tiež vhodné vytvoriť „forum referenčných údajov“ zložené z jednotlivcov pochádzajúcich z relevantných častí príslušnej organizácie, aby diskutovali o možnostiach a spôsoboch správy referenčných údajov.

Vytvorenie a používanie jediného dôveryhodného zdroja referenčných údajov

Referenčné údaje by mali byť vytvorené z jedného dôveryhodného zdroja údajov. A tento zdroj by mal byť umiestnený v úložnom systéme alebo databáze a mal byť známy ako systém záznamov.

V niektorých situáciách možno bude nutné vytvoriť referenčnú množinu údajov kombináciou niekoľkých zdrojov údajov. V týchto prípadoch by mal mať každý publikovaný súbor referenčných údajov svoj vlastný systém záznamov vytvorený spojením akýchkoľvek existujúcich systémov záznamov, ktoré k nemu prispeli.

Každý záznam v referenčnej množine údajov by mal byť označený jedinečným identifikátorom, aby mohol byť priradený k rovnakému systému záznamov počas životnosti množiny údajov. To uľahčí používateľom indexovanie, vyhľadávanie a správu referenčných údajov, ako aj sledovanie zmien medzi publikovanými verziami.

Jedinečný identifikátor pomáha efektívne odlišiť jeden záznam od iného záznamu v množine údajov. Syntax jedinečného identifikátora by mala pozostávať z písmen, čísel alebo ich kombinácie. Príklady zahŕňajú sériové čísla, skladové jednotky, ktoré sa nachádzajú v rámci čiarových kódov na položkách na predaj, alebo kódy mien, ktoré sa nachádzajú v službách medzinárodnej meny. Mali by tiež zabezpečené, aby jedinečné identifikátory boli trvalé, čo znamená, že zostanú nezmenené počas celej životnosti referenčných údajov, aby sa zaistila presnosť a konzistentnosť pre používateľov.

Pri vytváraní nových systémov záznamov a referenčných súborov údajov by sa malo postupovať podľa všeobecne záväzných pravidiel týkajúcich sa opätovného použitia údajov vždy, keď je to možné. Pomôže to znížiť plytvanie a duplicitu.

Zverejnenie referenčné údaje spôsobom, aby boli použiteľné a bezpečné

Publikovaný súbor referenčných údajov musí byť čitateľný pre ľudí aj stroje. Odporúča sa prístup založený na rozhraní API, ktorý zverejňuje referenčné údaje vo formáte JSON. Druhou možnosťou je publikovanie vo formáte CSV (CSV na webe), ak by používatelia potrebovali súbor CSV. Pri publikovaní vo forme dokumentu by mal byť použitý štandard Open Document Format (ODF). Najdôležitejším kritériom pri výbere formátu je, či je formát, v ktorom majú byť údaje publikované, najvhodnejší pre potreby používateľov.

Do množiny referenčných údajov by mali byť zahrnuté metadáta, ktoré poskytujú:

- prehľad obsahu referenčného súboru údajov
- kontaktné údaje svojho správcu
- informáciu kedy bola množina vytvorená
- informáciu kedy bola množina naposledy aktualizovaná
- stručný popis všetkých nových zmien v najnovšej verzii.

V rámci publikovania sa môžu vyskytnúť rôzne komplikácie a problémy. Napríklad, je možné, že aktualizácia množiny referenčných údajov si bude nakoniec vyžadovať jej zverejnenie ako úplne novej, samostatnej množiny referenčných údajov. Alebo môže nastať prípad, že zmena referenčných údajov, ktoré sa už používajú, spôsobí poruchu systémov alebo platforiem, ktoré ich používajú. Ďalším príkladom komplikácie je, že

štandardná priemyselná klasifikácia ekonomických činností Spojeného kráľovstva musí stále poskytovať referenčné údaje za roky 2003 aj 2007, aby vyhovovali rôznym prípadom použitia.

Pri publikovaní novej verzie súboru referenčných údajov spolu s existujúcou súpravou by malo byť zabezpečené nasledovné:

- nová verzia je zverejnená ako nový, samostatný referenčný údaj a nie ako zmena alebo variant existujúceho,
- používateľom je k dispozícii existujúca aj nová verzia,
- korelácia medzi existujúcimi a novými verziami je používateľom objasnená, najlepšie v sprievodnom korelačnom dokumente.

Publikované referenčné údaje musia byť zabezpečené. To znamená, že by sa mali nachádzať v zabezpečenom prostredí a prístup k tomuto prostrediu by mal byť spravovaný bezpečne pomocou HTTPS.

Poskytovanie užívateľskej podpory pre používanie referenčných údajov

Pre riešenie užívateľských problémov ale aj pre ďalší rozvoj referenčných údajov je dôležité poskytnúť jednoduchý spôsob, ako používatelia poskytnú spätnú väzbu alebo nahlásia chyby pri používaní referenčných údajov, napríklad prostredníctvom e-mailového odkazu alebo webového formulára.

Je dôležité vedieť, kto používa referenčné údaje. Umožňuje vám poskytovať lepšiu podporu používateľom a ich komunite, a to aj počas aktualizácií, údržby a neočakávaných prestojov. Dobrým spôsobom, ako zistiť, kto používa referenčné údaje, je povzbudiť používateľov, aby sa prihlásili na ich odber, napríklad tým, že používateľom poskytnete možnosť zadať e-mailovú adresu pri ich sťahovaní.

2.5.3 Základné registre s registrom práv a povinností (CZ)

Základné registre sa v Českej republike používajú ako základný (referenčný) dátový zdroj údajov o subjektoch a objektoch evidencie a o verejnej správe. Tieto základné registre obsahujú referenčné údaje o fyzických osobách, právnických osobách, adresách a územných prvkoch a nehnuteľnostiach, orgánoch verejnej správy a súkromnoprávných používateľoch údajov, ako aj o agendách a pôsobnosti výkonu verejnej správy.

Základné registre tvoria kľúčovú infraštruktúru dátového fondu verejnej správy vrátane mechanizmu pseudonymizácie a prepájania identifikácií z jednotlivých agend. Okrem toho poskytujú najmä fyzickým osobám prehľad o využívaní ich údajov jednotlivými používateľmi (OVM, SPUÚ, atď.) a zároveň zaznamenávajú ich poskytovanie.

Za základné registre sa považuje register obyvateľov (ROB), register osôb (ROS), register práv a povinností (RPP) a register územnej identifikácie, adres a nehnuteľností (RUIAN).

Tabuľka 3: Prehľad základných registrov

Názov základného registra	Vlastník	Druhy referenčných údajov	Druhy referenčných údajov	Druhy referenčných údajov
Register obyvateľov (ROB)	Ministerstvo vnútra ČR	Referenčné údaje o občanoch ČR	Referenčné údaje o cudzincoch žijúcich v ČR	Referenčné údaje o iných fyzických osobách majúcich práva povinnosti v ČR
Register práv a povinností (RPP)	Ministerstvo vnútra ČR	Referenčné údaje o právnych spisoch eSbírka	Referenčné údaje o pôsobnosti orgánov verejnej moci	Referenčné údaje o právach a povinnostiach osôb
Register osôb (ROS)	Český štatistický úrad	Referenčné údaje o právnických osobách	Referenčné údaje o podnikajúcich fyzických osobách	Referenčné údaje o orgánoch verejnej moci
Register územnej identifikácie, adres a nehnuteľností (RUIAN)	Český úrad pôdohospodársky a katastrálny	Referenčné údaje o územnej identifikácii	Referenčné údaje o nehnuteľnostiach	Referenčné údaje o adresách

Produkty základných registrov

Občania majú právo na:

- výpis údajov, ktoré sú o ňom vedené v základnom registri
- výpis osôb, oprávnených čítať v jeho registri a tiež osôb, ktoré tak urobili
- výpis záznamu o referenčných väzbách s inými registrami.

Implementácia a správa základných registrov

Správa základných registrov zriadená zákonom č. 111/2009 Zb., o základných registroch je správcom informačného systému základných registrov, zaisťuje ich prevádzku a bezpečnosť. Tiež sprístupňuje referenčné údaje podľa oprávnenia, ktoré je definované v registri práv a povinností. Správa základných registrov je úrad, ktorý je samostatnou organizačnou zložkou štátu a patrí do rozpočtovej kapitoly Ministerstva vnútra.

Harmonogram naplnenia základných registrov referenčnými údajmi v štyroch etapách s konkrétnymi termínmi splnenia bol v Českej republike určený zákonom, čo považujeme za nasledoviahodný príklad.

- Prvá etapa sa skladala z overenia správnosti a úplnosti údajov v editorských informačných systémoch voči sebe navzájom aj voči ďalším významným informačným systémom verejnej správy. Napríklad pre register obyvateľov je potrebné overiť osobné údaje v evidencii obyvateľov a informačnom systéme cudzincov oproti informačnému systému územnej identifikácie, registra poistencov zdravotného a sociálneho poistenia.
- V druhej etape ohraničenej 31. marcom 2012 malo dôjsť k prvotnému zápisu údajov do základných registrov. Vyčistené editorské systémy sa prepojili s informačným systémom základných registrov a došlo k naplneniu jednotlivých registrov.
- V tretej etape ohraničenej 30. aprílom 2012 malo dôjsť k pripojeniu jednotlivých užívateľských agendových informačných systémov do základných registrov. Budú to najvyužívanejšie evidencie, vykazujúce vysokú mieru validácie pre overenie obsahu aj technického zabezpečenia.
- Štvrtá etapa je etapou pripojenia ostatných informačných systémov spravovaných orgánmi verejnej moci a mala byť ukončená samotným spustením projektu základných registrov k 1. júlu 2012.

Oproti systému vyhlasovania referenčných údajov platnému v Slovenskej republike, je takýto harmonogram oveľa efektívnejším nástrojom, o čom svedčí aj porovnanie výsledkov oboch systémov.

Príklady fungovania základných registrov

A/ V prípade narodenia nového občana:

1. register obyvateľov požiada ÚOOÚ o vytvorenie nového ZIFO,
2. ÚOOÚ ho vygeneruje, uloží a vytvorí AIFO, ktoré odošle do registra obyvateľov
3. Register obyvateľov priradí AIFO k menu občana
4. Zároveň informačný systém základných registrov informuje ostatné agendy o vytvorení novej položky.
5. ÚOOÚ generuje zo ZIFO pre každú agendu špeciálnu AIFO a odošle im ich, občan je tak vedený v každej agende pod iným číslom.

B/ V prípade zmeny mena občana:

1. matričný úrad informuje register obyvateľov
2. v registri obyvateľov dôjde k aktualizovaniu údajov - je zapísané nové meno
3. tento údaj sa stáva referenčným údajom a nie je nutné iné úrady informovať.

2.5.4 Systémové katalógy (NL)

Holandsko má vysokú výkonnosť elektronickej verejnej správy, ktorá pramení z takmer dvoch desaťročí politiky elektronickej verejnej správy. Súčasne platnými politikami sú „Digitálna vláda 2017“ a „Digitálna agenda 2020“. Dokument o vízii Digitálna vláda 2017 podporuje digitálny štandardný prístup, pričom sa zaoberá potrebou digitálnej komunikácie medzi občanmi a vládou, optimálnym využívaním údajov vládou a zlepšenie spoločnej infraštruktúry.

Dôležitým faktorom pri rozvoji holandskej elektronickej verejnej správy je systém základných registrov, ktorý implementuje princíp „jeden krát a dost“. Tento systém má poveseť, že nielen zefektívňuje vládu, ale aj znižuje administratívnu záťaž pre občanov a podniky.

Na realizácii súčasného systému zloženého z 12 základných registrov sa pracovalo od roku 2000. Každý z týchto základných registrov je zakotvený zákonom a musí byť v súlade s 12 princípmi. Už v roku 2003 boli dohodnuté zásady výberu základných registrov a požiadavky na legislatívu, ktoré navrhuje orgán zodpovedný za príslušný základný register.

Vzájomné prepojenie medzi holandskými registrami bolo vytvorené prostredníctvom systému „Stelsel catalogus“ (Systémový katalóg) a miera zdieľania údajov sa neustále zvyšuje. Systémový katalóg je online katalóg, register registrácií, ktorý je súčasťou holandského systému základných registrov. Poskytuje prehľad o tom:

- aké údaje sa v registráciách uchovávajú
- aké sú definície týchto údajov,
- aké sú ich vzájomné vzťahy a
- aký je ich právny základ.

Systémový katalóg teda neobsahuje samotné údaje, ale poskytuje prehľad o tom, ktoré údaje kde možno nájsť. Pomocou týchto informácií môžu používatelia systémového katalógu zistiť, ktoré údaje z registrácií sú dôležité pre ich procesy.

Systémový katalóg ponúka prehľad a interpretáciu v rámci údajov vo vlastníctve verejnej správy. Ktoré údaje sa uchovávajú v ktorých registráciách, aký význam majú tieto údaje a ako spolu údaje z rôznych registrácií súvisia. Systémový katalóg ponúka každému, kto potrebuje informácie, východiskový bod na nájdenie miesta, kde možno nájsť požadované informácie.

Systémový katalóg zahŕňa údaje, ktoré sa uchovávajú v rámci registrácií, ako sú: **adresa, osobné alebo agendové údaje.**

Systémový katalóg obsahuje popisy informácií dostupných v pripojených registroch. Tieto metadáta sú dodávané do Systémového katalógu na zodpovednosť držiteľa

registrácie. Pri tomto **federatívnom modeli** sú správcovia registrácií spoločne zodpovední za obsah systémového katalógu. Správcovia majú uzatvorené dohody o doručovaní, správnosti, úplnosti, dostupnosti a aktuálnosti údajov v Systémovom katalógu.

Dáta je v Systémovom katalógu možné vyhľadávať, napríklad pomocou filtrov a ľahko porovnávať. Systémový katalóg zvýšil konzistentnosť údajov okrem iného pomocou klastrových konceptov a systémového štítku. Okrem webovej stránky je možné údaje vyhľadávať aj prostredníctvom rozhraní API. Systémový katalóg je voľne prístupný každému.

Systémový katalóg má viaceré výhody. **Pomocou Systémového katalógu je možné:**

- Získať prehľad o rôznych registráciách s ich vzájomnými vzťahmi
- Získať prehľad o všetkých konceptoch a dátových prvkoch z registrácií
- Porovnať pojmy a dátové prvky
- Pozrieť sa, ktoré údaje sú autentické
- Prehľad z akých právnych predpisov boli odvodené
- Zobrazit' prepojené zdroje údajov
- Použiť obsah systémového katalógu ako prepojené údaje (linked data) vo svojich vlastných pracovných procesoch.

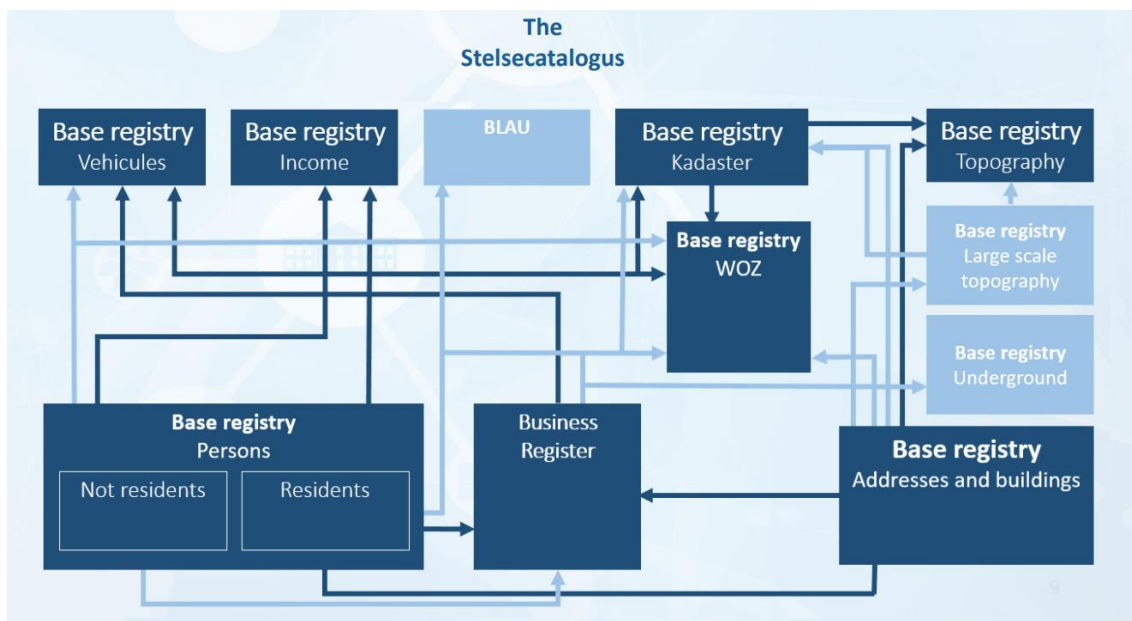
Systémový katalóg ponúka integrovaný prehľad a prehľad širokej škály údajov, ktorými disponuje verejná správa. Správcom Systémového katalógu je Logius, ktorý je súčasťou holandského ministerstva vnútra.

Orgány verejnej správy nežiadajú od občanov, firiem a inštitúcií informácie, ktoré už má. Pre realizáciu tohto princípu v praxi bol vytvorený Systém kľúčových registrov, ktorý obsahuje základné údaje, ktoré sú potrebné pre realizáciu rôznych (rozhodovacích) procesov. Časť údajov v týchto kľúčových registroch bola označená ako „autentická“. To znamená, že vládna organizácia môže v zásade predpokladať, že informácie sú správne, a je povinná ich použiť pri výkone svojich povinností.

Zdrojový kód pre Systémový katalóg možno nájsť ako softvér s otvoreným zdrojovým kódom v Github cez OSSG61 (OpenSourceSoftwareGegevenscatalogus). Je pripravený na inštaláciu a používanie po nastavení prístupu ku koncovému bodu SPARQL, ktorý obsahuje údaje. Technológie a špecifikácie tohto riešenia sú: PHP 5.3 (pdo, mysql, mcrypt, memcached, curl, xsl), MySQL Server 5.x, Apache 2.x (php, rewrite, memcache), štandard W3C LOD a NORA.

Z uvedených informácií je zrejmé, že Systémový katalóg uľahčuje jeho používateľom oveľa ľahšiu orientáciu, vyhľadávanie potrebných informácií a údajov. Tento systém ukazuje dostupnosť informácií a ich význam, zvyšuje použiteľnosť informácií a vytvára pôdu pre harmonizáciu dát v niektorých prípadoch.

Obrázok 2: Konceptuálna architektúra holandského riešenia pre základné registre



3 Definície a účel referenčných údajov

Pre pochopenie definície a účelu referenčných údajov a vylúčenie pochybností alebo nejasností, je potrebné presne definovať aj súvisiace pojmy ako sú informácie, dáta, základné registre, kmeňové dáta, referencovanie údajov, ako aj pojmy registrátor a správca registra. Tieto pojmy, ako sú definované na tomto mieste, vychádzajú z platnej legislatívy účinnej na území Slovenskej republiky, ako aj z legislatívy a materiálov vypracovaných v rámci EÚ.

Informácie a dáta

Informácie sú definované ako dáta dané do kontextu, ktorý vytvára určitý význam. Tak dáta ako aj informácie môžu byť uložené a spracovávané elektronicky v menej štruktúrovanom systéme alebo vo vysoko štruktúrovanom informačnom systéme. Pojmy dáta a informácie sa často zamieňajú, alebo sú používané v rovnakých súvislostiach, ale ich význam sa líši. Navyše, ich význam je odlišný aj pre ľudí pracujúcich v rôznych odvetviach.

Dáta (alebo údaje) sú najčastejšie definované ako základné objektívne fakty, ktoré sú nespojitelné a často merateľné. Dáta môžu, ale nemusia mať samé o sebe význam. Dáta sú zvyčajne zachytené a uložené v počítačových systémoch takmer výlučne v textovej alebo numerickej forme, ale nájdeme ich aj v iných formách, napríklad ako obrázok alebo symbol. Dáta sa stávajú informáciami v momente, keď ich dáme do určitého kontextu a keď sa začnú vzťahovať k iným dátam. Vtedy informácia nadobudne význam a stane sa užitočnou pre nejaký typ rozhodnutia.

Poznáme množstvo druhov dát, ako aj množstvo klasifikačných kritérií, podľa ktorých môžeme dáta rozdeľovať do na rôzne druhy. Diferencovanie dát podľa prístupu k nim, resp. ich zdieľania, je nasledovné:

- **Otvorené dáta** sú dáta zdieľané verejne s minimom obmedzení alebo bez obmedzenia, sú prístupné komukoľvek, kto má záujem o ich opätovné publikovanie, ide napríklad o dáta, ktoré nie sú osobnými údajmi
- **Zdieľané dáta** sú zdieľané medzi štátnymi orgánmi a môžu byť zdieľané aj so súkromnoprávnymi entitami, a to na základe a v súlade s pravidlami ochrany súkromia a dohodami o interoperabilite
- **Zatvorené dáta** nemôžu byť zdieľané s nikým, ide o tajné alebo vysoko chránené dáta.

Metadáta sú v podstate dátami o dátach. Metadáta sú dáta, ktoré označujú popisujú alebo charakterizujú iné dáta, čím uľahčujú obnovovanie, interpretovanie alebo použitie informácií. Metadáta slúžia na to, aby bol splnený cieľ zbierania dát. Akákoľvek disciplína alebo aktivita, pri ktorej sa využívajú dáta by mala byť asociovaná s metadátami.

Historické dáta obsahujú významné fakty, ktoré by nemali byť menené okrem prípadov, kedy dochádza k oprave chýb. Tieto dáta sú veľmi dôležité z hľadiska bezpečnosti a súladu s danými pravidlami (compliance). Operačné systémy môžu tiež obsahovať historické tabuľky pre účely reportingu alebo analytické účely. Príkladmi historických dát sú reporty o udalostiach, ktoré sa stali v určitom čase, informácie o verziách alebo časové snímky.

Dočasné dáta sa udržiavajú v pamäti, aby zrýchľovali procesy. Nie sú viditeľné a slúžia výlučne pre technické účely. Ide napríklad o kópie tabuliek vytvorených počas spracovateľskej relácie pre zrýchlenie vyhľadávania.

Základné číselníky

Základné číselníky (code list) – sú údaje, ktoré sa používajú na kategorizáciu a klasifikáciu iných údajov. Hodnota je väčšinou vyberaná zo zoznamu, ktorý sa mení len vo výnimočných prípadoch. Číselníky slúžia na štandardizáciu, centralizáciu a uchovávanie kľúčových informácií, ktoré sa používajú v rôznych častiach systému alebo aplikácie. Základné číselníky môžu obsahovať hodnoty, ktoré sa využívajú na viacerých miestach, ako napríklad:

- Kódy krajín a regiónov: Základný číselník môže obsahovať kódy jednotlivých krajín alebo regiónov, ktoré sa používajú pri identifikácii alebo adresovaní.
- Typy produktov alebo služieb: Číselník môže obsahovať kódy alebo popisy rôznych typov produktov alebo služieb, ktoré sa používajú pri katalógoch alebo objednávkach.
- Stavy objednávok alebo procesov: Základný číselník môže obsahovať rôzne stavy, ako napríklad "v spracovaní", "dokončené" alebo "zamietnuté", ktoré sa používajú na sledovanie priebehu objednávok alebo procesov.
- Kategórie alebo triedy: Číselník môže obsahovať rôzne kategórie alebo triedy, ktoré sa používajú na klasifikáciu objektov, príkladmi môžu byť kategórie produktov, zamestnancov, projektov a podobne.

Výhody základných číselníkov zahŕňajú jednotnosť, konzistenciu a správu dát na centrálnej úrovni. Zabezpečujú, že rovnaké hodnoty sa používajú v celom informačnom prostredí verejnej správy, čo znižuje chyby, duplicity a nekonzistencie v dátach. Ďalšou výhodou je jednoduchšie aktualizovanie hodnôt, keď je potrebné vykonať zmeny.

Registre, kmeňové dáta a transakcie

Register je štruktúrovaný zoznam jednotiek obsahujúci pre každú jednotku niekoľko atribútov. Register je štruktúrovaný a systematický záznam, ktorý obsahuje informácie o určitých objektoch, subjektoch, udalostiach alebo údajoch. Je to centrálna alebo decentralizovaná databáza, ktorá slúži na zhromažďovanie, uchovávanie a správu údajov o konkrétnych entitách. Register by mal obsahovať mechanizmus pravidelnej aktualizácie.

Registre majú často presne definovanú štruktúru a polia, ktoré určujú typy údajov, ktoré majú byť zaznamenané. Údaje v registri môžu byť buď verejného charakteru a prístupné širokej verejnosti, alebo môžu byť prístupné len určitým oprávneným subjektom.

Registre majú viaceré využitie v rôznych oblastiach, vrátane verejnej správy, obchodu, zdravotníctva, dopravy a mnohých ďalších. Môžu slúžiť na evidenciu občanov, vozidiel, nehnuteľností, podnikov, právnych subjektov, súdnych sporov a mnoho ďalších údajov.

Na aplikačnej úrovni býva register implementovaný

Základné registre (na Slovensku známe ako referenčné registre) sú dôveryhodné a spoľahlivé administratívne zdroje základných informácií o dátových prvkoch, napríklad ľuďoch, spoločnostiach, budovách, vozidlách, licenciách alebo cestách. Tento typ informácií vytvára tzv. kmeňové dáta (angl. master data) využiteľných vo verejnej správe.

Kmeňové dáta sú základné a stabilné informácie o jednotlivých subjektoch alebo objektoch v registri alebo informačnom systéme verejnej správy. Tieto dáta predstavujú statické a trvalé charakteristiky a atribúty, ktoré sa menia zriedka alebo vôbec.

Kmeňové dáta zahŕňajú identifikačné údaje, ako napríklad meno, adresa, dátum narodenia, identifikačné čísla a podobne. Môžu obsahovať aj informácie o právnom postavení, štatúte, historických údajoch alebo iných relevantných charakteristikách jednotlivých subjektov. Tieto dáta slúžia ako základ pre ďalšie spracovanie a analýzu. Sú nevyhnutné pre identifikáciu a rozlišovanie jednotlivých entít a zabezpečujú konzistenciu a spoľahlivosť údajov v systéme.

Kmeňové dáta popisujú ľudí, miesta a veci, ktoré sú nejakým spôsobom zahrnuté v predmete činnosti organizácie. Napríklad ľudia môžu byť zoskupení ako zákazníci, zamestnanci, dodávatelia alebo predajcovia, miesta ako lokality, predajné teritória alebo kancelárie a veci ako produkty, aktíva alebo súbory dokumentov. Typ kmeňových dát môže byť syntetizovaný s odlišnými typmi dát, ktoré poskytujú celostný pohľad informácie verejnej správy. Keďže tieto dáta bývajú používané v rámci viacerých procesov a informačných systémov, normalizačné formáty kmeňových dát synchronizačné hodnoty sú kriticky dôležité pre úspešnosť integrácie systémov. Kmeňové dáta taktiež bývajú zoskupené do kmeňových záznamov, ktoré môžu zahŕňať pridružené referenčné dáta.

Transakčné dáta popisujú internú alebo externú udalosť alebo "transakciu", ktorá sa uskutočňuje v určitej organizácii v rámci jej činnosti. Príklady zahŕňajú objednávky, faktúry, doručovacie dokumenty, platby kreditnými kartami alebo nároky z poistenia. Tieto dáta sú obvykle zgrupované do transakčných záznamov, ktoré zahŕňajú kmeňové údaje a príslušný kontext transakcie. Transakčné dáta obsahujú informácie, ako sú čas a dátum transakcie, identifikátory subjektov alebo objektov zapojených v transakcii, hodnoty, kvantitatívne údaje, popisy a ďalšie podrobnosti súvisiace s konkrétnymi operáciami. Tieto dáta môžu byť zaznamenané a uchovávané v rôznych formátoch, ako sú tabuľky, záznamy, žurnály alebo logy.

V rámci verejnej správy nazývame transakčné dáta aj agendové dáta.

Agendové údaje popisujú jednotlivé udalosti a fakty v rámci danej agendy verejnej správy, ktoré sa evidujú v informačnom systéme verejnej správy, napríklad rozhodnutia, daňové priznania, colné deklarácie, žiadosti, platobné príkazy, zoznamy neplatičov a podobne. Agendové údaje majú vždy časovú dimenziu.

Pre účely tejto kapitoly je tiež dôležité poznať rozdiel medzi zákonom definovanými pojmami **správca registra a registrátor**. Správca registra je orgán verejnej moci, ktorý register vedie a registrátorom je orgán verejnej moci, ktorý je podľa osobitného predpisu oprávnený zapisovať, meniť a vymazávať údaje objektov evidencie z registra.

Referenčné údaje

Subjekt evidencie je osoba, údaje o ktorej sú predmetom evidovania orgánom verejnej moci v rámci jeho pôsobnosti podľa osobitných predpisov. Subjektami evidencie sú tak právnické alebo fyzické osoby.

Objekt evidencie je množina údajov o subjekte evidencie, ktorá je predmetom evidovania orgánom verejnej moci v rámci jeho pôsobnosti a ktorá je jednoznačne identifikovaná identifikátorom objektu evidencie. Objekt evidencie má jasnú štruktúru a môže tak obsahovať referencované údaje zo základných registrov.

Referenčné dáta (údaje) predstavujú súbory hodnôt alebo klasifikačných schém, ktoré sa používajú napríklad v rámci systémov, aplikácií alebo procesov. **Referenčné údaje nesú význam.** Stanovujú prípustné hodnoty, uľahčujú konzistenciu, mapujú interné údaje oproti externým údajom alebo štandardom a ovplyvňujú presnosť vykazovania a správu údajov.

Podľa zákona č. 305/2013 Z. z. o elektronickej podobe výkonu pôsobnosti orgánov verejnej moci (zákon o e-Governmente) je referenčným údajom údaj objektu evidencie, ktorý je uvedený v zozname referenčných údajov, pričom objektom evidencie je podľa tohto zákona množina údajov o subjekte evidencie, ktorá je (v rámci jeho kompetencií) predmetom evidovania orgánom verejnej moci, a ktorá je jednoznačne identifikovaná identifikátorom objektu evidencie. Subjektom evidencie *podľa zákona dnes* môže byť osoba, vec, právo, povinnosť alebo skutočnosť, ktorá je evidovaná orgánom verejnej moci a registrom je podľa tohto zákona miesto elektronického uchovávaného objektov evidencie, ktoré vytvára, uchováva a vykonáva vo vzťahu k nim ďalšie orgány verejnej moci, a to v rámci svojej pôsobnosti podľa osobitných predpisov.

Príklady referenčných údajov zahŕňajú zoznamy platných hodnôt, zoznamy kódov, statusové kódy, skratky, typy produktov alebo ich hierarchia. Štandardizované referenčné dáta sú kľúčové pre integráciu údajov a ich interoperabilitu, pričom umožňujú zdieľanie a reportovanie o informáciách.

Referenčné údaje charakterizujú údaje a spájajú údaje s informáciami v interných aj externých databázach. Referenčné údaje môžu byť využívané napríklad pre odlíšenie jedného typu záznamu od iného pre účely kategorizácie alebo analýzy. Môžu byť taktiež významným faktom ako napríklad krajina, ktorý sa objaví v rámci širšieho informačného súboru akým je adresa.

Referenčné údaje sú teda údaje používané na definovanie a klasifikáciu iných údajov.

Referencovanie údajov takisto priamo definuje zákon o e-Governmente. Referencovaním je použitie údaje objektu evidencie v podobe, ktorá je zhodná s podobou, v ktorej je referenčný údaj vedený v referenčnom registri.

Používanie referenčných údajov resp. referencovanie údajov sa používa za účelom efektívneho zdieľania údajov, ktorého predpokladom sú (základné) registre obsahujúce (referenčné) údaje, ktoré stanovujú štruktúru, podobu a formu, teda akýsi štandard ostatných údajov rovnakého druhu tak, aby boli spôsobilé pre zdieľanie medzi subjektami na to určenými alebo oprávnenými, a to naprieč inštitúciami verejnej správy.

Pôvodný koncept počítal, že sa väčšina kmeňových a agendových údajov vo verejnej správe vyhlási ako referenčné údaje a budú sa následne referencovať, čím sa zabezpečí uplatnenie princípu „jeden-krát a dost“. Takýto predpoklad sa naplniť nepodarilo a súčasný model vyhlasovania a uplatňovania referenčných údajov a referencovania na území Slovenskej republiky nefunguje optimálne, preto je nevyhnutné ho prehodnotiť, čo je aj cieľom tejto publikácie vrátane navrhnutia potrebných zmien. Toto prehodnotenie a navrhnutie zmien bude zohľadňovať nasledovné východiská, ktoré dlhodobý vývoj v tejto oblasti definoval ako kľúčové:

- Základným **nástrojom pre referencovanie údajov budú základné registre** obsahujúce údaje určené na opakované zdieľanie, a to v jednoznačne definovanej štruktúre, ktoré sú získavané a spracovávané v súlade určeným právnym rámcem, stratégiou a pravidlami.
- Základné registre by mali integrovať a prepájať viaceré registre, ktoré musia spĺňať napríklad stanovené podmienky interoperability (interoperability governance) a pravidiel pre správu údajov (data management), ktorých cieľom je vytvoriť vhodné prostredie pre úspešnú implementáciu základných registrov. Jedným zo spôsobov ako toto dosiahnuť je používanie referenčných údajov, ktoré určujú štruktúru, obsah, podobu a formu údajov.
- Pre fungovanie základných registrov je kľúčové správne nastavenie a používanie dátovej architektúry, teda stanovenie štandardov zohľadňujúcich vyššie uvedené kritériá a vytvorenie primeraných procesov.
- Pre efektívnu správu dát je potrebné využívať spoločné dátové modely, uplatňovať jednotný manažment kmeňových dát a metadát a sústrediť pozornosť napríklad na kvalitu a bezpečnosť dát (v cieľovom stave v podobe ontológií v súlade s centrálnym modelom údajov).
- Pre ostatné údaje (mimo objektov evidencie z), ktoré je potrebné v rámci verejnej správy zdieľať je nevyhnuté

Dátové výstupy

Dátové výstupy dostupné cez Centrálnu integračnú platformu predstavujú transformáciu objektu evidencie do reálnej digitálnej podoby, s ktorou je možné ďalej pracovať.

Výpis je elektronický dokument, ktorý obsahuje záznamy alebo údaje z evidencií, registrov, kartoték alebo iných systémov verejnej správy k stanovanému dátumu. Výpis sa používa na poskytnutie informácií o určitých skutočnostiach, právach, povinnostiach, povoleniach alebo iných relevantných údajoch. Môže sa jednať o výpis z obchodného registra, katastra nehnuteľností, registra trestov, registra vozidiel a podobne. Znamená to, že výpis súvisí s kmeňovými údajmi pre subjekt evidencie. Výpis sa vydáva na žiadosť subjektu, ktorý má právo na prístup k týmto informáciám. Formátom výpisu môže byť podpísaný elektronický dokument v PDF, ale aj dátový objekt v strojo-spracovateľnom formáte.

Potvrdenie je elektronický dokument, ktorý slúži na overenie, potvrdenie alebo overenie určitej skutočnosti, udalosti alebo stavu. V rámci verejnej správy môže sa jednať o potvrdenie o zaplatení poplatkov, potvrdenie o prijatí žiadosti, potvrdenie o návšteve školy, potvrdenie o povolení, potvrdenie o trvalom pobyte a podobne. Potvrdenie sa vydáva na základe žiadosti subjektu, ktorý potrebuje potvrdenie na účely osobných, právnych, obchodných alebo iných záujmov. Potvrdenie súvisí s agendovými údajmi pre

objekty evidencie. Formátom potvrdenia môže byť podpísaný elektronický dokument v PDF, ale aj dátový objekt v strojovo-spracovateľnom formáte.

Dátový objekt je výber údajov z objektu evidencie (to znamená týkajúce sa vybraného subjektu evidencie), s ktorým je možné strojovo pracovať. Dátový objekt môžu tvoriť kmeňové alebo transakčné údaje a mohol byť transformovaný do požadovanej podoby. Dôležitou vlastnosťou dátového objektu je jeho formát: CSV, XML, JSON, pričom cieľovým a žiadaným stavom je použitie ontológií prepojených údajov RDF, čo znamená formát JSON-LD.

Nulový dôkaz (zero-knowledge proof) je kryptografický protokol, ktorý umožňuje jednej strane preukázať platnosť nejakej tvrdenia druhej strane bez odhalenia konkrétnych detailov alebo informácií, ktoré by mohli byť považované za citlivé alebo dôverné. Cez CIP je možné distribuovať nulové dôkazy pre konzumentov, napríklad pre overenie veku, identity, právomocí.

Elektronický doklad (podľa konceptu Verifiable Credentials) sú digitálne objekty, ktoré obsahujú identifikačné údaje o jednotlivcovi, ako napríklad meno, dátum narodenia, adresu a ďalšie informácie. Tieto doklady sú založené na otvorených štandardoch a používajú kryptografické techniky na zabezpečenie integrity, autenticity a súkromia údajov. Jednou z hlavných vlastností elektronických dokladov Verifiable Credentials je možnosť overiť pravosť a platnosť týchto dokladov bez potreby centrálného autorizačného orgánu. Overovanie je založené na digitálnych podpisoch, ktoré umožňujú overovateľovi (napríklad poskytovateľovi služby) overiť, že elektronický doklad bol vydávaný dôveryhodným vydavateľom a že údaje v doklade sú pravdivé a nepodvrhnuté.

Konsolidovaný dátový objekt spája (konsoliduje, zlučuje) údaje z viacerých objektov evidencie (ktoré sú vedené vo viacerých registroch) do jednotného dátového objektu (z definovanou štruktúrou) pre vybraný prípad použitia. Štruktúra konsolidovaného objektu je preto daná prípadom použitia (v akom kontexte sú údaje potrebné a aký je právny základ na ich spracovanie) a reálnou dostupnosťou údajov cez Centrálnu integračnú platformu. Konsolidovaný dátový objekt tak obsahuje všetky dostupné a dôležité informácie, ktoré sú potrebné pre rozhodovanie v danom kontexte a časovú pečiatku, ktorá určuje čas, kedy bol vygenerovaný. Použitie konsolidovaných dátových objektov je vhodné pri podpore princípu jeden-krát a dosť v správnom konaní respektíve pri riešení životných situácií.

Dataset môže byť vytvorený z rôznych zdrojov, vrátane meraní, dotazníkov, údajov z webových stránok, databáz, senzorov alebo iných prostriedkov. Sú používané na podporu analytických a výskumných procesov, kde sa tieto dáta skúmajú, spracovávajú, vyhodnocujú a využívajú na získanie poznatkov, identifikáciu vzorov, predikciu budúcich udalostí alebo iných relevantných výstupov. Dataset sa neviaže ku konkrétnemu subjektu evidencie ale spravidla zoskupuje dostupné údaje pre daný fenomén.

4 Návrh nového modelu

Návrh nového modelu definuje spôsob fungovania zdieľania údajov (či už kmeňových alebo agendových) vo verejnej správe v budúcnosti, použitie referenčných údajov a základných registrov, pričom sa sústredíme na:

- výber štandardov a konvencií,
- spôsob modelovania dátových prvkov,
- referenčné údaje vo vzťahu k centrálnemu modelu dátových prvkov.

4.1 Strategické rozhodnutia

Účelom tejto kapitoly je výber prístupu pre zlepšovanie „jeden-krát a dost“, z ohľadom na súčasný stav to znamená najmä: zrýchlenie zdieľania údajov, zlepšenie kvality zdieľaných údajov a podpora automatizácie a strojového spracovania údajov.

Víziou je dôveryhodným spôsobom sprístupniť údaje, ktoré sú vedené v agendových informačných systémoch verejnej správy v okamihu keď sú potrebné, a to v rámci konania alebo riešenie životnej situácie. **Podporí sa tak možnosť automatizácie rozhodovacích procesov získaním potrebných údajov v strojovo-spracovateľnom formáte.**

Jednotlivé témy sú podrobnejšie vysvetlené v kontexte slovenského prostredia e-Governmentu. Identifikované sú základné varianty, ako k problematike pristúpiť a výber vhodného riešenia. Stratégiu je možné chápať ako sériu rozhodnutí a výber prístupov pre jednotlivé témy:

- Právny rámec pre zdieľanie údajov.
- Mechanizmy pre zdieľanie údajov a princíp „jeden-krát a dost“.
- Manažment kmeňových údajov.

4.1.1 Právny rámec pre zdieľanie údajov

Prepojenie a fungovanie základných registrov a zdieľanie údajov potrebuje právny rámec premietnutý do súboru zákonných a podzákonných noriem, ktorými sa následne riadia všetky zúčastnené strany. Stanovenie právneho rámca je dôležité jednak pre právnu istotu, ktorá garantuje predvídateľnosť celého systému a stanovuje záväznú hranicu, ktoré nesmú byť pri využívaní tohto systému.

Vzhľadom na uvedené, ako aj vzhľadom na potrebu zavedenia detailnejších postupov a procesov, definíciu práv a povinností, je potrebné sa v rámci navrhnutia konceptu venovať.

Pre vytvorenie právneho rámca je potrebné zodpovedať nasledujúce kľúčové otázky:

Ako by malo vyzerat' stanovenie celoštátnej správy a riadenia?

V tejto oblasti je prioritou určiť **orgán s celoštátnou pôsobnosťou**, ktorý bude centrálny **zodpovedný za správu a riadenie** digitálnej agendy, ako aj orgán, ktorý bude **kontrolovať dodržiavanie pravidiel** vychádzajúcich z právneho rámca a dohliadať na dohody o zdieľaní dát. Takýmto orgánom je **Dátová kancelária** Ministerstva investícií, regionálneho rozvoja a informatizácie SR.

Z rôznych dôvodov, napríklad pre vyhnutie sa duplicitám, alebo pre využitie doterajších skúseností a vzniku synergií navrhujeme, aby sa centrálny zodpovednou inštitúciou stalo Ministerstvom informatizácie, regionálneho rozvoja a investícií SR a kontrolným orgánom Úrad na ochranu osobných údajov SR, ktorý má vhodné postavenie nezávislého orgánu, a ktorému by sa tak rozšírili kompetencie na kontrolu nakladania so všetkými údajmi. Prípadnou alternatívou je aj vznik nového kontrolného úradu, ktorý by mal na starosti aj monitorovaciu a kontrolnú úlohu vyplývajúcu z aktu o správe údajov².

Tieto aj ďalšie detaily správy a riadenia by mali byť upravené **v novom zákone o údajoch**, ktorý by mal reflektovať celkový koncept a zjednotiť pojmy, postupy, procesy, práva a povinnosti všetkých aktérov pôsobiacich alebo benefitujúcich z tohto systému. Tento zákon by mal tiež určiť hierarchiu všetkých noriem a vzťahov v tejto oblasti.

Ako by mali byť definované použité legislatívne pojmy?

Všetky pojmy už definované v iných všeobecných nástrojoch (t. j. spracovateľ, sprostredkovateľ, dotknutá osoba, osobné údaje, osobitné kategórie údajov atď.) by sa mali odzrkadliť a začleniť do konkrétnych právnych predpisov zameraných na reguláciu zdieľania údajov vrátane pravidiel pre základné registre.

V slovenskom právnom poriadku definované pojmy z tejto oblasti už v súčasnosti neplnia funkciu jasnej a zrozumiteľnej definície a navyše, v prípade prijatia nového konceptu bude musieť dôjsť **k redefinícii pojmov a tiež k vytvoreniu nových**, ktoré budú vychádzať z reálií systému a budú v súlade s ustálenými pojmami na medzinárodnej úrovni (pojmy sú definované v kapitole 2.6 Definície a účel referenčných údajov).

Ako sprístupniť údaje?

Údaje ako verejné aktíva by mali byť prístupné štyrmi spôsobmi:

- ostatným verejným inštitúciám cez Centrálnu integračnú platformu na základe zákonného oprávnenia (ktoré sa bude definovať v Metainformačnom systéme),
- oprávneným osobám, ktorých sa údaje týkajú cez Manažment osobných údajov,
- verejnosti v otvorenom formáte,
- subjektom, ktoré získajú prístup k údajom na základe aktu o správe údajov.

Z pohľadu formy je pre všetky spôsoby sprístupnenia cieľovým stavom použitie prepojených údajov na základe ontológií z Centrálného modelu údajov (RDF v JSON-

² [NARIADENIE EURÓPSKEHO PARLAMENTU A RADY o európskej správe údajov \(akt o správe údajov\)](#)

LD). Takéto údaje bude možné strojovo spracovať a zároveň sémanticky interpretovať ich význam.

Zo sprístupnením otvorených údajov úzko súvisí otázka licencie. S cieľom podporiť distribúciu a opätovné použitie aktív verejného sektora by sa malo podporovať **zverejňovanie špecifikácií s otvoreným zdrojom**, a to na základe uznanej licencie k otvoreným zdrojom. Toto by sa malo prednostne vykonávať bezplatne a nekomerčným spôsobom.

Hlavné príklady licencií s otvoreným zdrojovým kódom zvyčajne spadajú pod jeden z nasledujúcich dvoch typov:

- **Recipročné licencie** – ako napr. verejná licencia Európskej únie (EURL)²⁹ – ktorá bola špeciálne navrhnutá na opätovné použitie, overenie, úpravu a distribúciu softvéru s otvoreným zdrojovým kódom potrebného na sledovanie ďalších distribuovaných vylepšení. Je to preto, že jeho „reciprocita“ znamená, že akýkoľvek derivát vyvinutý opätovným používateľom/treťou stranou sa nestane majetkom takejto tretej strany, ale akýkoľvek zdrojový kód by mal byť tiež sprístupnený a opätovne použiteľný.
- **Povolené softvérové licencie** – ako napr. licencia Apache alebo MIT – ktoré sú na rozdiel od recipročných licencií „tolerantnejšie“ v tom zmysle, že umožňujú vytváranie vlastníctva derivátov generovaných licenciami s otvoreným zdrojom.

V oboch prípadoch je zverejnenie špecifikácií s otvoreným zdrojom pod licenciou prospešné, pretože uprednostňuje **harmonizáciu prístupov** vzhľadom na absenciu nákladov na prístup a možnosť opätovného použitia riešení, ktoré už podobné verejné služby používajú.

Ako by mali byť definované zásady zdieľania?

Pri prepojení základných registrov sa budú brať do úvahy nasledujúce zásady zdieľania:

- **Zásada potreby a proporcionality**, ktorá hovorí, že zdieľanie údajov s iným verejným orgánom musí byť potrebné, užitočné a primerané na vykonávanie verejnej služby, napríklad jeho výsledkom je zníženie administratívnej záťaže, obmedzenie praxe keď jedna inštitúcia žiada o informácie dotknutých osôb, ktoré už vlastní iná inštitúcia – podľa „princípu jeden-krát a dost“).
- **Zásada transparentnosti** môže byť uvedená do praxe napríklad na základe povinnosti predkladať ex-ante písomné dohody o zdieľaní údajov medzi verejnými orgánmi orgánu správy a riadenia na overenie, či je v súlade s podmienkami dohodnutého prístupu a s architektúrou systému, alebo či je otvorená pre potenciálne prístupy viacerých strán. V tejto súvislosti by mala existovať vzorová dohoda, ktorá by mala byť orgánom správy a riadenia riadne uverejnená. Tento nástroj môže tiež špecifikovať, že orgán správy a riadenia vytvorí portál prístupu k osobným údajom, ktorého cieľom bude uľahčiť výkon práva dotknutých subjektov.
- **Zásada vedenia** na základe ktorej orgán správy a riadenia môže po splnení podmienok označiť konkrétnu databázu ako „základný register“ a priradiť ju ako jedinečný zdroj pre kategóriu informácií.

- **Zásada riadenia prostredníctvom dohôd**, v rámci ktorej sa ustanoví vedúci zástupca a následne sa na pravidelnej báze (napríklad každých 5 rokov) monitorujú a kontrolujú dohody o zdieľaní, či sú v súlade so zverejnenou vzorovou dohodou a ostatnými podmienkami.
- **Zásada správy a riadenia na niekoľkých úrovniach**, ktorej obsahom sú nasledovné povinnosti alebo možnosti správy riadenia:
 - každý správca registra je zodpovedný za to, že informácie sú presné, aktuálne a úplné a tiež za to, že sa implementujú, monitorujú a dokumentujú prístupy.
 - Dátová kancelária je zodpovedná za riadenie prostredníctvom dohôd.
 - dohodou o použití údajov je stanovený prevádzkovateľ v zmysle GDPR a zákona o ochrane osobných údajov.
 - existuje orgán správy a riadenia, ktorý bude mať právomoc auditovať alebo kontrolovať, či sú prípadné poplatky za správu prístupu k údajom férové, primerané a nediskriminačné.

4.1.2 Mechanizmy pre princíp „jeden-krát a dost“

Zdieľanie údajov je kľúčové pre zabezpečenie princípu „jeden-krát a dost“ – požiadavku aby pri interakcii s verejnou správou sa od žiadateľa vyžadovali len údaje, ktoré sú nové a verejná správa nimi ešte nedisponuje.

Navrhovaný model počíta s dvoma nástrojmi: referenčné údaje a základné číselníky, ktoré používa celá verejná správa spoločne a zdieľanie kmeňových a agendových údajov na základe dohodnutých pravidiel.

Referencovanie

Referenčné údaje a základné číselníky tvoria kontext pre agendové údaje v informačných systémoch verejnej správy. Vyhlásenie údajov za referenčný dáva verejným inštitúciám jasný postup a časové rámcové pre plnenie princípu 1x a dost vrátane právnej istoty. Zámerom je, aby referenčné údaje boli definované v základných registroch a aby tak tento kontext bol v celej verejnej správe spoločný. V tomto modeli referenčné údaje vyjadrujú:

- základné fakty o subjektoch evidencie Slovenskej republiky (právnické osoby a fyzické osoby),
- základné fakty o území a priestore (meta-údaje o priestore a adresy),
- práva a povinnosti v dátovom priestore (za účelom riadenia zdieľania údajov).
- Referenčné údaje majú slúžiť na synchronizáciu údajov a stotožňovanie subjektov evidencie tak, aby inštitúcia, ktorá referencuje referenčné údaje mala právnu istotu pri ich používaní.

Zdieľanie kmeňových a agendových údajov

Agendové údaje a kmeňové údaje môžu slúžiť na výmenu a zdieľanie údajov (faktov) medzi inštitúciami verejnej správy pre potreby výkonu verejnej moci podľa zákona proti

byrokracii. Zdieľať by sa mali kmeňové a agendové údaje, ktoré potrebuje iná inštitúcia verejnej správy pre svoj výkon kompetencií. Orgán verejnej moci môže pristupovať k dátovému výstupu svojim informačným systémom a následne ho spracovať.

Na to, aby bolo možné kmeňový alebo agendový údaj zdieľať inej inštitúcii, je potrebné zabezpečiť:

- aby bolo jasné, kto je za údaj zodpovedný, do ktorého objektu evidencie je údaj zaradený,
- aby bol vybraný subjekt evidencie jednoznačne identifikovaný (a stotožnený),
- aby bol vybraný objekt evidencie prístupný pre použitie (zvolený vhodný dátový výstup dostupný cez Centrálnu integračnú platformu),
- aby bolo overené oprávnenie pristupovať k dátovému výstupu.

Používaný centrálny systém riadenia referenčných (zdieľaných) údajov verejnej správy obsahuje nástroje pre prístup k referenčným a agendovým údajom vo verejnej správe. Tento systém je súčasťou Centrálny integračnej platformy.

Zdieľané agendové údaje sa nachádzajú:

- vo forme zdrojových údajov (čo sú údaje na strane zdrojových systémov poskytovateľov),
- vo forme odvodených údajov (čo sú údaje na strane systémov konzumentov).

Keď systém konzumenta (užívateľa údajov) pri podpore svojej agendy potrebuje agendové údaje s iných informačných systémov, odporúčaným postupom je konsolidácia informácií do jedného konsolidovaného dátového objektu pre daný účel, ktorý spája vybrané údaje z viacerých zdrojov, pričom dodržiava zásadu minimálnosti. Konsolidovaný dátový objekt by malo byť možné transformovať do podoby prepojených údajov v súlade s príslušnými ontológiami Centrálny modelu údajov.

Pre výber vhodného modelu pre zdieľanie údajov je dôležité primárne zodpovedať na nasledujúce otázky:

Aké nástroje by mali byť použité pre naplnenie princípu „jeden-krát a dosť“?

- **Stotožňovanie referenčných údajov.** Centrálna integračná platforma by mala poskytovať nástroje pre stotožnenie referenčných údajov so svojou evidenciou a upozorňovať na zmeny referenčných údajov.
- **Zdieľanie agendových údajov a kmeňových údajov** (ktoré nie sú referenčné) získaných cez Centrálnu integračnú platformu v konsolidovanej podobe. Ak sa takéto údaje použijú ako podklad pre ďalšie rozhodnutie, je dôležité, aby bola zaznamenaná informácia o hodnote v konkrétnom čase.
- **Zasielanie údajov na základe súhlasu**, teda na základe vyjadrenia vôle subjektu resp. osoby, o ktorej sú údaje spracúvané prostredníctvom MOU.

Aké druhy základných registrov by mali byť implementované?

Ako bolo spomínané vyššie prístup k zriaďovaniu základných registrov zvolený v Českej republike považujeme za najvhodnejšiu cestu, rovnako je to aj pri výbere druhov základných registrov, preto navrhujeme v Slovenskej republike zriadiť nasledovné druhy základných registrov:

- Register fyzických osôb.
- Register právnických osôb.
- Register adries.
- Register priestorových informácií.
- **Register práv a povinností (ako súčasť Metainformačného systému).**

Ktoré údaje môžu byť referenčné?

Referenčnými údajmi budú atribúty zo základných registrov a základné číselníky. Z pohľadu ďalšieho postupu je potrebné:

- Zrušiť vyhlásenie referenčných údajov pre Register daňových subjektov a Informačný systém služieb zamestnanosti. Spomenuté údaje bude možné získavať cez mechanizmus zdieľania agenodových a kmeňových údajov.
- Ako referenčné údaje vyhlásiť celý Register adries (adresné body).
- Vyhlásiť referenčné údaje pre Register priestorových informácií.
- Vyhlásiť referenčné údaje z Meta-informačného systému verejnej správy (pre Centrálny model údajov a práva a povinnosti).

Akým spôsobom by mali byť referenčné údaje vyhlasované?

Súčasný systém vyhlasovania referenčných údajov sa ukázal ako neefektívny.

Súčasne platné a účinné metodické usmernenie vyhlasovania referenčných údajov stanovil procesy, ktoré sa ukázali ako nedostatočné, a to najmä preto že prvotný impulz vo forme podnetu necháva iniciatívu skôr na náhodu a metodické usmernenie nestanovuje jasnú stratégiu a plán, ktoré údaje majú byť referencované a v akom časovom horizonte. Tento problém sa s odstupom času ukazuje ako kľúčový a majúci zásadný dopad na stav referencovania údajov v prostredí Slovenskej republiky.

Fakt, že zákon proti byrokracii presadzujúci princíp „jeden krát a dost“ umožnil ľahšiu cestu, prispel k tomu, že vyhlasovanie referenčných údajov podľa metodického usmernenia sa stalo málo využívanou možnosťou aj napriek tomu, že má svoje nesporné výhody.

Tretím faktorom, ktorý nás núti uvažovať o zmene spôsobu vyhlasovania referenčných údajov je, že procesy stanovené metodickým usmernením sú koncipované spôsobom vhodnejším pre analógový alebo papierový svet a nie pre digitálny.

Vzhľadom k uvedenému navrhujeme, aby sa pre vyhlasovanie údajov použil tzv. projektový prístup na začiatku ktorého by bolo autoritatívne a jasné rozhodnutie o tom:

- Aké základné registre majú byť zriadené.
- Aké referenčné údaje majú základné registre obsahovať.
- Akým spôsobom (procesy) a v akom čase majú byť základné registre referenčnými údajmi naplnené (harmonogram).
- Ktoré orgány sú zodpovedné za manažment údajov v základných registroch, dodržiavanie procesov a splnenie harmonogramu.

V súvislosti s tým navrhujeme, aby sa vyššie uvedené stanovilo zákonom (napríklad novým zákonom o údajoch), podobne ako v Českej republike, pričom hlavným **zodpovedným orgánom by sa mala stať Dátová kancelária** Ministerstva investícií, regionálneho rozvoja a informatizácie SR, ktorá by sa mala tiež stať orgánom správy a riadenia celej digitálnej agendy.

4.1.3 Manažment kmeňových údajov

Digitalizácia a štruktúrovanie informácií základných registrov a agendových informačných systémov verejnej správy je kľúčom k uľahčeniu interoperability a maximalizácii dostupnosti verejných údajov a ich opätovného použitia. Vytvorené štruktúrované údaje je potrebné správne spravovať a manažovať, aby sa zabezpečila ich kvalita, platnosť, autenticita a v konečnom dôsledku aj dostupnosť.

Manažment kmeňových údajov (Master Data Management - MDM) predstavuje „jednotný“ prístup a riešenie problémov, ktorým čelia základné registre a služby verejnej správy, ktoré s nimi musia spolupracovať. MDM je zameraný na konzistentnosť a kvalitu kmeňových dát a zahŕňa všetky procesy a nástroje, ktoré sú k tomu funkčné.

Správa kmeňových údajov je efektívny jednotný prístup k riadeniu životného cyklu údajov vo verejnej správe. Rieši výzvy, ktorým čelia základné registre a služby, ktoré s nimi musia spolupracovať. Smernice MDM nepokrývajú len sémantické aspekty, ale aj technické opatrenia, keďže tieto dve dimenzie spolu úzko súvisia. Nástroje MDM, ktoré sú zvyčajne proprietárnym komerčným vývojom, môžu pomôcť s mnohými úlohami súvisiacimi s údajmi. Pred pokusom o vývoj vlastných riešení MDM sa odporúča preskúmať existujúce postupy a riešenia a posúdiť, ako ich môžu opätovne použiť a aplikovať orgány verejnej správy.

Nástroje kmeňových údajov (Master Data Tools)

Nástroje kmeňových údajov sú proprietárne komerčné nástroje užitočné na správu kmeňových údajov. Normálne pomáhajú vytvoriť jednotný pohľad na hlavnú entitu pre organizáciu (napr. verejnú správu) naprieč všetkými prevádzkovými a analytickými použitiami a nezávisle od akéhokoľvek iného úložiska podobných údajov.

Tri z najväčších výziev, ktoré MDM rieši, sú teda:

- Správa údajov (Data Governance) – Riadenie údajov organizácie je vytvorenie a presadzovanie zásad a postupov pre obchodné a technické zaobchádzanie s údajmi, na ktoré zvyčajne dohliada výkonná rada alebo výbor. Jeho pokrytie môže siahť od údajov týkajúcich sa jedinej aplikácie až po všetky údaje organizácie.
- Správcovstvo údajov (Data Stewardship) – Správca údajov je zvyčajne manažér, ktorý chápe, ako údaje ovplyvňujú výkon spoločnosti alebo konkrétnej jednotky. V

rámci svojich povinností spolupracujú s odborníkmi na správu údajov, aby viedli snahu o správu hlavných údajov s cieľom splniť obchodné ciele a ciele.

- Kvalita údajov (Data Quality) – Kvalita údajov je súborom súvisiacich postupov správy údajov a procesov podnikovej kvality vytvorených na zaručenie presnosti údajov, aktuálnosti a relevantnosti pre určený cieľ. Najbežnejšie používané techniky kvality údajov sú čistenie údajov a štandardizácia údajov; okrem toho sa využíva overovanie, profilovanie, monitorovanie, porovnávanie, zlučovanie, geokódovanie a obohacovanie údajov.

Pre zavedenie princípov manažmentu kmeňových údajov je potrebné zodpovedať nasledujúce kľúčové otázky:

Aké mechanizmy je potrebné zvoliť pre sémantickú interoperabilitu údajov?

- **Používanie Centrálného modelu údajov** (spoločných slovníkov a ontologického modelu, metadát, číselníkov a podobne) v rámci celej verejnej správy, aby si informačné systémy verejnej správy medzi sebou „rozumeli“ na úrovni sémantickej interoperability a strojového-spracovania údajov.
- **Zavedenie princípov „správy kmeňových údajov“** (Master Data Managementu), aby sa naprieč informačným prostredím používali rovnaké hodnoty pre identifikované kmeňové záznamy, a aby sa tiež používali len:
 - kompletne údaje, teda všetky dostupné informácie existujúce vo verejnej správe,
 - časovo aktuálne údaje, teda vyjadrujúce poslednú platnú informáciu zo zdroja dát,
 - preukázateľné a konzistentné údaje.

Aký štýl manažmentu kmeňových údajov by mal byť v rámci verejnej správy použitý?

Existujú rôzne štýly MDM, z ktorých je možné vybrať ten správny pre konkrétny systém. Voľba závisí od toho, či je potrebné mať na správu svojich údajov nejaký centrálny bod alebo je potrebné dáta synchronizovať s existujúcimi zdrojmi.

Kľúčovými bodmi záujmu pri vytváraní modelu by sa mali týkať správy údajov, zvyšovania kvality údajov a zabezpečení toho, aby sa údaje dali ľahko spravovať a sprístupniť. Nasledujúca tabuľka sumarizuje štyri najčastejšie používané štýly manažmentu kmeňových údajov:

Registrový štýl

V tomto prístupe je základným konceptom centralizovaný register, ktorý slúži ako jednotný zdroj pre ukladanie a správu kritickej množiny kmeňových údajov.

Ak by sme zavádzali registrový štýl, vytvoril by sa pre všetky základné registre jednotný register centrálny úložný priestor, ktorý by využívali všetky ostatné informačné systémy verejnej správy. Tento register by obsahoval obsahuje jedinečné záznamy pre jednotlivé subjekty evidencie, ako napríklad fyzické osoby, právnické osoby, ale aj pre vybrané

objekty evidencie. Každý záznam subjektu evidencie obsahuje základné údaje a atribúty prislúchajúce danej entite.

Hlavným cieľom registrového štýlu je dosiahnuť jednoznačnosť a kvalitu kmeňových dát prostredníctvom centrálného registra. V tomto prístupe je dôležitá aj štruktúra údajov a definície jednotlivých entít. Kmeňové dáta sú v registri uchovávané a spravované tak, aby sa minimalizovali duplikáty, chybné údaje a nesúlad medzi rôznymi zdrojmi dát.

Pri použití registrového štýlu je kľúčové zabezpečiť aktualizáciu a správu kmeňových v centrálny registračnej databáze. Správca kmeňových údajov musí byť zodpovedný aj za kvalitu kmeňových údajov. Súčasne sa na centrálnej úrovni implementujú špecifické procesy, ako napríklad kontroly údajov, schvaľovacie mechanizmy a sledovanie zmien.

Pred zavedením centrálnej správy referenčných údajov sa na Slovensku používal hybridný variant registrového štýlu, keď základné registre boli spravované decentralizovane a jednotlivé informačné systémy verejnej správy mohli pristupovať k údajom zo základného registra na základe vlastného uváženia prostredníctvom API.

Konsolidačný štýl

Konsolidačný štýl je prístup, ktorý sa zameriava na zosúladenie a konsolidáciu kmeňových dát z viacerých zdrojov do centrálného úložiska. Hlavným cieľom konsolidačného štýlu je dosiahnuť jednoznačnosť a kvalitu kmeňových dát prostredníctvom procesu zlučovania, transformovania a spravovania dát z rôznych zdrojov.

V konsolidačnom štýle sa využívajú rôzne mechanizmy a nástroje na integráciu a transformáciu dát, aby sa zabezpečilo, že kmeňové dáta sú jednotné a konzistentné. Typicky sa používajú techniky ako zlučovanie záznamov, deduplikácia, normalizácia údajov a mapovanie dát, aby sa dosiahol jednotný formát a štruktúra kmeňových údajov.

Proces konsolidácie zahŕňa identifikáciu a zosúladenie duplicitných záznamov z rôznych zdrojov dát, ako aj vyriešenie nesúladov a rozdielov medzi údajmi. Konsolidačný štýl sa zameriava na zabezpečenie, že výsledné kmeňové údaje sú presné, aktuálne a použiteľné pre ďalšie procesy a aplikácie.

V konsolidačnom štýle je dôležité mať dobre definované pravidlá a postupy na zlučovanie a transformáciu dát. Taktiež je potrebné zabezpečiť dostatočnú kvalitu dát, kontrolu chýb a sledovanie zmien v dátach. Správa konsolidovaných kmeňových dát je zvyčajne centrálny riadená a spravovaná, aby sa zabezpečila konzistencia a integrita údajov.

Pôvodný koncept správy referenčných údajov a platforma CIP bola postavená na konsolidačnom štýle. V praxi sa však nepoužívali všetky funkcionality a riešenie sa postupne používalo na výmenu správ medzi informačnými systémami verejnej správy v podobe, v akej ich poskytuje zdrojový systém alebo register.

Riešenie pre manažment osobných údajov (MOU) už využíva centrálny nastavené pravidlá pre transformáciu a zlučovanie údajov, pričom sa využíva Centrálny model údajov a na výstupe s pracuje s jednotným formátom pre prepojené dáta JSON-LD. V budúcnosti bude možné rovnaké mechanizmy (transformačný modul) využívať pre prístup ku konsolidovaným dátovým produktom aj pre konzumentov z verejnej správy.

Koexistenčný štýl

Koexistenčný štýl je prístup, ktorý sa zameriava na správu kmeňových dát v prostredí, kde existuje viacero nezávislých systémov, ktoré udržiavajú vlastné verzie kmeňových dát. Verejná správa s obrovským množstvom inštitúcií má prirodzené tendencie smerovať k takémuto prostrediu. V tomto prístupe nie je cieľom zosúladiť a centralizovať dáta do jedného úložiska, ale skôr umožniť koexistenciu a súlad medzi rôznymi verziami kmeňových dát v rôznych systémoch.

V koexistenčnom štýle je základnou myšlienkou ponechať existujúce systémy a ich vlastné zdroje kmeňových dát nedotknuté a spravovať kmeňové dáta ako sémantické vrstvy, ktoré poskytujú zjednotený pohľad na dáta pre rôzne aplikácie a procesy. Tieto sémantické vrstvy môžu byť vytvorené pomocou metadát, ontológií, pravidiel a služieb na poskytovanie jednotného a konzistentného pohľadu na kmeňové údaje. Takéto služby sú už dnes dostupné v rámci CIP a Centrálného modelu údajov. CIP tiež v súčasnosti ponúka službu pre distribúciu údajov, ktorá dokáže notifikovať počúvajúce systémy konzumentov o zmenách vybraných kmeňových údajov v reálnom čase.

Koexistenčný štýl umožňuje zachovať autonómiu a nezávislosť jednotlivých systémov, pričom zabezpečuje koordináciu a konzistenciu medzi nimi. Sémantické vrstvy umožňujú harmonizáciu údajov a poskytujú jednotný model dát pre rôzne aplikácie a procesy, čím sa minimalizuje duplicita a nesúlad v kmeňových dátach.

Pri implementácii koexistenčného štýlu je dôležité mať dobre definované metadáta a metadátový model, ktoré opisujú štruktúru a vzťahy medzi kmeňovými dátami v rôznych systémoch, čo má za úlohu práve Centrálny model údajov. Taktiež je potrebné zabezpečiť interoperabilitu a možnosť vzájomnej komunikácie medzi systémami, aby sa zabezpečila konzistencia a aktuálnosť dát.

Transakčno-centralizačný štýl

Transakčno-centralizačný štýl je prístup, ktorý kombinuje prvky transakčného a centralizačného prístupu. V tomto štýle sa dôraz kladie na centralizované úložisko kmeňových dát, kde sa spravujú a udržiavajú dáta vo forme transakcií.

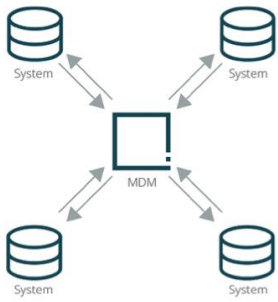
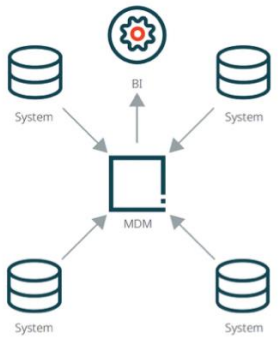
V transakčno-centralizačnom štýle sú všetky operácie na kmeňových dátach realizované cez transakcie, čo zaručuje konzistentnosť a integritu dát. Zmeny a aktualizácie kmeňových dát sa vykonávajú prostredníctvom transakčných procesov, ktoré zabezpečujú správne usporiadanie a atomickosť zmien. Tieto transakcie sú vykonávané v centrálnej platforme, ktorá je zodpovedná za správu a synchronizáciu dát v rôznych systémoch.

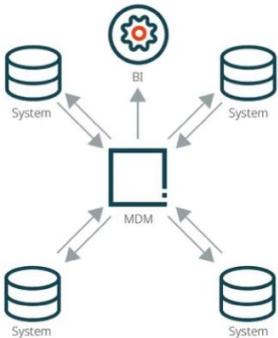
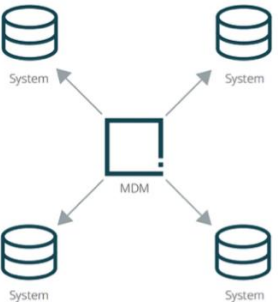
V tomto štýle sa zabezpečuje, aby sa všetky systémy a aplikácie, ktoré pracujú s kmeňovými dátami, obracali na centrálné úložisko, aby získali aktuálnu verziu dát. Tým sa minimalizuje duplicita a rozdiely v dátach a zabezpečuje sa konzistentný a jednotný pohľad na kmeňové dáta pre všetky systémy.

Implementácia transakčno-centralizačného štýlu vyžaduje vytvorenie centrálnej platformy pre správu kmeňových dát, ktorá bude zodpovedná za správu a spracovanie transakcií s kmeňovými dátami. Je tiež potrebné mať definované pravidlá a procesy na správu zmien a aktualizácií dát, aby sa zabezpečila konzistencia a integrita kmeňových dát. Prakticky to znamená, že centrálna platforma zasahovala do fungovania

jednotlivých informačných systémov verejnej správy. Nasadenie takéhoto mechanizmu v podmienkach verejnej správy na Slovensku by bolo mimoriadne komplikované až nerealizovateľné.

Tabuľka 4: Porovnanie výhod a nevýhod rôznych štýlov MDM

Štýl MDM	Výhody	Nevýhody
<p>Registrový štýl</p> 	<p>Výhodou tohto štýlu je, že rieši zložitosti, ktoré súvisia s veľkým počtom zdrojových systémov, z ktorých každý má svoje vlastné pravidlá, ide o nízkonákladový / jednoduchší spôsob implementácie MDM.</p>	<p>Nevýhodou tohto štýlu je, že údaje zo zdrojových systémov sa nemenia (neaktualizujú sa), ťažkosti určiť dôveryhodný zdroj a latencia údajov.</p>
<p>Konsolidačný štýl</p> 	<p>Výhodou konsolidačného štýlu pre MDM je zlepšenie kvality dát a eliminácia duplicitných a nekonzistentných záznamov. Centrálne konsolidácia kmeňových dát umožňuje lepšiu kontrolu a správu dát, a tým zvyšuje ich hodnotu a využiteľnosť pre organizáciu.</p>	<p>Nevýhody sú totožné ako u registrového štýlu.</p>

Štýl MDM	Výhody	Nevýhody
<p>Koexistenčný štýl</p> 	<p>Výhodou koexistenčného štýlu je schopnosť zachovať existujúce systémy a ich vlastné zdroje dát, čo môže byť vhodné v prípade organizácií s komplexným ekosystémom systémov a procesov. Koexistenčný štýl umožňuje flexibilitu a agilitu v správe kmeňových dát, zatiaľ čo zabezpečuje konzistenciu a súlad v rámci organizácie.</p> <p>Synchronizácia medzi MDM hubom a zdrojmi v reálnom čase odoslaním zlatého záznamu do každého príslušného zdrojového systému, výrazné zlepšenie kvality kmeňových dát aktualizáciou kmeňových dát v zdrojových systémoch a MDM hube a možnosť postupného rozširovania (postupne pridávať ďalšie domény).</p>	<p>Nevýhody sú takisto rovnaké ako u predchádzajúcich štýlov a navyše sa u tohto štýlu môžu vyskytnúť problémy so synchronizáciou a takisto fakt, že nasadenie tohto štýlu je finančne náročnejšie ako pri ostatných štýloch MDM.</p>
<p>Transakčný - centralizovaný štýl</p> 	<p>Výhodou je, že centrálny komponent/systém sa stáva jediným poskytovateľom kmeňových dát (funguje ako druh úložiska), aktualizácie sa vykonávajú priamo v centrálnom systéme a následne sa distribuujú do ostatných – žiadne systémy mimo centrálného systému už nemôžu vytvárať alebo upravovať kmeňové dáta.</p> <p>Transakčno-centralizačný štýl prináša výhody, ako je zlepšená konzistencia dát, lepšia sledovateľnosť zmien a jednotná správa kmeňových dát. Centrálna správa dát umožňuje aj lepšiu kontrolu a riadenie prístupu k dátam, čo prispieva k zvýšenej bezpečnosti.</p>	<p>Nevýhodami sú náročnosť na zdroje a potenciálne náklady a čas potrebný na implementáciu tohto štýlu.</p>

Na záver je potrebné zdôrazniť, že je možné a v praxi sa to deje veľmi často, že sa prijímú rôzne štýly alebo kombinácie štýlov na základe lokálnych potrieb a právnych požiadaviek.

Slovensko tak bude implementovať konsolidačný štýl pre manažment osobných údajov a koexistenčný štýl pre správu kmeňových údajov.

Aké identifikátory by mali byť použité?

Identifikátor uľahčuje prístup ku všetkým údajom, ktoré sa týkajú príslušnej entity. Je dôležité rozlišovať medzi „jedinečnou“ a „kontextovo špecifickou“ identifikáciou (alebo základnými údajmi):

- **Jedinečná identifikácia** – je univerzálna a týka sa multikontextových údajov; dá sa použiť v každej krajine bez ohľadu na kontext. Príkladom je zoznam poznávacích značiek vozidiel v Základnom registri vozidiel alebo zoznam čísiel občianskych preukazov v Základnom registri civilných osôb.
- **Špecifická identifikácia** – týka sa údajov, ktoré sú prispôbené kontextu jednej krajiny alebo služby, ale nie nevyhnutne inej krajiny alebo služby. To je napríklad prípad stredných mien (alebo druhých priezvisk) v registri civilného obyvateľstva v Španielsku.

Preto by sa mali vytvárať alebo uchovávať univerzálne a jedinečné identifikátory (UUID). Toto je obzvlášť dôležité pre cezhraničné a medzisektorové iniciatívy, pretože kmeňové údaje – aj keď je to nepravdepodobné – by mohli mať rovnaké identifikátory v rôznych základných registroch.

Vytvorenie nového systému UUID môže mať dopad na základné registre, preto sa odporúča – pred vynájdением nového identifikačného systému – vyhľadať existujúce a pokúsiť sa ich opätovne použiť. V každom prípade by orgány verejnej správy pred prijatím existujúceho riešenia alebo vyvinutím nového mali posúdiť vplyv, ktorý by to mohlo mať na systémy zainteresovaných strán, a prípadne sa pokúsiť ho minimalizovať.

Nižšie sú uvedené základné odporúčania týkajúce sa používania identifikátorov:

- Pri navrhovaní modelov výmenných údajov je potrebné zväžiť použitie viacerých identifikátorov pre jednu inštanciu objektu. To je vhodné pre medzisektorové iniciatívy, kde rôzne orgány môžu identifikovať tú istú inštanciu rôznymi spôsobmi, a tiež to umožňuje automatizovanú produkciu mapovania identifikátorov.
- Identifikátor by sa nikdy nemal meniť v priebehu času, keď už bol pridelený jednej konkrétnej entite. Jedným zo spôsobov, ako zabezpečiť dlhodobú existenciu identifikátora, je priradiť mu URI, čo tiež uľahčuje popis identifikovanej entity.

Výzva jedinečnej (a univerzálnej) identifikácie entít je spoločná pre mnoho rôznych obchodných domén, medzi nimi aj základné registre. Na úrovni EÚ prišiel projekt BRIS s riešením: EUID (jedinečné európske ID pre firmy). Toto riešenie bolo inšpirované výskumným projektom¹⁰⁴ financovaným EK a je dosť podobné riešeniu IBAN prijatému na jedinečnú identifikáciu bankových účtov.

Jedným zo spôsobov, ako zabezpečiť dlhodobé prežitie identifikátorov, je priradiť ich k trvalým jednotným identifikátorom zdrojov (PURI). SEMIC odporúča používanie PURI¹⁰⁵ a definuje princípy a postupy ich používania. Jedným zaujímavým príkladom je iniciatíva EULF¹⁰⁶, kde sa adresy PURL používajú na identifikáciu miest, overenie, že používatelia sú tí, za ktorých sa vydávajú, a že sú oprávnení používať požadované informácie alebo funkcie.

4.2 Prepojenie s manažmentom údajov

Zavedenie zdieľania údajov a správy kmeňových údajov si vyžaduje úzke prepojenie s ďalšími aktivitami v rámci programu manažment údajov:

- Kvalita údajov: aby bolo zmysluplné údaje zdieľať, je potrebné neustále zvyšovať ich kvalitu.
- Klasifikácia a kategorizácia údajov: aby bola jasná dátová architektúra vo verejnej správe, je potrebné vedieť, kto a aké údaje spravuje.
- Integrácia údajov: funkčná platforma pre dátovú integráciu.

4.2.1 Zdieľané údaje musia byť kvalitné

Kvalita údajov má priamy dopad na možnosti využitia údajov v oblastiach a ovplyvňuje kvalitu rozhodovania vo verejnej správe, spoľahlivosť navrhovaných politik a regulácií vo verejnej správe, jednoduchosť a komfort služieb pre podnikateľov a občanov, ako aj na platnosť prípadného vedeckého výskumu, ktorý by sa vykonával použitím údajov verejnej správy.

Systematické meranie kvality sa vykonáva s využitím kritérií ako presnosť, kompletnosť, aktuálnosť, unikátnosť, referenčná integrita, strojová spracovateľnosť, správnosť a podobne. Kvalitu údajov je potrebné tiež chápať v dvoch rozmeroch:

- celková kvalita údajov v informačnom prostredí verejnej správy (vrátane možnosti vynucovania zverejňovania dát príslušnými inštitúciami verejnej správy)
- kvalita jednotlivých údajov evidovaných v jednotlivých informačných systémoch.

Pre splnenie kritérií kvality je potrebné realizovať procesy riadenia a správy dátovej kvality: okrem samotného sprístupnenia údajov, bude potrebné aj zrozumiteľne zdokumentovať dátové štruktúry, proces tvorby dát, štatistické metodológie (ak boli použité), dátové zdroje, kontext a mnoho ďalších aspektov. Riadenie kvality koordinuje Dátová kancelária verejnej správy, ale samotné zlepšovanie kvality údajov je potrebné riešiť lokálne. Metodický postup pre riadenie dátovej kvality je definovaný v dokumente *1.1.1 Štandardizácia dátovej kvality* a praktické návody pre riadenie kvality sú pripravené v rámci dokumentov *1.2.1 Opatrenia pre zvýšenie dátovej kvality*.

4.2.2 Je potrebné vedieť, kto a aké údaje spravuje

Aby bolo možné nastaviť jednotné pravidlá pre prácu s údajmi, je potrebné vychádzať z jasnej klasifikácie. Klasifikácia údajov umožňuje údaje kategorizovať podľa vybraných kritérií tak, aby bolo možné nastaviť efektívne a účelné pravidlá pre prácu s takýmito údajmi a aby bol pokrytý potrebný rozsah údajov.

V prostredí verejnej správy navrhujeme jednotnú kategorizáciu, ktorá rozlišuje podľa charakteru údajov, podľa vlastníctva údajov, podľa pôvodu údajov, podľa zdroja, podľa typu a formy, podľa úrovne strojového spracovania a podľa typu interoperability zverejnených dát.

Klasifikácia údajov by mala byť dostupná na centrálnom mieste, ako súčasť mapovania dátových zdrojov jednotlivých verejných inštitúcií, pričom údaje budú dostupné v rámci Metainformačného systému verejnej správy.

Klasifikácia a kategorizácia údajov je podrobne vysvetlená v dokumente 1.1.8 *Štandardizácia dôveryhodných údajov*.

4.2.3 Funkčná platforma pre integráciu údajov

Integráciu údajov je možné chápať ako prepojenie agendových systémov, ich sekundárnych registrov a evidencií na úrovni údajov. Ide o integráciu na úrovni jednotlivých dátových úložísk, synchronizáciu, aktualizáciu a referencovanie. **Cieľom integrácie údajov je v čo najväčšej miere uľahčiť tok dát a informácií.** Preto je na začiatku nevyhnutné identifikovať požadované dátové toky a následne vytvoriť nástroje pre referencovanie údajov, synchronizáciu údajov, riešenie konfliktov a podobne.

Vďaka integráciám, si môžu inštitúcie navzájom poskytovať údaje o subjekte, ktoré vlastní a spravuje iný orgán verejnej správy, a to bez potreby interakcie s týmto subjektom. Pri tom, je však nevyhnutné zabezpečiť nasledovné:

- mal by byť vytvorený jednotný dátový model údajov verejnej správy (ontológie) a používané URI identifikátory
- pre každý proces musí existovať biznis kontext (zoznam typov dátových objektov z iných inštitúcií verejnej správy, ku ktorým je možné priamo pristupovať) a na základe tohto kontextu by mali byť nastavené oprávnenia pre prístup k údajom
- biznis kontext by mal byť centrálné spravovaný
- subjekt bude vždy informovaný o výmene s ním súvisiacich údajov medzi orgánmi verejnej moci
- vo vybraných situáciách (v súlade so zákonom) by mal klient dávať súhlas so zdieľaním jeho údajov
- mala by fungovať automatizácia zdieľania údajov medzi jednotlivými informačnými systémami verejnej správy a odstránenie manuálnych činností pracovníkov verejnej správy (prepísovanie údajov z jednej aplikácie do druhej), ktoré môžu byť zautomatizované.

Integrácia údajov bude riadená centrálné a definovaná v dátovej architektúre. Bez integrácie údajov a architektúry by nebolo možné zabezpečiť požadovanú kvalitu údajov a neodstránili by sa duplicity a nekonzistentnosti údajov.

Následne je kľúčové zabezpečiť robustné **dátové integračné väzby** pre platformy Konsolidovanej analytickej vrstvy (KAV) a Manažmentu osobných údajov (MOU).

Dátová integrácia je proces získavania údajov z viacerých zdrojov a ich kombinovanie s cieľom vytvoriť jeden jednotný pohľad. Tieto platformy vyžadujú objekty evidencie v dlhom časovom období (spravidla v celom rozsahu histórie), ktoré sa dajú stiahnuť prostredníctvom procedúr ETL alebo ELT pre všetky právnické a fyzické osoby so pseudonymizáciou pre KAV a prostredníctvom procedúr ETL pre vybranú právnickú alebo fyzickú osobu pre MOU. Zároveň tieto platformy potrebujú získavať **zmenové hlásenia/dávky**, ak sa niečo v už stiahnutých objektoch evidencie zmenilo,

alebo ak vznikli nové objekty evidencie pre danú fyzickú či právnickú osobu. Riešenia ETL/ELT využívajú dátový integračný vzorec - agregáciu - na spojenie údajov z rôznych zdrojov do jedného miesta, napríklad do dátového skladu.

Agregácia je proces spájania údajov z mnohých systémov do jedného cieľového systému. Tento vzor poskytuje centralizovaný pohľad na údaje v reálnom čase z viacerých systémov a zabezpečuje, aby sa údaje neduplikovali. Konsolidovaný charakter agregovaných údajov je optimálny na analýzu údajov. Agregovaný vzor môže byť cenný pri vytváraní rozhraní API na získavanie údajov z viacerých systémov a ich poskytovanie v jednej odpovedi.

Integrácia údajov je dôležitou technologickou podmienkou pre zdieľanie údajov. Konceptuálne je definovaná v rámci dokumentu 1.1.6 Štandardizácia dátovej transformácie.

5 Odporúčania a usmernenia

5.1 Odporúčania pre Dátovú kanceláriu

5.1.1 Zosúladiť definíciu pojmov a procesov s technológiami a nástrojmi manažmentu údajov („data governance“)

Referenčné údaje majú vplyv na každodennú činnosť agendových aj technických používateľov informačných systémov. Na optimalizáciu používania a dostupnosti údajov musia organizácie vedieť, čo sú referenčné údaje, a čo nie sú (nie sú to kmeňové údaje („master data“)), prečo sú dôležité a ako ich efektívne spravovať pomocou technológie. Referenčné údaje zvyčajne pozostávajú z malého, diskretného súboru hodnôt, ktoré sa neaktualizujú ako súčasť biznis transakcií, ale zvyčajne sa používajú na zavedenie konzistentnej klasifikácie a charakterizácie údajov. Referenčné údaje majú zvyčajne nízku frekvenciu aktualizácie. Referenčné údaje sú relevantné vo viacerých IS VS patriacich rôznym organizáciám a sektorom. Malcolm Chisholm, svetovo uznávaný líder v oblasti správy údajov, vo svojej knihe *Managing Reference Data in Enterprise Databases* (Správa referenčných údajov v podnikových databázach) definuje referenčné údaje ako "akékoľvek údaje používané výlučne na kategorizáciu iných údajov nachádzajúcich sa v databáze alebo výlučne na prepojenie údajov v databáze s informáciami mimo hraníc organizácie".

Referenčné údaje majú význam. Stanovujú prípustné hodnoty, uľahčujú konzistenciu a mapujú interné údaje s externými údajmi a/alebo štandardmi. Kategorizujú údaje do skupín predtým, ako ich konzumenti údajov analyzujú. Hoci predstavujú malý podiel na celkovom objeme údajov, referenčné údaje predstavujú 25 % až 50 % tabuliek v databázach a ovplyvňujú presnosť vykazovania, analýz a manažmentu údajov.

Referenčné údaje sú:

- súbory hodnôt ako základné číselníky a adresné body v registri adres,
- alebo klasifikačné schémy,

na ktoré sa odvolávajú informačné systémy, aplikácie, dátové úložiská, procesy a reporty, ako aj transakčné a kmeňové záznamy.

Kmeňové údaje opisujú fyzické a právnické osoby, miesta, objekty ako napríklad vozidlá a financie zapojené do agendy organizácie. Organizácie používajú kmeňové údaje na uplatňovanie pravidiel kvality, správu svojich údajov podľa transakčnej štruktúry a „enterprise“ štruktúry na vytvorenie jedného „zlatého záznamu“ („golden record“). Tiež tu treba odlišovať, kedy je takýto „zlatý záznam“ potrebný len v rámci jednej agendy a kedy je skutočne potrebný medzirezortne a na aké prípady použitia, čo ovplyvnia rozsah dátových prvkov, ktoré budú jeho súčasťou. Technológie použité na vytváranie kmeňových údajov musia zabezpečiť, aby tieto dáta boli:

1. **Konsolidované** - čo znamená, že kmeňové údaje spájajú údaje z rôznych zdrojov do jednotného pohľadu.
2. **Očistené** – teda je pri nich extra dôraz na presnosť a kvalitu bez chýb a nadbytočností.
3. **Komplexné** - kmeňové údaje zahŕňajú všetky potrebné podrobnosti, aby poskytli úplný obraz o subjekte, ktorý reprezentujú.

© yyyy Copyright owned by one or more of the KPMG International entities. KPMG International entities provide no services to clients. All rights reserved.

Hoci referenčné údaje možno vnímať ako podmnožinu kmeňových údajov, zásadný rozdiel je v tom, že **referenčné údaje neobsahujú osobné údaje**. Pritom kmeňové údaje napríklad o fyzických osobách, vozidlách alebo pozemkoch spravidla obsahujú citlivé osobné údaje, preto s nimi treba úplne inak nakladať a **zabezpečiť ich dôslednú pseudonymizáciu** (viď dokument 1.1.5 Štandardizácia anonymizácie údajov). Aj vzhľadom na tri charakteristiky definované vyššie je potrebné minimalizovať rozsah informácií obsiahnutých v medzirezortnom „zlatom zázname“, ktorý treba opodstatniť prípadmi použitia zvyšujúcimi efektívnosťou a hospodárnosťou verejnej správy, pretože zabezpečiť tieto tri charakteristiky a zároveň bezpečnosť zlatých záznamov si vyžaduje vysoké náklady na technológie a ľudské zdroje.

Zásadný rozdiel v procese vytvárania kmeňových záznamov je v tom, že master data management spravidla začína zhromažďovaním údajov z rôznych zdrojov, ako sú databázy, tabuľky a informačné systémy verejnej správy. Pritom aktuálne sa základné registre spravujú ako samostatné informačné systémy, ktoré síce ponúkajú dáta na integráciu, ale nesynchronizujú si potenciálne aktuálnejšie informácie z iných informačných systémov. Aj toto je však cesta, kedy dáta a záznamy sú k dispozícii pre iné OVM buď k nahliadnutiu alebo k dátovej integrácii s tým, že sa ale negarantujú tri charakteristiky kmeňových údajov uvedené vyššie. K odhaleniu prípadných nezrovnalostí následne vedú iné procesy, ako napríklad biznis logika systému na strane konzumenta alebo nahlásenie chyby od občana či podnikateľa.

Odporúčame teda aktualizovať vyhlasovanie údajov ako skutočne kmeňových, alebo len údajov pripravených na dátovú integráciu či nahliadnutie v rámci procesov vybavovania agendy.

5.1.2 Posilniť význam manažmentu metadát cez Centrálny model údajov

Metadáta, často označované ako „údaje o údajoch“, poskytujú kontext, význam a použiteľnosť iným typom údajov. Môžu obsahovať informácie o tom, kedy a kým bol konkrétny súbor údajov vytvorený a kto k nemu môže pristupovať. Metadáta môžu obsahovať aj podrobnosti o formátoch a štandardoch údajov. Metadáta v podstate slúžia ako návod na pochopenie údajov a prácu s nimi. Sú kľúčové pre činnosti manažmentu údajov, ako je integrácia, správa a riadenie.

Centrálny model údajov, ako bolo uvedené v dokumente 1.1.2 Štandardizácia pre modelovanie údajov, je založený na štandarde RDF a ontológiach. Ak však aj údaje, ktoré majú byť zdieľané v rámci verejnej správy, nie sú transformované do tohto štandardu, stále je veľmi užitočné zosúladiť (stotožniť) jednotlivé dátové prvky s metadátami ontológií³, ak je to logicky možné. Cieľom sémantiky, a teda ontológií je vytvoriť spoločný význam dátových prvkov pre všetky orgány verejnej správy prostredníctvom prispôsobenia a implementácie spoločných dátových modelov. Na zdokumentovanie významu jednotlivých pojmov a definovanie a rozlišovanie medzi typmi metaúdajov je dôležité používať čo najviac metaúdajov. Takto je možné odstrániť nejednoznačnosť významu poskytnutím spoločnej terminológie a slovníka pre každý pojem či dátový prvok. Celkovo sa odporúča opätovne použiť existujúce sémantické štandardy a odkazovať na ne pri postupe vytvárania aplikačných profilov. Aplikačný profil je špecifikácia údajov na uľahčenie výmeny údajov v presne definovanom aplikačnom kontexte aplikácie alebo informačného systému. Opätovne využíva koncepty z jednej

³ Zdroj: <https://znanosti.gov.sk/metadata>, Dátum referencie: 28.06. 2023

alebo viacerých sémantických špecifikácií údajov (ontológií), pričom pridáva viac špecifikým, že identifikuje povinné, odporúčané a voliteľné prvky, rieši aplikačné potreby a odporúčania pre slovníky, ktoré sa majú používať.

Aktuálne manažment metadát v rámci verejnej správy komplikuje aj to, že spoločne existujú dva svety: svet XML údajov a XSD schém a svet RDF údajov podľa Centrálného modelu údajov (dokument 1.1.2 Štandardizácia pre modelovanie údajov). Tiež pre tieto dva svety existujú rôzne sady nástrojov, ktoré nie sú navzájom previazané a kompatibilné. Preto pri nasledovaní nižšie uvedených krokov pre manažment metadát je potrebné udržiavať oba svety v čo najväčšej možnej kompatibilitate cez dátové schémy XSD vychádzajúce z Centrálného modelu údajov a cez mapovacie tabuľky medzi schémami XML a RDF či SKOS:

1. Krok: Pri vývoji dátových modelov nielen kmeňových a referenčných údajov by sa mali v čo najväčšej miere opätovne používať štandardné schémy a slovníky; ak sa používajú lokálne schémy a slovníky, čo najviac ich treba mapovať na štandardné prvky.
2. Krok: Vo všetkých prípadoch je potrebné určiť, v ktorom súbore je potrebné dátový prvok pridať, zmeniť alebo vymazať. Ak sú dátové štandardy súčasťou definície schémy XML, ide o XSD schému, ktorú treba zmeniť; ak sú dátové štandardy spravované ako schéma RDF, ide o menný priestor RDF; ak ide o riadený slovník vyjadrený pomocou SKOS, ide o koncepcnú schému SKOS.
3. Krok: Pre dátový prvok v modeli XSD je potrebné definovať jeho názov a štruktúrnu definíciu. Na to, aby bol kód zahrnutý do zoznamu kódov na báze XSD, je potrebné definovať jeho názov, atribúty a umiestnenie v hierarchii. Pre prvky (trieda, vlastnosť) v schéme RDF a pre pojem v schéme pojmov SKOS je potrebné vyčleniť URI v kontexte menného priestoru RDF alebo schémy pojmov SKOS spolu s jednoznačnou definíciou prvku alebo pojmu.
4. Krok: V tomto kroku sa dátový prvok, ktorý bol pripravený v predchádzajúcom kroku, začlení do existujúcej schémy alebo sa z nej odstráni. V XSD sa pridáva nový prvok alebo kód; zmenou existujúceho prvku alebo kódu sa prepíše stará verzia; vymazaním sa jednoducho odstráni z definície schémy. V menných priestoroch RDF a pojmových schémach SKOS sa zmena sémantiky môže uplatniť len vtedy, ak nemá vplyv na existujúce aplikácie. Vo všeobecnosti možno sémantický význam rozšíriť (keďže existujúce údaje zostávajú platné), ale nikdy nie zúžiť (čo by mohlo spôsobiť neplatnosť existujúcich údajov).

Vymazaniu prvkov alebo termínov slovníka by sa malo zabrániť, pokiaľ sa nedá s úplnou istotou overiť, že takéto prvky sa nikde nepoužívajú; v opačnom prípade by sa mali prvky anotovať, napríklad pomocou vlastnosti status (napríklad adms:status) s hodnotou „Deprecated“.

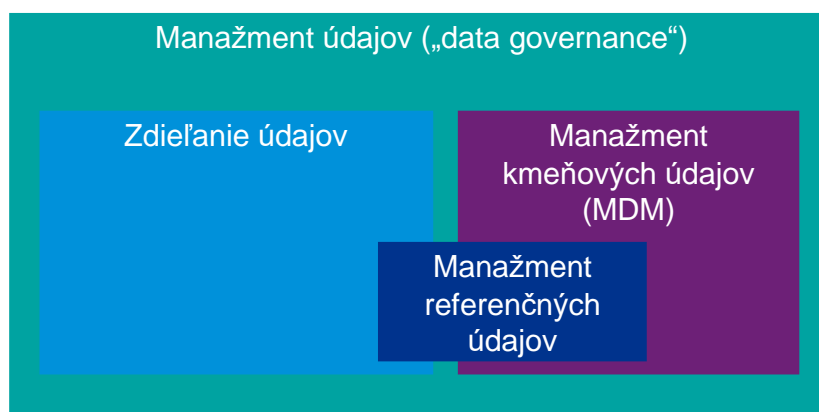
Talend sa dá tiež použiť na manažment metadát – na centralizáciu XSD schém pre XML údaje⁴. – v XML-XSD štandarde. Implementovalo sa v ňom aj čiastočné mapovanie na dátové prvky v CMÚ v jeho dátovom slovníku. Tento proces mapovania a nástrojov preň

⁴ Zdroj: <https://help.talend.com/r/en-US/7.3/open-studio-user-guide/centralizing-xml-file-metadata>, Dátum referencie: 28.06.2023

však treba vylepšiť a intenzívnejšie zaviesť do praxe. Existujú aj dedikované nástroje na vytváranie a manažment XSD schém ako napríklad Altova⁵.

Sémantické metadáta sa spravujú v nástroji VocBench⁶, z ktorého sa publikujú na znalosti.gov.sk.

5.1.3 Vybudovať nástroje pre správu referenčných údajov ako súčasť celkového manažmentu údajov



Obrázok 3: Manažment referenčných údajov ako podmnožina MDM

Referenčné údaje ako podmnožina kmeňových údajov možno manažovať obdobne ako kmeňové údaje, respektíve ich možno zahrnúť do riešení pre MDM (kapitola 5.1.5). Referenčné údaje sú však kľúčové aj pri akomkoľvek zdieľaní údajov vo verejnej správe. Opäť je tu však problém koexistencie dvoch svetov – XML a RDF, ako bolo popísané v kapitole 5.1.2. Typickým príkladom pre túto výzvu je základný Register adries⁷, ktorý by mal byť založený na ontológii adresy, ale zároveň by sa mali s ním dať stotožňovať adresy aj v XML údajoch. Dobrou praxou pre základné číselníky teda musí byť, aby sa samotná hodnota / kľúč sa používala ako posledná časť URI:{základná-časť} / {kód}.

5.1.4 Zvážiť ako alternatívu decentralizovaný kmeňový záznam pomocou MOU

MOU centralizuje údaje evidované o fyzickej alebo právnickej osobe vo verejnej správe v jej osobnom úložisku. Cieľom je, aby tieto údaje boli konsolidované, očistené a komplexné, tým pádom spĺňajú charakteristiky kmeňových údajov, uvedené v kapitole 5.1.1. Keďže obsahujú aj elektronickú pečať OVM zodpovedného za ich integritu, mohli by sa považovať za „Moje kmeňové údaje“ („My Master Data“) minimálne pre vybrané prípady použitia. Aj Európska komisia z hľadiska bezpečnosti údajov podporuje decentralizované riešenia⁸: „ak neexistuje špecifická potreba centrálne založených

⁵ Zdroj: <https://www.altova.com/xml-schema-tools>, Dátum referencie: 28.06.2023

⁶ Zdroj: <http://vocbench.uniroma2.it/>, Dátum referencie: 28.06.2023

⁷ Zdroj: <https://data.gov.sk/en/organization/2DF13D50-0B6F-48BA-884E-BE66DC0A2934?tags=register+adres>, Dátum referencie: 28.06.2023

⁸ Zdroj: https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2023-04/Guidelines%20on%20Access%20to%20Base%20Registries_v3.00.pdf, Dátum referencie: 28.06.2023

funkcií, decentralizovaná architektúra minimalizuje riziká pre jednotlivcov, ktorých údaje sa spracúvajú, pretože sa vyhýba jedinému miestu zlyhania a lepšie dodržiava zásadu minimalizácie údajov“.

5.1.5 Využívať viac platformu Talend aj pre oblasť správy kmeňových údajov (“MDM”) alebo prejsť inú platformu, viac kompatibilnú s CMÚ

Riešenia pre MDM musia pokryť tri oblasti:

1. Softvér pre MDM a zabezpečenie „zlatého záznamu“: Softvér MDM je špecializovaný typ softvéru, ktorý je určený na pomoc organizáciám pri správe ich kmeňových údajov. Zvyčajne zahŕňa funkcie, ako je zber údajov, validácia, kontrola kvality, správa, údržba a integrácia.
2. Nástroje na manažment údajov („data governance tools“): Tieto nástroje sú navrhnuté tak, aby pomáhali organizáciám spravovať aspekty manažmentu ich údajov. Tieto nástroje môžu zahŕňať funkcie, ako je bezpečnosť údajov, ochrana osobných údajov a súlad s regulačnými požiadavkami.
3. Nástroje na zvyšovanie kvality údajov: Pomáhajú organizáciám zlepšovať kvalitu ich údajov. Tieto nástroje môžu zahŕňať funkcie, ako je validácia údajov, čistenie údajov a porovnávanie údajov.

Medzi príklady riešení pre MDM patrí Ataccama⁹ (dá sa napojiť aj na Apache Kafka, poskytuje aj manažment metadát), Collibra¹⁰ (okrem bodu 1, dá sa však prepojiť s ETL nástrojmi alebo databázami), Talend MDM¹¹, Tibco¹², Informatica MDM a SAP Master Data Governance. Ak sa identifikujú závažné slabé miesta platformy Talend, možno zvážiť aj iné alternatívne platformy, ktoré môžu priniesť výhody aj do celkového konceptu dátovej integrácie. Tieto platformy môžu podporovať rôzne typy MDM, ako sumarizuje **Error! Reference source not found.**¹³.

Grafové databázy a sémantické štandardy ako RDF poskytujú výhodu flexibilného dátového modelu oproti relačným databázam, preto sa zvažujú aj pre využitie na MDM. Momentálne však ešte nejde o dominantné platformy, ktoré by sa využívali v podnikoch, čo sa však môže o pár rokov radikálne zmeniť, keďže tieto technológie už prekonalí fázu skepticizmu a sú na vzostupe.

⁹ Zdroj: <https://www.ataccama.com/platform/master-data-management>, Dátum referencie: 30.06.2023

¹⁰ Zdroj: <https://www.collibra.com/>, Dátum referencie: 30.06.2023

¹¹ Zdroj: <https://help.talend.com/r/Lfrx8uSM~Ud8zWd9BsWB5g/C3zLOZz8nALpuSkTdk~PIQ>, Dátum referencie: 28.06.2023

¹² Zdroj: <https://www.tibco.com/products/tibco-ebx-software>, Dátum referencie: 28.06.2023

¹³ Zdroj: https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2023-04/Guidelines%20on%20Access%20to%20Base%20Registries_v3.00.pdf, Dátum referencie: 28.06.2023

5.2 Aktualizované metodické usmernenie

5.2.1 Postup vyhlasovania referenčných údajov (procesná legislatívna procesná),

Vyhlasovať nielen referenčné údaje (základné číselníky a údaje slúžiace na klasifikáciu), ale aj kmeňové údaje alebo údaje pripravené pre zdieľanie cez CIP. Vyhlásiť údaje pripravené na zdieľanie môže akýkoľvek OVM bez udania účelu. Podmienkou je, aby v relevantných dátových prvkoch a metadátach boli použité referenčné údaje a metadáta podľa CMÚ (alebo aby minimálne existovalo mapovanie na dátové prvky a metadáta CMÚ). Vyhlásenie referenčných a kmeňových údajov musí byť súčasťou procesu správy ontológií popísaného v dokumente 1.1.2: Štandardizácia pre modelovanie údajov, aby sa vytvoril všeobecný konsenzus ohľadom dátového modelu a metadát. Okrem toho musí byť vyhlásenie kmeňových údajov doplnené solídnymi prípadmi použitia, ktoré odobria vytváranie „zlatého záznamu“ z pohľadu rizika bezpečnosti údajov a nákladov. Pre tieto prípady použitia musí byť tiež zvolený správny štýl MDM a technológia. Dátová kancelária by pre tento účel mala byť posilnená skupinou dátových architektov a bezpečnostných expertov (môže ísť aj o poskytovanie ad hoc služieb externými zdrojmi), aby vedeli dané prípady použitia a realizovateľnosť „zlatého záznamu“ posúdiť.

Error!

Reference

Contact us

Rudolf Sedmina

partner

Management Consulting

E rsedmina@kpmg.sk

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

www.kpmg.com

© yyyy Copyright owned by one or more of the KPMG International entities. KPMG International entities provide no services to clients. All rights reserved.

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavour to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation.

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.

6 Mapovanie na Centrálny model údajov

6.1 Dátové modely pre referenčné údaje

Všetky objekty evidencie, základné číselníky alebo iné datasety (napríklad analytický dataset) musia mať svoju kartu s dokumentáciou, aká sa momentálne vypracováva pre objekty evidencie dostupné v MOU v rámci aktivity 3.2. Dátové modely sú súčasťou Centrálneho modelu údajov a dostupné v Meta-informačnom systéme (v rámci stránky interoperabilita.gov.sk).

6.2 Prehľad údajov

Referenčné údaje:

- Základné číselníky,
- Ontológie a metadáta evidované na znalosti.gov.sk,
- Register adries.

Základné registre sú primárnym zdrojom „kmeňových“ údajov. Medzi základné entity kmeňových údajov, o ktorých by mali byť vedené „zlaté záznamy“ patria:

- Subjekty:
 - Občania,
 - Podnikatelia,
 - Zamestnanci verejného sektora,
 - Dodávatelia verejného sektora,
- Miesta:
 - Adresné body,
 - Parcely,
 - Geografické oblasti definované podľa smernice INSPIRE (Register priestorových informácií¹⁴, podľa technických metodických pokynov vznikol aj validátor datasetov a metadát¹⁵),
- Objekty:

¹⁴ Zdroj: <https://rpi.gov.sk/>, Dátum referencie: 03.07.2023

¹⁵ Zdroj: <https://inspire.ec.europa.eu/validator/home/index.html>, Dátum referencie: 03.07.2023

Error!

Reference

- Vozidlá (v súkromnom vlastníctve – Register vozidiel, alebo vo vlastníctve štátnej správy),
- Nehnuteľnosti vo vlastníctve alebo prenájme štátnej správy,
- Financie (vybrané pohľadávky, nedoplatky občanov a podnikateľov, položky štátneho rozpočtu),
- Softvérové licencie,
- Služby poskytované alebo nakupované štátom.

Presné vymedzenie pravidiel, aké záznamy sa dostanú do zlatého záznamu a za akých podmienok, a aké dátové prvky alebo premenné budú obsahom zlatého záznamu, musí byť predmetom dôslednej, odbornej a medzirezortnej diskusie pod vedením Dátovej kancelárie.