

Výstup č. 1.3.3: Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

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	Ceľuchová Bošanská Bárdy	3.8.2023	

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		

ZOZNAM SKRATIEK	
Skratka	Význam
AdES	Zdokonalený elektronické podpisy (<i>Advanced Electronic Signature</i>)
AI	Umelá inteligencia (<i>Artificial Intelligence</i>)
API	Rozhranie pre programovanie aplikácií (<i>Application Programming Interface</i>)
CC	Creative Commons
CSV	Hodnoty oddelené čiarkou (<i>Comma-Separated Values</i>)
DCAT	Slovník katalógu údajov (<i>Data Catalog Vocabulary</i>)
DOC	Formát textového editora MS Office
DLT	Technológia distribuovaných záznamov (<i>Distributed Ledger Technology</i>)
DTD	Definícia typu dokumentu (<i>Document Type Definition</i>)
eIDAS	Nariadenie Európskej únie č. 910/2014 o elektronickej identifikácii a dôveryhodných službách pre elektronické transakcie na vnútornom európskom trhu.
ETL	Proces extrakcie, transformácie a nahrania údajov (<i>Extract-Transform-Load</i>)
GDPR	Všeobecné nariadenie o ochrane osobných údajov (<i>General Data Protection Regulation</i>)
GPU	Grafický procesor (<i>Graphical Processing Unit</i>)
HVD	Súbory údajov s vysokou hodnotou (<i>High-Value Datasets</i>)
ISO	Medzinárodná organizácia pre štandardizáciu (<i>International Organization for Standardization</i>)
JSON-LD	JSON pre linkované údaje (<i>JSON for Linking Data</i>)
JSON	<i>JavaScript Object Notation</i>
LOD	Prepojené otvorené údaje (<i>Linked Open Data</i>)
MDM	Správa kmeňových údajov (<i>Master Data Management</i>)
MIRRI SR	Ministerstvo investícií, regionálneho rozvoja a informatizácie SR
MS	<i>Microsoft</i>
MSP	Malé a stredné podniky
NFT	Nezameniteľný token (<i>Non-Fungible Token</i>)
ODbL	<i>Open Data License</i>

ZOZNAM SKRATIEK	
Skratka	Význam
ODC-BY	<i>Open Data Commons Attribution License</i>
ODI	<i>Open Data Institute</i>
ODS	Formát textového editora OpenDocument
PKI	Infraštruktúra verejných kľúčov (<i>Public Key Infrastructure</i>)
PoS	dôkaz podielu (<i>Proof of Stake</i>)
PoW	Dôkaz práce (<i>Proof of Work</i>)
RDF	Špecifikácia formátu, podporujúca popis zdrojov pomocou abstraktného modelu metaúdajov (<i>Resource Description Framework</i>)
SSI	Samosprávna identita (<i>Self-Sovereign Identity</i>)
QES	Kvalifikovaný elektronický podpis (<i>Qualified Electronic Signature</i>)
TTL	Syntax a formát súboru na vyjadrenie údajov v dátovom modeli RDF (<i>Turtle</i>)
URI	Jednotný referencovateľný identifikátor (<i>Uniform Resource Identifier</i>)
URL	Jednotná adresa zdroja (<i>Uniform Resource Locator</i>)
VoID	Slovník prepojených súborov údajov (<i>Vocabulary of Interlinked Datasets</i>)
W3C	<i>World Wide Web Consortium</i>
XLSX	Formát tabuľkového editora MS Office
XML	<i>Extensible Markup Language</i>

Obsah

1	Úvod a zhrnutie	6
1.1	Kontext	6
1.2	Metodika realizácie výstupu	6
1.3	Manažérske zhrnutie	7
2	Dôveryhodnosť otvorených údajov	10
2.1	Dôveryhodnosť podpisovaním záznamov	11
2.1.1	Všeobecný kontext / eIDAS legislatíva	11
2.1.2	Infraštruktúra súkromných kľúčov	12
2.1.3	Hašovanie	13
2.2	Dôveryhodnosť pomocou distribuovanej siete	13
2.3	Dôveryhodnosť správcovstvom údajov	14
3	Archivácia otvorených údajov	18
3.1	Správa záznamov a elektronické podpisy	18
3.2	Správa záznamov a DLT, Blockchain	19
3.2.1	Ethereum a konsenzuálne algoritmy	20
3.2.2	Inteligentné zmluvy	21
3.2.3	Riziko DLT: Obavy o udržateľnosť	22
3.2.4	Riziko DLT: Súlad s GDPR	23
3.2.5	Riziko DLT: Potreba uchovávať samotné informácie	23
3.3	Správa záznamov a metaúdaje	24
4	Publikácia otvorených údajov	26
5	Postup publikácie informácií verejného sektora	28
5.1	Identifikácia súborov údajov na otvorenie	29
5.1.1	Stanovenie cieľovej úrovne otvorenosti	32
5.1.2	Definícia plánu zverejňovania otvorených údajov	34
5.2	Príprava súborov údajov na publikovanie	36
5.2.1	Opis súborov údajov	37
5.2.2	Výber a implementácia softvérových nástrojov	38
5.2.3	Definícia prístupu k publikovaniu súboru údajov	40

5.3	Publikácia súborov údajov	42
5.3.1	Licencovanie otvorených údajov	43
5.3.2	Poskytovanie komplexných metaúdajov	45
5.3.3	Geolokácia údajov	45
5.3.4	Spolupráca s komunitou používateľov	46
5.3.5	Meranie využitia údajov	47
6	Osvedčené postupy pre zdieľanie údajov verejného sektora	49
6.1	Implementácia celoorganizačnej stratégie	49
6.1.1	Príklad najlepšej praxe: Belgicko	49
6.2	Podpora crowdsourcingu	50
6.2.1	Príklad najlepšej praxe: Česká republika	51
6.3	Vytvorenie rámca kvality údajov	51
6.3.1	Príklad najlepšej praxe: Spojené kráľovstvo	53
6.4	Publikovanie prehľadu spravovaných údajov	54
6.4.1	Príklad najlepšej praxe: Fínsko	55
6.5	Monitorovanie a meranie opätovného použitia otvorených údajov	57
6.5.1	Príklad najlepšej praxe: Estónsko	58
7	Zahraničný prístup k archivácii údajov verejného sektora – eARD Projekt (SE)	60

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

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 údajov.

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).

Výstup Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov reflektuje požiadavky Dátovej kancelárie a nadväzuje na agendu otvorených údajov. Dokument rozvíja nasledovné témy:

- dôveryhodnosť otvorených údajov,
- archivácia otvorených údajov,
- publikácia otvorených údajov,
- výber vhodných postupov a metód,
- návod na procesy s otvorenými údajmi,
- príklady najlepšej praxe.

1.2 Metodika realizácie výstupu

Príručka je orientovaná na publikovanie otvorených údajov a ich následnú archiváciu. Okrem klasických otázok procesu publikovania otázok sa zameriame na inovatívne aspekty témy a to najmä na problematiku dôveryhodnosti údajov: za akých podmienok je možné otvorené údaje používať na právne účely a ako publikovať údaje, ktorým môže verejnosť dôverovať (je možné overiť hodnotu údajov, pravdivosť, zdroj a spôsob, akým spôsobom boli údaje menené). Dôležitá je aj archivácia otvorených údajov, keďže sa dá predpokladať, že najmä veľké datasety nebudú dostupné neustále. Koncept archivácie musí riešiť otázku, akým spôsobom sa konzumenti dostanú k dôveryhodnej historickej kópii údajov, v prípade, že údaje boli archivované. Zároveň je potrebné jasne nastaviť pravidlá, kedy je možné údaje archivovať a akým spôsobom.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

1.3 Manažérske zhrnutie

Koncept otvorených údajov

Podľa Open Knowledge Foundation (2012) otvorené údaje sú údaje, ktoré môže ktokoľvek voľne a opätovne používať a redistribuovať. Existujú dve hlavné dimenzie otvorenosti otvorených údajov: právna a technická.

Otvorené údaje sú neoceniteľným zdrojom pre verejnosť a ponúkajú neobmedzený potenciál pre jednotlivcov a organizácie. Poskytnutím neobmedzeného prístupu k údajom môžu vlády a iné organizácie podporovať transparentnosť a zlepšenie verejnej politiky. Tieto údaje vrátane súborov údajov o miere kriminality, populačných štatistik, softvérových projektov s otvoreným zdrojovým kódom, údajov o počasí, regionálnych hraníc, miery neobsadenosti bytov a údajov zo sčítania ľudu môže tiež použiť ktokoľvek na lepšie pochopenie sveta okolo seba alebo na vytváranie inovatívnych produktov, a služieb. Otvorené údaje sa líšia od súkromných údajov, pretože sú voľne dostupné verejnosti a možno ich zdieľať, upravovať a opätovne používať na rôzne účely. Vďaka sile otvorených údajov môže verejnosť profitovať napríklad aj z lepšieho vedeckého výskumu, zdokonaleného vývoja produktov a pokročilých rozhodnutí verejnej politiky.

Údaje môžu byť chránené právami duševného vlastníctva apod., preto je potrebné zabezpečiť, aby používatelia mali povolenia na legálne opätovné použitie a redistribúciu otvorených súborov údajov. Bežným spôsobom, ako dosiahnuť tento cieľ, je správne licencovať otvorené súbory údajov.

Technická otvorenosť zabezpečuje, že otvorené údaje sú prístupné a ľahko použiteľné. Otvorené údaje by mali byť prístupné cez API (alebo dostupné na stiahnutie z internetu) v otvorených a strojovo čitateľných formátoch. Aj keď sa používajú strojovo čitateľné formáty, medzi rôznymi formátmi existujú rozdiely. Tim Berners-Lee (2006) navrhol 5-hviezdičkovú schému nasadenia. Každá hviezdička predstavuje jednu úroveň otvorenosti súboru údajov na základe použitých formátov.

Tabuľka 1: Popis úrovni otvorenosti údajov (Hausenblas, 2012)

Úroveň otvorenosti	Popis
1*	— Údaje sú dostupné pod otvorenou licenciou. — Používa sa akýkoľvek formát.
2*	— Údaje sú dostupné pod otvorenou licenciou. — Používa sa formát štruktúrovaných údajov, napr. MS Excel.
3*	— Údaje sú dostupné pod otvorenou licenciou. — Používa sa štruktúrovaný a otvorený dátový formát, napr. CSV.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

4*	<ul style="list-style-type: none">— Údaje sú dostupné pod otvorenou licenciou.— Používa sa štruktúrovaný a otvorený dátový formát.— Objekty sú identifikované pomocou URI.
5*	<ul style="list-style-type: none">— Údaje sú dostupné pod otvorenou licenciou.— Používa sa štruktúrovaný a otvorený dátový formát.— Objekty sú identifikované pomocou URI.— Údaje sú prepojené s inými údajmi s cieľom poskytnúť kontext.

Dôveryhodnosť otvorených údajov

Otvorené údaje sú založené na myšlienke, že určité údaje by mali byť voľne dostupné každému, aby ich mohol použiť a znova publikovať podľa vlastného uváženia, bez obmedzení autorských práv, patentov alebo iných mechanizmov kontroly. Využitie tohto druhu údajov si vyžaduje znalosti o ich pôvode, kvalite a dôveryhodnosti, aby sa zabezpečilo, že údaje budú také, aké sa od nich očakávajú. Kvalitu údajov možno definovať ako údaje, ktoré sú vhodné na ďalšie použitie verejnosťou. Dôveru v otvorené údaje definujeme ako presvedčenie, že údaje sú správne, úplné, platné a spoľahlivé. Pre papierové aj elektronické záznamy si dosiahnutie cieľa presných údajov vyžaduje primerané postupy a systémy kontroly, ktoré tvoria systém manažmentu kvality.

Sľubným prístupom k dosiahnutiu objektívnej dôveryhodnosti otvorených údajov je obohatenie údajov o digitálne podpisy. V súčasnej dobe existuje niekoľko prístupov k podpisovaniu otvorených údajov. Kryptografické opatrenia, ako sú (kvalifikované) elektronické podpisy a pečate od poskytovateľov dôveryhodných služieb umožňujú nepopierateľnosť záznamov, a teda jedinečný dôkaz ich pravosti a integrity dôveryhodnou treťou stranou. Spolu s kvalifikovanou časovou pečiatkou poskytujú aj platný dôkaz o existencii v danom čase. Kryptografické prístupy však majú v súčasnosti obmedzenú škálovateľnosť, najmä kvôli náročnosti na komunikáciu a úložisko (spôsobenej veľkým počtom potrebných podpisov). Navyše záruky spoľahlivosti založené na podpisoch sú výrazne slabšie ako prístupy založené na technológii distribuovaných záznamov (*Distributed Ledger Technology, DLT*), ktoré je omnoho ťažšie pozmeňovať, no trpia obmedzenou škálovateľnosťou kvôli obmedzenej priepustnosti a vysokým nákladom na infraštruktúru.

Inherentné kryptografické protokoly DLT (napríklad Blockchain) podľa eIDAS v súčasnosti nespĺňajú požiadavky na (kvalifikované) podpisy, pečate alebo časové pečiatky a musia byť doplnené o dôveryhodné služby, ktoré sú potrebné na zabezpečenie skutočnej overiteľnosti digitálnych transakcií, ako aj dôkazu o existencii. Efektívne prístupy preto kombinujú oba prístupy – dlhodobé uchovávanie záznamov sa v súčasnosti vykonáva pomocou mechanizmov založených na kryptografických metódach opätovným podpisovaním a rehašovaním (kvalifikovaných) podpisov (alebo pečatí) v kombinácii s kvalifikovanou elektronickou časovou pečiatkou alebo dôkazovými záznamami.

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

Použitá literatúra:

BERNERS-LEE, T. 2006. Linked Data - Design Issues. In: *Design Issues*.

HAUSENBLAS, M. 2012. 5 star Open Data. In: 5 star Open Data.

Open Knowledge Foundation, 2012. The Open Data Handbook. In: *Open Data Handbook*.

2 Dôveryhodnosť otvorených údajov

Literatúra ponúka rôzne rámce na definíciu dôvery, dôveryhodnosti a toho, čo znamená byť dôveryhodný v rôznych prostrediach. Tieto rámce rozdeľujú dôveru na kľúčové zložky, ako sú „dôveryhodnosť“, „spoľahlivosť“ a „kompetentnosť“. Niekoľko štúdií sa tiež pokúsilo určiť, čo znamená dôvera v kontexte údajových ekosystémov. Gupta (2015) definuje dôveru ako „ochotu strany byť zraniteľnou voči činom druhej strany na základe očakávania, že druhá strana vykoná určitú činnosť“. Preto dôvera môže čiastočne určovať mieru, do akej sú jednotlivci a organizácie ochotní uzatvárať medzi sebou dohody. Dôvera ovplyvňuje aj to, ako sa aktéri správajú k ostatným účastníkom ekosystému, vrátane výmeny údajov v rámci ekosystému.

Dôveryhodnosť aktéra možno hodnotiť pomocou kombinácie interných (povesť, existujúci vzťah dôvery) a externých (zákony, predpisy) hnacích síl. Dôveryhodnosť bude závisieť od kontextu tvrdení, činov a sľubov aktéra, ako aj od protistrán, s ktorými aktér v systéme komunikuje. Aj keď sú externé pravidlá správania a nariadenia nevyhnutné, aktéri majú určitý stupeň autonómie na zlepšenie dôveryhodnosti, ako napríklad získanie externej certifikácie alebo účasť na auditoch. S dôveryhodnosťou je spojená aj autenticita. ISO 15489-1 definuje autenticitu a spoľahlivosť ako: „Autentický záznam je záznam, o ktorom sa dá dokázať, že a) je taký, za aký sa vydáva, b) že bol vytvorený alebo odoslaný osobou, ktorá ho údajne vytvorila alebo odoslala, a c) boli vytvorené alebo odoslané v predpokladanom čase“ (International Standards Organization, 2016).

Autenticita je dôležitá, pretože vďaka nej ľudia dôverujú záznamom. Jednou z výziev pre moderné archívy, t. j. tie, ktoré sa zaoberajú elektronickými záznamami, je, ako zachovať autenticitu záznamov v priebehu času. Dôveryhodnosť záznamov pramení z pravosti a dôkaznej hodnoty záznamov. Dôveryhodnosť záznamov často zaručujú dôveryhodné digitálne úložiská, v ktorých sa záznamy uchovávajú. Archivár je často vnímaný ako správca dôveryhodných záznamov, ktorý by mal garantovať, že pravosť záznamov bude zachovaná v priebehu času a nikto nemôže s uloženými záznamami manipulovať.

Otvorené údaje sú o opätovnom používaní verejných údajov. Keď sú otvorené údaje vo verejnom úložisku údajov, možno ich spravovať a uchovávať tak, aby sa zachovala ich autenticita. Keď sa však údaje presunú mimo úložiska, s údajmi sa bude stále zaobchádzať a budú sa interpretovať ako dôveryhodné verejné údaje, ale ich pravosť už nie je zaručená. V praxi existuje niekoľko základných princípov, ktoré zabezpečujú dôveryhodnosť a autenticitu otvorených údajov:

1. Údaje musia byť úplné,
2. Údaje musia byť primárne,
3. Údaje musia byť včasné,
4. Údaje musia byť prístupné,

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

5. Údaje musia byť strojovo spracovateľné,
6. Prístup k údajom musí byť nediskriminačný,
7. Formáty údajov musia byť neproprietárne,
8. Údaje musia mať licenciu, ktorá neobmedzuje ich použitie, vrátane komerčného použitia.

Na prepojenie otvorených údajov sa používa koncept prepojených otvorených údajov (*Linked Open Data*, ďalej len LOD), ktoré umožňujú jednoduchšie zdieľanie údajov a prístup k nim, čo znamená, že k údajom môže pristupovať viac ľudí a využívať ich vo svoj prospech. LOD je spôsob publikovania údajov pomocou otvorených štandardov a formátov, vďaka ktorým sú údaje prístupnejšie a interpretovateľnejšie a ľuďom poskytuje kontrolu nad údajmi. Prepojené otvorené údaje pridávajú k údajom kontext a umožňujú webovým aplikáciám porozumieť údajom a interpretovať ich s cieľom poskytovať pokročilé služby a funkcie. Umožňujú tiež zdieľanie údajov medzi rôznymi typmi softvéru a platforiem a ich integráciu oveľa efektívnejším spôsobom. Na základe týchto výhod možno povedať, že dôveru k prepojeným otvoreným údajom možno definovať v dvoch častiach:

- Autentifikácia pôvodu súboru údajov: Dôveryhodnosť a spoľahlivosť súboru údajov by mala byť preukázaná užívateľom, aby mohli údajom dôverovať. V skutočnosti používatelia potrebujú vidieť pôvod údajov pri získavaní údajov. Takže určenie vlastníctva údajov a overenie vlastníctva je veľmi dôležité.
- Validácia súborov údajov: Validácia súborov údajov je dôležitá, pretože zabezpečuje, že používané údaje sú presné a spoľahlivé. Okrem toho overovanie súborov údajov zabezpečuje, že všetky zdroje údajov sú aktuálne a že akékoľvek transformácie vykonané na údajoch nespôsobili žiadne chyby alebo skreslenia.

2.1 Dôveryhodnosť podpisovaním záznamov

2.1.1 Všeobecný kontext / eIDAS legislatíva

Nariadenie EÚ č. 910/20141 (ďalej len eIDAS), o elektronickej identifikácii a dôveryhodných službách pre elektronické transakcie na vnútornom trhu, definuje predvídateľné regulačné prostredie pre elektronickú identifikáciu a súbor elektronických dôveryhodných služieb, konkrétne elektronické podpisy, pečate, časové pečiatky, služby doporučeného doručovania a certifikáty na overenie webových stránok. Cieľom tohto nariadenia je harmonizovať štandardy elektronického podpisu v rámci členských krajín EÚ, aby organizácie mohli vykonávať najmä cezhraničné transakcie.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

eIDAS umožňuje vybrať si, akú úroveň právneho uznania chcú organizácie použiť, bez vynucovania digitálnych podpisov pre všetky prípady použitia. V rámci eIDAS existujú tri typy podpisov:

- Základné elektronické podpisy,
- Pokročilé elektronické podpisy (AdES), ktoré si vyžadujú niektoré bezpečnostné prvky, ktoré zaručujú, že je jedinečne prepojený s podpisovateľom, je schopný identifikovať podpisovateľa a je spojený s údajmi takým spôsobom, že je možné zistiť akúkoľvek následnú zmenu údajov,
- Kvalifikované elektronické podpisy (QES), ktoré poskytujú dodatočnú úroveň uistenia o totožnosti podpisovateľa a zvýšenú ochranu a úroveň uistenia o vytvorení podpisu. Na vytvorenie je potrebné špeciálne zariadenie.

2.1.2 Infraštruktúra súkromných kľúčov

V súčasnom stave techniky sú QES implementované pomocou asymetrickej kryptografie. S touto technológiou každý signatár vlastní kľúčový pár vytvorený zo súkromného a verejného kľúča (technológia sa nazýva kryptografia verejného kľúča) a takto vytvorené elektronické podpisy sa nazývajú digitálne podpisy.

Infraštruktúra súkromných kľúčov (*Private Key Infrastructure*, ďalej len PKI) je jedným z mechanizmov, ktoré možno použiť na zabezpečenie dôvery v prenos údajov. Používanie PKI zaisťuje dôveru v záznamy a má štyri dôležité princípy:

- Autentifikácia – proces, ktorým je strana schopná preukázať identitu subjektu alebo jednotlivca. Autentifikácia sa používa na zabezpečenie toho, že subjekt je tým, za koho sa vydáva, a že má právo na prístup k určitým informáciám alebo zdrojom.
- Nepopierateľnosť – mechanizmus, ktorý zabezpečuje, že jednotlivec alebo organizácia nemôže poprieť vykonanie konkrétnej transakcie. Často sa používa na zaručenie toho, že odosielateľ záznamu nemôže neskôr poprieť jej odoslanie. Nepopierateľnosť sa dosahuje použitím digitálnych podpisov, špecializovaného hardvéru a softvéru a autentifikačných služieb tretích strán.
- Šifrovanie – skrytie obsahu záznamu pred tretími stranami tým, že sa stane nečitateľnou bez zodpovedajúceho dešifrovacieho kľúča. Šifrovanie používa matematický algoritmus na pomiešanie údajov, vďaka čomu sú nečitateľné pre každého, kto nevlastní zodpovedajúci dešifrovací kľúč.
- Ochrana osobných údajov – používateľ alebo organizácia by si mala udržiavať kontrolu nad svojim súkromným kľúčom, uchovávať ho v tajnosti a v bezpečí. Zabezpečuje, že prístup k údajom alebo zdrojom má len organizácia, ktorá vlastní súkromný kľúč.

Proces vytvárania a overovania podpisu je nasledujúci:

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

1. Signatár používa súkromný kľúč na podpísanie dokumentu – Súkromný kľúč je v skutočnosti tajný kód používaný matematickou funkciou, aby sa údaje stali nezrozumiteľnými (šifrovanie údajov).
2. Overovateľ používa na overenie podpisu verejný kľúč signatára – Verejný kľúč je v skutočnosti kód používaný reverznou matematickou funkciou na získanie počítačových údajov zo zašifrovaných údajov.
3. Overenie podpisu verejným kľúčom – podpis bol vypočítaný so zodpovedajúcim súkromným kľúčom. Pôvodcom podpisu môže byť iba podpisovateľ, ktorý vlastní súkromný kľúč. V dôsledku toho osoba, ktorá vlastní súkromný kľúč zhodný s verejným kľúčom (t. j. signatár), nemôže poprieť, že je pôvodcom takéhoto podpisu; táto vlastnosť nepopierateľnosti je základom každého podpisu.

2.1.3 Hašovanie

Hash hodnota je štandardizovaný a jedinečný súhrn správy (dokumentu), ktorý sa získa aplikáciou špecifického kryptografického nástroja nazývaného kryptografická hašovacia funkcia. Hašovacia funkcia je akákoľvek funkcia, ktorú možno použiť na mapovanie digitálnych údajov ľubovoľnej veľkosti na digitálne údaje pevnej veľkosti, pričom malé rozdiely vo vstupných údajoch spôsobujú veľmi veľké rozdiely vo výstupných údajoch. Kryptografická hašovacia funkcia je hašovacia funkcia, ktorá má špecifické bezpečnostné vlastnosti:

- Považuje sa za prakticky nemožné znovu vytvoriť vstupnú správu z jej hašovacej hodnoty;
- Považuje sa za prakticky nemožné vypočítať z konkrétnej správy druhú správu, ktorá má rovnakú hodnotu hash (t. j. rôzne správy vedú k rôznym hodnotám hash);
- Považuje sa za prakticky nemožné nájsť dve rôzne správy, ktoré by viedli k rovnakej hodnote hash.

Pri takýchto vlastnostiach je pri opakovanom použití na rovnakú správu hodnota hash vždy rovnaká, ak je správa mierne upravená (aj o jeden jediný bit) hodnota hash bude vždy iná. To umožňuje overiť integritu správy v porovnaní so správou, pre ktorú bol predtým hash vypočítaný – keď sú hodnoty hash identické, potom sú identické aj správy. Keďže kryptografické hodnoty hash predstavujú veľké množstvo údajov ako oveľa menšie číselné hodnoty, často sa používajú s digitálnymi podpismi.

2.2 Dôveryhodnosť pomocou distribuovanej siete

V posledných rokoch technológia distribuovaných záznamov (*Distributed Ledger Technology*, DLT) a jej najznámejší predstaviteľ blockchain sa dostali do povedomia verejnosti, najmä vďaka NFT umeniu, kryptomenám a kryptoburzám ako FTX. DLT je typ decentralizovanej digitálnej technológie, ktorá umožňuje bezpečné a transparentné zdieľanie údajov a transakcií medzi viacerými uzlami (*nodes*). Využíva mechanizmus konsenzu, ktorý zabezpečuje, že všetky transakcie sú platné a nezmenené.

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

DLT overovanie autenticity funguje tak, že distribuované siete umožňujú prenos údajov alebo hodnôt z jednej strany na druhú a to bez zapojenia sprostredkovateľov. Po zapísaní do distribuovaných záznamov (*ledger*) sú transakcie nemenné, hlavne na základe hašovacej ochrany údajov uložených v reťazci. Každá transakcia sa preto dá spoľahlivo sledovať v reťazci. Ak sú faktické distribuované údaje alebo transakcie spojené do sekvenčných prepojených blokov, nazýva sa to blockchain, čo je špeciálny druh DLT. Bloky môžu obsahovať aj hash predchádzajúceho bloku a tak vybudovať spomínanú hash-ochranu a takzvanú časovú pečiatku (*Timestamp*). Túto časovú pečiatku DLT odlišujeme od časových pečiatok definovaných v eIDAS a súvisiacich štandardoch z dôvodu nedostatku dôveryhodného zdroja času, chýbajúceho vytvárania a overovania digitálnych podpisov poskytovateľom dôveryhodných služieb a chýbajúceho dôkazu o existencii vytvoreného treťou stranou (Sorge a Leicht, 2022).

DLT je forma technológie, ktorá si nevyhnutne nevyžaduje sieťového operátora alebo konzorcium pre riadenie. V závislosti od typu DLT môžu mať transakcie a údaje rôzne úrovne prístupu a viditeľnosti. Vo verejných DLT sú všetky transakcie a údaje viditeľné pre všetkých, zatiaľ čo súkromné DLT umožňuje ich zobrazenie iba oprávneným používateľom. Verejné DLT dávajú každému možnosť overiť a ukladať transakcie, zatiaľ čo súkromné DLT obmedzujú, kto to má povolené na základe udelených prístupových práv.

DLT preto možno chápať ako distribuovaný systém, ktorý poskytuje dôveru pomocou kryptografickej ochrany a nemennosti, ako aj mechanizmu konsenzu, ktorý umožňuje, aby boli neoprávnené zmeny transparentné. DLT je užitočná iba v distribuovaných ekosystémoch s viac ako 2 stranami, kde dôvera v technológiu je nevyhnutná pre úspech. V súčasnosti EÚ zlepšuje európsku infraštruktúru blockchainových služieb ako paneurópsku sieť DLT so zameraním na prípady použitia, ako napríklad:

- Notárske overenie alebo overenie údajov,
- Samosprávna identita (*Self-Sovereign Identity*, SSI),
- Digitálne dôkazy,
- Elektronické registratúry a tokenizácia,
- Medziodvetvové obchodné platformy alebo platformy na výmenu údajov.

Vo väčšine prípadov DLT funguje ako transakčná vrstva alebo v prípade SSI aj ako kotviaca vrstva, kde budú záznamy väčšinou uložené mimo reťazec (z dôvodu obmedzenia výkonu a dôvodov ochrany osobných údajov, napr. GDPR), len anonymizované alebo pseudonymizované ekvivalenty sú v reťazci.

2.3 Dôveryhodnosť správcovstvom údajov

V poslednej dobe sa medzi popredné mechanizmy zvyšovania dôveryhodnosti údajov (vrátane otvorených údajov) radí aj správcovstvo údajov prostredníctvom organizačnej jednotky s názvom dátový trust (*Data Trust*).

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

Trust (angl. doslova dôvera) je v krajinách angloamerického práva (Common Law) právny vzťah medzi osobami, z ktorých jedna majetok spravuje tak, aby z neho druhá osoba (beneficient) mala úžitok. Je to jedna z foriem fiduciárneho vzťahu, čiže plnej správy cudzieho majetku. Najčastejšie sa trusty používajú na usporiadanie rodinného majetku a dobročinných darov. Trust je teda unikátnym právnym inštitútom angloamerického práva a na európskom kontinente mu je najbližšia fundácia (nadácia).

ODI definuje dôveryhodnosť údajov ako právnu štruktúru, ktorá poskytuje nezávislé spravovanie údajov. Správcovia údajov rozhodujú o tom, kto má prístup k údajom, za akých podmienok a kto z nich môže mať úžitok. Túto úlohu zvyčajne majú organizácie, ktoré zhromažďujú a uchovávajú údaje. Vďaka trustom údajov umožňujú organizácie, ktoré zhromažďujú a uchovávajú údaje, nezávislej inštitúcii rozhodovať o tom, ako sa tieto údaje používajú a zdieľajú na dohodnutý účel. Správcovia dátového trustu preberajú zodpovednosť a majú slobodu zdieľať údaje spôsobmi, ktoré odomknú ich hodnotu (Open Data Institute, 2020). Dátový trust preto možno chápať ako:

- nezávislú inštitúciu, ktorá by mohla pomôcť zlepšiť zdieľanie údajov a kto k nim môže pristupovať,
- nástroj na zníženie nákladov a zručností potrebných na spravovanie a zdieľanie údajov,
- príležitosť pre podniky na prístup k údajom,
- inovačný potenciál, ktorý vie podporiť nové technológie, ako je AI, ktoré pomáhajú ľuďom robiť informovanejšie rozhodnutia, vytvárať pracovné miesta a stimulovať rast,
- Prostriedok na zabezpečenie, aby sa výhody údajov distribuovali širšie, eticky a spravodlivo.

Stručne povedané, dátové trusty sú nezávislé organizačné štruktúry, ktoré sú právne zodpovedné za spravovanie údajov od vlastníkov alebo používateľov údajov. Majú schopnosť rozhodovať o prístupe k údajom a ich používaní s tým, že hlavnou úlohou je zabezpečiť, aby sa údaje spravovali a využívali spôsobom, ktorý bude prínosom pre zainteresované strany.

Príkladom je aj najväčšia otvorená databáza spoločností na svete – OpenCorporates, ktorá v roku 2017 oznámila OpenCorporates Trust – firemnú štruktúru na ochranu svojho verejnoprospešného poslania vytvárať globálny archív verejne dostupných záznamov o firmách pre širší verejný prospech, vrátane boja proti praniu špinavých peňazí, korupcii, podvodom a organizovanému zločinu. Údaje OpenCorporates boli ústredné pre množstvo prelomových vyšetrení, vrátane ICIJ Panama a Paradise Papers.

Tak ako niektoré krajiny majú národné trusty na starostlivosť o zdroje, ako sú historické budovy a životné prostredie, je isté, že v budúcnosti bude existovať čím ďalej viac sietí národných a medzinárodných trustov pre údaje, ktoré spravujú základnú dátovú infraštruktúru. Príkladom môže byť aj správcovstvo údajov Partnerstva pre výskum administratívnych údajov (*Administrative Data Research Partnership*), ktoré spravuje

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

údaje ministerstiev Spojeného kráľovstva, miestnych samospráv a zdravotníckych orgánov vo Veľkej Británii.

Tabuľka 2: Vzťah medzi rôznymi prístupmi k zdieľaniu alebo prístupu k údajom

Prístup	Rozlišovacie znaky
Dátový trust	Správcovia dátového trustu prevezmú zodpovednosť (s určitými záväzkami) za spravovanie údajov na dohodnutý účel.
Dátové družstvá (Data Cooperatives)	Vzájomná organizácia vlastnená a demokraticky riadená členmi, ktorí delegujú kontrolu nad spoločnými údajmi.
Spoločný súbor údajov (Data Commons)	Zhromažďuje údaje zo širokého spektra zdrojov do jednotnej databázy, aby boli údaje prístupnejšie a užitočnejšie.
Úložisko osobných údajov (Personal Data Stores)	Uchováva údaje poskytnuté jednotlivcom v ich mene a poskytuje prístup k týmto údajom tretím stranám, keď ich jednotliviec nasmeruje.
Výskumné partnerstvá (Research partnerships)	Vzájomné poskytovanie prístupu k údajom univerzít a výskumných organizácií.

Použitá literatúra:

AEPD. 2019: *Introduction To The Hash Function As A Personal Data Pseudonymisation Technique*.

BAUER, F. 2012. *Linked Open Data: The Essentials*. reep.org

EC. 2017. *eIDAS A Quickstart Guide To Electronic Identification (Eid) And Trust Service Solutions For Businesses Made Easy!*.

GUPTA, A. a DHAMI, A., 2015. *Measuring the impact of security, trust and privacy in information sharing: A study on social networking sites*. Journal of Direct, Data and Digital Marketing Practice, 17(1).

International Standards Organization. 2016. *Information and Documentation – Records Management – Part 1: Concepts and Principles*. Iso 15489-1:2016, ldt.

KUHN. R. 2001: *Introduction to Public Key Technology and the Federal PKI Infrastructure*.

Open Data Institute. 2020. *Designing trustworthy data institutions*.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

SORGE, C. a LEICHT, M. 2022. *Blockchain-based electronic time stamps and the eIDAS regulation: The best of both worlds*. 10.2966/scip.190122.61.

3 Archivácia otvorených údajov

Digitálne záznamy sa stávajú primárnym zdrojom spoľahlivých informácií v digitálnom svete. Presné, úplné a autentické záznamy zásadne poskytujú základ dôvery medzi jednotlivcami a organizáciami. Vytváranie a ochrana týchto záznamov, ako aj zachovanie ich integrity je však komplikovanou úlohou pre podniky a verejné inštitúcie na celom svete, pretože digitálne záznamy sú vystavené veľkému riziku pre krehkosť digitálnych nosičov, absenciu presných a úplných metaúdajov a rýchle zastarávanie softvéru a počítačových systémov. Tieto výzvy je obzvlášť ťažké riešiť v prostrediach, kde sú zdroje, kontrolné systémy, informovanosť a odborná kapacita obmedzené. Ak záznamy nie sú chránené a uchovávané, môže to predstavovať vysoké riziko pre všetkých používateľov a zainteresované strany.

V súlade s platnými medzinárodnými štandardmi, správa záznamov poskytuje potrebné procesy, úlohy a zodpovednosti, riadenie a technické riešenia pre správu elektronických záznamov. Základnými charakteristikami sú autenticita, integrita a sledovateľnosť elektronických záznamov, ako aj ich dostupnosť a prenosnosť. Tieto prirodzené vlastnosti musia byť zabezpečené a zachované dovtedy, kým sú záznamy potrebné. To si vyžaduje dostupnosť a prenosnosť záznamov. Záznamy spĺňajúce tieto požiadavky sa nazývajú smerodajné záznamy a ich hodnovernosť, teda ich pravosť, celistvosť a sledovateľnosť, musí byť zachovaná až do konca doby uchovávania. Systém vytvárania, a uchovávania záznamov až do ich dispozície sa nazýva záznamový systém. Dôveryhodnosť záznamov sa dá zaistiť pomocou rôznych metód – v súčasnej dobe sa dôveryhodnosť záznamových systémov najviac zvyšuje pomocou elektronických podpisov a hašovaním, technológiou distribuovaných záznamov (*Distributed Ledger Technology*, DLT), a efektívnou správou metaúdajov.

3.1 Správa záznamov a elektronické podpisy

Aby záznam zostal spoľahlivý, autentický, použiteľný a so zachovanou integritou, je potrebné zachovať jeho obsah, kontext a niekedy aj štruktúru. Dôveryhodný záznam uchováva skutočný obsah samotného záznamu a informácie o zázname, ktoré súvisia s kontextom, v ktorom bol vytvorený a použitý – nezachovanie štruktúry záznamu naruší jeho štruktúrnu integritu. To zase môže podkopať spoľahlivosť a autenticosť záznamu. Pri archivácii záznamov, ktoré sú zabezpečené elektronickými podpismi, sú potrebné osobitné úvahy o:

- **Obsahu:** Elektronický podpis alebo podpisy v zázname sú súčasťou obsahu. Uvádzajú, kto podpísal záznam a či táto osoba schválila obsah záznamu. Viacnásobné podpisy môžu označovať počiatočné schválenie a následné súbehy. Podpisy sú často sprevádzané dátumami a inými identifikátormi. Nedostatok týchto informácií vážne ovplyvňuje spoľahlivosť a správnosť dokumentu.
- **Kontexte:** Niektoré technológie elektronického podpisu sa spoliehajú na jednotlivé identifikátory, ktoré nie sú vložené do obsahu záznamu, dôveryhodné cesty a iné prostriedky na vytvorenie a overenie platnosti elektronického podpisu. Tieto

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

informácie sú mimo obsahu záznamu, ale sú dôležité pre kontext záznamu, pretože poskytujú ďalší dôkaz na podporu spoľahlivosti a pravosti záznamu. Nedostatok týchto kontextových záznamov vážne ovplyvňuje schopnosť overiť platnosť podpísaného obsahu.

- Štruktúre: Zachovanie štruktúry záznamu znamená jeho fyzický a logický formát a vzťahy medzi dátovými prvkami tvoriacimi záznam zostávajú fyzicky a logicky nedotknuté. Organizácia môže určiť, že je potrebné zachovať štruktúru elektronického podpisu, v takom prípade je potrebné uchovať hardvér aj softvér, ktorý vytvoril podpis (napr. čipy alebo šifrovacie algoritmy), aby bolo možné celý záznam neskôr podľa potreby znovu overiť.

3.2 Správa záznamov a DLT, Blockchain

DLT a Blockchain ponúkajú unikátne riešenie, ktoré môžu archívy použiť na dokazovanie autenticity ich záznamov. Porovnanie kontrolného súčtu (*checksum*) a metaúdajov záznamu s údajmi zaznamenanými v reťazci, slúži ako dôkaz, že v zázname neboli vykonané žiadne zmeny (úmyselné alebo náhodné), odkedy bol uložený v archíve. DLT a Blockchain týmto spôsobom poskytujú prostriedok na vytvorenie dôvery v platnosť záznamu – tento aspekt technológie je veľmi atraktívnym pre každú inštitúciu alebo subjekt, ktorý hľadá spôsob, ako dokázať, že ich záznamy sú autentické a nezmenené.

Hlavnou vlastnosťou blockchainu je, že údaje v reťazci nemožno zmeniť, čo zaručuje najmä reťazová štruktúra hašovacích hodnôt. Algoritmus hash možno vysvetliť tak, že sa vygeneruje reťazec pevnej dĺžky, ak sa zadá ľubovoľná dĺžka vstupnej informácie (napríklad binárny súbor) a vypočíta sa určitý algoritmus (napríklad algoritmus hash). Tento reťazec sa nazýva hash hodnota zodpovedajúca vstupu v rámci hash algoritmu. Pre ideálny model hash algoritmu je priestor vstupnej vzorky nekonečný, zatiaľ čo výstupný priestor je konečný. Bežne používaný algoritmus SHA256 má tieto vlastnosti a používa sa v systéme Bitcoin. Hašovacia hodnota sa používa ako ukazovateľ a je vnorená vrstvu po vrstve. Ak chcete zmeniť údaj na rozpoznanom reťazci, ovplyvní to celý výpočet a spotrebuje veľa výpočtových zdrojov. Zároveň hash hodnota tvorí Merkelov strom, ktorý dokáže rýchlo reagovať na zmeny obsahu v reťazci.

Šifrovací algoritmus sa vyvinul zo symetrického šifrovania na asymetrické šifrovanie, ktoré je princípom algoritmu digitálneho podpisu. Digitálny podpis dosahuje podobný efekt ako tradičný fyzický podpis. V sieťovom systéme blockchain existuje mnoho uzlov s rôznymi oprávneniami. Blockchain používa digitálne podpisy na dosiahnutie kontroly oprávnení a identifikáciu, či je identita správna. Každý uzol má svoj vlastný pár verejného a súkromného kľúča. Verejný kľúč, ako verejná identifikácia samotného uzla, bude vysielaný v sieti blockchain a zaznamenaný všetkými uzlami. V systéme blockchain, keď uzol iniciuje proces obchodnej logiky, najprv zašifruje obsah súkromným kľúčom a pripojí podpis k obchodnej logike. Po prijatí vysielanej správy ostatné uzly najprv overia pripojený digitálny podpis po dokončení kontroly integrity a overení platnosti identity odosielajúceho uzla, čím sa spustí následný proces.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

V systéme blockchain sa na zaznamenávaní podieľajú všetky uzly, takže záznam ktorého uzla má prednosť? Inými slovami, ako dosiahnuť konsenzus medzi uzlami je problém, ktorý musí systém blockchain vyriešiť. Tradičný centralizovaný systém so súhlasom centrálného uzla zabezpečuje autoritu a konzistenciu údajov. V decentralizovanom systéme, akým je blockchain, nie je vlastný stav a sieťové prostredie každého uzla rovnaké, pričom jeho identita sa ťažko kontroluje. Môžu existovať škodlivé uzly, ktoré bránia doručeniu správ alebo dokonca posielajú nesprávne informácie, čím narúšajú konzistenciu celého systému. Preto je otázka konsenzu veľmi dôležitou otázkou. Algoritmy konsenzu používané v blockchaine sa delia najmä do týchto kategórií, t. j. algoritmy konsenzu dôkazov práce (*Proof-of-work*, PoW), distribuované algoritmy konsenzu dôkazov o podiele (*Proof-of-Stake*, PoS) byzantské algoritmy odolné voči chybám (*Byzantine fault-tolerance Algorithms*) a inteligentné zmluvy (*Smart Contracts*).

3.2.1 Ethereum a konsenzuálne algoritmy

Algoritmy konsenzu dôkazov práce

Blockchain Ethereum začal fungovať 30. júla 2015. Viac ako sedem rokov bol reťazec zabezpečený protokolom Proof-of-Work (PoW). Ten zahŕňal veľké klastre špecializovaných počítačov, známe ako ťažobné siete, ktoré medzi sebou súťažili o náhodné číslo nazývané *nonce*. Prvému ťažiarovi (tiež *miner*), ktorý preukázal, že hash *nonce* je nižší ako úroveň náročnosti siete, bolo umožnené pridať nový blok do reťazca. Miner dostal odmenu za blok a vybral poplatky za transakcie zahrnuté v bloku. Tradičné systémy PoW sa preto spoliehajú na minerov, ktorí pridávajú blok do blockchainu výmenou za malú odmenu, zvyčajne za cenu drahého GPU hardvéru a vysokej spotreby elektrickej energie, ktorá rýchlo rástla kvôli náročnosti siete. Overovanie transakcií sa vykonáva takto:

1. Každý účet má kópiu blockchainu,
2. Nová transakcia medzi dvoma účtami sa pridá do reťazca a musí sa potvrdiť. Táto transakcia sa uskutočňuje za poplatok pre transakčné strany,
3. Potvrdzovatelia, známi ako mineri, súťažia v riešení kryptografickej hádanky s cieľom overiť, či je transakcia platná podľa stavu blockchainu,
4. Ak transakciu potvrdí dostatočný počet validátorov, prevod hodnoty sa dokončí.

Tento mechanizmus je decentralizovaný, pretože mnoho rôznych minerov musí prispieť svojou hašovacou silou (a teda výpočtovými zdrojmi), aby mali nárok na odmenu za ťažbu bloku transakcií (bloková odmena). S rastúcim počtom rôznych minerov sa sieť stáva čoraz decentralizovanejšou. Existuje však niekoľko kritických slabín tohto protokolu, a to plytvanie prácou (mineri, ktorí súťažia, ale nedostanú blokovú odmenu, plytvajú energiou), nízka rýchlosť a zraniteľnosť voči útokom, ak sieť nie je dostatočne

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

decentralizovaná – ak niektorý jednotlivý miner kontroluje väčšinu hašovacej sily (viac ako 50 %), má moc úplne prepísať blockchain a generovať falošné transakcie. To je jeden z najväčších dôvodov, prečo je potrebný bezpečný mechanizmus konsenzu.

Algoritmy konsenzu dôkazov podielu

15. septembra 2022 prijala sieť Ethereum konsenzuálny mechanizmus *proof-of-stake* (PoS). PoS je trieda konsenzuálnych algoritmov, v ktorých sa namiesto odmeňovania účastníkov na základe ťažobného výkonu odmeňujú podľa ich podielu v sieti. Títo účastníci sú validátori, ktorí pri schvaľovaní novej transakcie hlasujú v kvóre (*quorum*), čím sa znižujú náklady na schvaľovanie blokov, keďže validátori nemusia na validáciu riešiť výpočtovo náročné kryptografické hádanky.

Táto zmena je základom úplne nového konsenzuálneho protokolu, ktorý do Etherea prináša byzantskú odolnosť voči chybám (*Byzantine fault-tolerance*). Hlavným cieľom návrhu je naďalej používať konsenzus v Nakamotovom štýle (*Nakamoto-style consensus*), t. j. protokol, ktorý neustále vytvára bloky v tvare stromu a vyberá vetvu ako aktuálny reťazec pomocou pravidla fork-chain. Nový mechanizmus (nazývaný *finality gadget*) však na rozdiel od čistého konsenzu v Nakamotovom štýle inkrementálne finalizuje bloky v reťazci. Finalizovaný blok je blok, za ktorý hlasovali aspoň dve tretiny validátorov. V systéme s menej ako jednou tretinou byzantských validátorov sa finalizovaný blok nikdy neodvolá.

3.2.2 Inteligentné zmluvy

Inteligentná zmluva v distribuovanom zázname je funkcia alebo protokol, ktorý uľahčuje dohodu medzi dvoma stranami. Vo všeobecnosti sa dve strany zvyčajne dohodnú na sume prevodu (v natívnej mene blockchainu) a vykonajú ho na základe dohodnutých podmienok. Výhodou inteligentných zmlúv oproti "fyzickým" zmluvám je, že ich decentralizovaným spôsobom kontroluje tretia strana. Táto tretia strana, pôsobí ako rozhodca, ktorý vykonáva funkcie zmluvy po splnení určitých podmienok. Zvyčajne sa tieto zmluvy zaoberajú prevodom hodnoty medzi dvoma subjektmi a prevod hodnoty dokončia až po splnení podmienok inteligentnej zmluvy. V Ethereu je kontrakt uložený v sieti a je verejný, aby si ho mohol ktokoľvek prečítať, ale nie je možné určiť, medzi kým bol uzavretý.

Keďže identita osôb, ktoré tvoria inteligentnú zmluvu, nemusí byť nikdy zverejnená, prevod hodnoty medzi dvoma stranami môže prebehnúť anonymne bez toho, aby bolo potrebné obetovať dôveru, čím sa zachováva anonymita, autonómia a zmluva sa môže zálohovať v histórii blockchainu. Väčšina nových blockchainov/distribuovaných záznamov sa spolieha na inteligentné zmluvy, aby mohli fungovať pre svoj špecifický prípad použitia.

Súčasnú inteligentné zmluvy sú stále neúplné a sú náchylnejšie na logické medzery ako tradičné zmluvy. Vysokoúrovňový jazyk na písanie inteligentných zmlúv má v tomto

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

ohľade niektoré "neisté" inštrukcie, ktoré môžu spôsobiť divergenciu systému a ovplyvniť jeho konzistenciu.

Decentralizované transakcie aj napriek tomu podnietili záujem o vývoj množstva špecializovaných blockchainov, z ktorých každý navrhuje svoj vlastný špecifický konsenzuálny protokol. Konsenzus je možno najväčším faktorom pri určovaní toho, ako funguje decentralizovaný systém, keď čelí konfliktným transakciám. Rovnako dôležitá je však dátová štruktúra použitá na ukladanie a šírenie distribuovaných záznamov. Hoci blockchain, najbežnejšia dátová štruktúra, ktorá sa používa, je vhodný na zachovanie potrebných funkcií decentralizovanej siete, zlyháva z hľadiska výkonu a efektívnosti, najmä v:

- priepustnosti transakcií,
- veľkosti distribuovaných záznamov,
- celkovej účinnosti.

DLT technológia môže poskytnúť iba dôkaz o tom, že záznam nebol zmenený od jeho pridania alebo prepojenia, nemôže preto slúžiť k overeniu platnosti obsahu záznamu. Zásady správy záznamov a inštitucionálna integrita sú stále potrebné na zabezpečenie plnej dôvery v uložené záznamy. Oproti výhodám technológií DLT existuje aj niekoľko rizík, ktoré je potrebné adresovať:

- Obavy o udržateľnosť,
- Súlad s GDPR,
- Potreba uchovávať samotné informácie.

3.2.3 Riziko DLT: Obavy o udržateľnosť

Archivácia pomocou DLT systému si vyžaduje vytvorenie organizačnej siete a plánu implementácie, pretože čím dlhšie je systém ponechaný sám bez komunity aktívnych používateľov, tým menej spoľahlivé sa stávajú údaje v ňom uložené. DLT sa spolieha na jeho distribuovanú povahu pre vytváranie dôvery a bez dostatočného počtu aktívnych používateľov sa zvyšuje potenciál nečestných uzlov, ktoré teoreticky môžu ovplyvniť platnosť alebo dôveryhodnosť informácií uložených v reťazci.

Podľa dostupnej literatúry by bolo potrebné vytvoriť konzorcium najmenej o siedmich archívoch, aby sa zabezpečilo, že väčšina prispievateľov dosiahne konsenzus o obsahu bloku. Až potom môže DLT zaručiť, že záznam nebol sfalšovaný k spokojnosti verejnosti. Táto požiadavka je kľúčová – bez účasti iných organizácií táto technológia neposkytuje väčšiu istotu ako centralizovaná databáza (Green et al. 2018). Riešením je aj model s obmedzeným prístupom. To znamená, že iba pozvaní účastníci majú prístup k údajom na zápis. Vo všeobecnosti by zlomyseľný aktér musel nejakým spôsobom prevziať kontrolu nad najmenej 51 % účastníkov systému, aby poškodil údaje, a toto je jedna z najčastejších kritik DLT a blockchainu. V prípade modelu s obmedzeným prístupom je

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

však toto riziko zmiernené obmedzeným prístupom. To znamená, že prispievatelia sú známi (nie anonymní) a ak by sa 51 % z nich dohodlo na spáchaní podvodu zmenou blokov, túto aktivitu a zodpovedných partnerov by identifikovali všetci zvyšní partneri.

3.2.4 Riziko DLT: Súlad s GDPR

Ďalšie riziko vychádza z nemennosti DLT a blockchainu, čo je jednou zo silných stránok, ale aj jednou z jeho slabých stránok v tom zmysle, že čokoľvek, čo sa do bloku vloží, sa nikdy nedá odstrániť ani zmeniť. Zákony o ochrane osobných údajov v EÚ (GDPR) umožňujú jednotlivcom mať priamu kontrolu nad určitou časťou ich osobných údajov, čo im umožňuje napríklad aj odstrániť ich z online zdrojov. Pre údaje, ktoré žijú v DLT, to predstavuje potenciálne riziká pre samotný reťazec. Samotná nemennosť systému znamená, že odstránenie údajov by vyžadovalo nejakú verziu hard forku, aby sa vyhovel požiadavke o odstránenie údajov. Na blockchaine môžu nastať dva typy forkov, soft fork a hard fork. Oboje sú zmeny v pravidlách, ktorými sa riadi fungovanie blockchainu.

Hard fork je drastické opatrenie, ktoré možno použiť na zneplatnenie predtým platných údajov v sieti zmenou pravidiel toho, čo je v tejto sieti prijateľné, a špecifikovaním toho, aké staré údaje budú v budúcnosti dostupné v sieti. Zatiaľ čo hard fork môže zmeniť stav platnosti blokov údajov, nemôže ich v skutočnosti odstrániť z internetu a nespĺňať by požiadavky podľa GDPR. Aby sa údaje úplne odstránili, museli by sa odstrániť z každého uzla spusteného na DLT. To by v podstate urobilo DLT nepoužiteľným na ukladanie akýchkoľvek údajov, na ktoré sa vzťahujú zákony o ochrane súkromia.

Keďže v súčasnosti v DLT neexistujú žiadne štandardizované opatrenia na splnenie požiadaviek na ochranu súkromia podľa GDPR, odporúča sa uchovávať elektronické záznamy mimo reťazca, ak je to možné. V tomto prípade zostanú v reťazci iba záznamy o transakciách. To vedie k zložitosti pri prevádzke DLT, pretože prepojenie medzi záznamami v reťazci a mimo reťazca musí byť zachované tak dlho, ako sú záznamy potrebné.

3.2.5 Riziko DLT: Potreba uchovávať samotné informácie

Ďalšia výzva vyplýva zo skutočnosti, že neexistujú žiadne štandardizované opatrenia na zachovanie samotných informácií, t. j. ich dostupnosti alebo technickej interpretovateľnosti počas obdobia uchovávanía 10, 20 alebo viac rokov. Aby bola pravosť a integrita záznamov v reťazci (napr. transakcie) evidentná voči tretím stranám, je potrebné zabezpečiť krypto agilitu a uchovávanie evidencie záznamov a obnoviť základnú hašovaciu ochranu a spojiť ju s platným dôkazom existencie vrátane využitia najmodernejších hašových algoritmov. Typická aplikácia DLT zahŕňa ukladanie relevantných údajov (minimálne záznamy o transakciách) vo vyhradenom transakčnom objekte priamo v reťazci, prípadne prepojenom so záznamami mimo reťazca. Záznamy transakcií sú chránené Merkle tree (pomocou hash algoritmu H), ktorého koreň je

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

umiestnený v hlavičke bloku a spolu so stromom tvoria jeden blok v reťazci. Na druhej strane hlavička bloku spolu so stromom tvoria jeden blok na reťazi.

3.3 Správa záznamov a metaúdaje

Metaúdaje umožňujú overiť Integritu a dôveryhodnosť záznamov a preto sú rozhodujúce pri podpore správy záznamov počas ich životného cyklu. Metaúdaje sú štruktúrované alebo pološtruktúrované informácie, ktoré sa používajú na vytváranie, registráciu, klasifikáciu, prístup, uchovávanie a likvidáciu záznamov. Okrem tohto metaúdaje tiež:

- spájajú rôzne komponenty záznamu a spájajú záznam s inými relevantnými záznamami, čo pomáha pri porozumení a použití,
- pomáhajú organizáciám nájsť, pochopiť a použiť záznamy na rôzne účely,
- môžu byť vytvorené manuálne alebo automaticky,
- pomáhajú overovať záznamy obsiahnuté v systémoch a procesoch,
- poskytujú ochranu záznamov riadením prístupu.

Metaúdaje sú nevyhnutné pre úspech iniciatív otvoreného vládnutia. Poskytujú občanom možnosť nájsť a prístupovať k informáciám, ktoré im boli sprístupnené, porozumieť im a dôverovať im. Ak metaúdaje spojené s vládnymi digitálnymi záznamami nie sú dôveryhodné, spoľahlivé a presné, môže to viesť k nedostatku dôvery verejnosti v schopnosť vlády byť otvorený a zodpovedný. Zle vedené záznamy môžu viesť k nepresným alebo neúplným údajom, ktoré môžu viesť k niekoľkým vážnym následkom, ako je nepochopenie a zneužitie informácií, zakrývanie podvodov, skreslené zistenia a nepresné štatistiky, chybné politické odporúčania a nesprávne umiestnené financovanie. To všetko môže mať zásadný vplyv na životy občanov. Iniciatívy otvoreného vládnutia musia zabezpečiť, aby boli metaúdaje aktuálne, aby sa dosiahol maximálny úspech. Existujú tri typy metaúdajov:

- administratívne metaúdaje – metaúdaje súvisiace s používaním, riadením a procesmi kódovania digitálnych objektov za určité časové obdobie. Zahŕňajú podmnožiny technických metaúdajov, metaúdajov správy práv a metaúdajov ochrany,
- popisné metaúdaje – metaúdaje, ktoré opisujú dielo na účely objavenia a identifikácie, ako je tvorca, názov a predmet,
- štruktúrne metaúdaje – metaúdaje, ktoré naznačujú, ako sú objekty štruktúrované.

Použitá literatúra:

ETSI. 2022. An Introduction of Permissioned Distributed Ledger (PDL)

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

ETSI. 2015. Electronic Signatures and Infrastructures (ESI).

Green, A. et al. 2018. *Using Blockchain to Engender Trust in Public Digital Archives*.

LAU, K. 2020. Ethereum 2.0 An Introduction.

PEYROTT, S. 2017. An Introduction to Ethereum and Smart Contracts.

PORAT, A. 2015. Blockchain Consensus: An analysis of Proof-of-Work and its applications.

4 Publikácia otvorených údajov

Organizácie by mali zverejňovať otvorené údaje, aby zvýšili transparentnosť a stimulovali ekonomický rast. Zdieľanie údajov umožňuje občanom lepšie porozumieť politikám a činnostiam organizácií a podporuje rozsiahle sledovanie a analýzu verejnej správy s cieľom získať prehľad o rôznych aspektoch súvisiacich s tvorbou politiky. Otvorené údaje môžu tiež poskytnúť veľké množstvo poznatkov, ktoré môžu podniky využiť na rôzne účely, čo im pomáha prijímať informované rozhodnutia a vytvárať inovatívne riešenia, ktoré zase vytvárajú potenciálne nové oblasti ekonomického rastu. Otvorené údaje môžu tiež pomôcť zlepšiť vládne služby a schopnosť reagovať tým, že údaje sú ľahko dostupné úradom a lepšie rozhodnutia sa môžu riadiť štatistikou, a nie iba názormi. Zdieľanie otvorených údajov vo verejnej správe tiež pomôcť pri budovaní a udržiavaní spravodlivých, transparentných a efektívnych verejných služieb a v konečnom dôsledku umožňuje štátu lepšie reagovať na svojich občanov. Otvorené údaje tiež podporujú opätovné použitie, čo môže viesť k úsporám štátu a pomôcť znížiť pracovné zaťaženie úradníkov.

Okrem vyššie uvedených benefitov, môže publikácia otvorených údajov pomôcť aj zlepšiť vzťahy s verejnosťou, keďže sprístupnenie verejných údajov naznačuje, že štát je otvorený verejnej kontrole a je transparentný vo svojich operáciách. Sprístupnenie údajov a ich opätovné použitie môže pomôcť zlepšiť samotné verejné údaje, pretože umožňuje overenie, podporuje spoluprácu medzi odborníkmi a môže pomôcť pri vytváraní osvedčených postupov. Buchholtz et al. (2014) ako ďalší benefit otvorených údajov uvádza sieťové efekty – každý novovydaný otvorený súbor údajov v strojovo čitateľnom formáte, ktorý je ideálne prepojený s inými súbormi údajov, pridáva hodnotu celému ekosystému:

- transparentný spôsob informovania širokej verejnosti o porušovaní legislatívy – otvorené údaje môžu slúžiť na informovanie občanov o výsledkoch administratívneho dohľadu a subjektov, ktoré nedodržiavajú legislatívu. Napríklad údaje o hygienických kontrolách v reštauráciách a ďalšie údaje o výkonnosti podnikov a inštitúcií môžu motivovať tieto subjekty k poskytovaniu lepších služieb a môžu pomôcť zvýšiť konkurenciu,
- lepšie pochopenie a správa údajov v orgánoch verejného sektora – katalogizácia údajov a vývoj zoznamov súborov údajov na otvorenie môže pomôcť orgánom verejného sektora získať lepší prehľad o dostupných súboroch údajov a následne môže podporiť procesy správy a riadenia údajov,
- minimalizácia chýb pri práci s vládnymi údajmi – ak sú k dispozícii strojovo čitateľné údaje, používatelia nemusia ručne kopírovať údaje z dokumentov. Otvorené údaje preto pomáhajú minimalizovať chyby pri práci s údajmi,
- jednoduchšie preklady – ak sú grafy a tabuľky dostupné len ako obrázky v originálnom dokumente, môže byť ťažké preložiť takýto dokument. Dostupnosť strojovo čitateľných údajov umožňuje jednoduchý preklad grafov, tabuliek, hlavičiek stĺpcov alebo buniek tabuliek,

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- menej žiadostí o údaje – v situáciách, keď sú údaje dostupné len na požiadanie, môže zverejnenie otvorených údajov viesť k zníženiu takýchto žiadostí, pretože dostupnosť otvorených údajov umožňuje používateľom uspokojiť ich informačné potreby využívaním otvorených súborov údajov.

Použitá literatúra:

BUCHHOLTZ, S., BUKOWSKI, M., ŚNIEGOCKI, A. 2014. Big and open data in Europe. A growth engine or a missed opportunity? In: *Big and open data in Europe*.

MELIN, U. Challenges and Benefits in an Open Data Initiative.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

5 Postup publikácie informácií verejného sektora

Proces otvárania údajov začína vedením rozhodnutím, ktoré zaväzuje daný orgán verejnej správy k publikácii otvorených údajov. Po vydaní rozhodnutia poverí vedenie poskytovateľa údajov vybraný útvar (osobu) rolí koordinátora otvárania údajov. Koordinátor otvárania údajov zaisťuje súčinnosť a kontrolu výstupov všetkých ostatných rolí, ktoré sa na otváraní údajov podieľajú. Koordinátor otvárania údajov zodpovedá za:

- prípravu publikačného plánu,
- návrh kurátorov dát pre jednotlivé súbory údajov,
- poskytovanie metodologickej podpory kurátorom dát pri plnení povinností v oblasti otvorených údajov,
- zverejnenie publikačného plánu,
- rozhodnutie o spôsobe katalogizácie (centrálny katalóg otvorených údajov alebo lokálny katalóg otvorených údajov),
- zabezpečenie finálnej kontroly údajov pred publikáciou,
- zabezpečenie publikácie súborov údajov a katalogizačných záznamov (metaúdajov),
- komunikáciu pripomienok užívateľov otvorených údajov a reportovanie o publikácii otvorených údajov
- vedenie poskytovateľa údajov a príp. tlačovému oddeleniu poskytovateľa údajov.

Okrem koordinátora sú v procese publikácie informácií verejného sektora dôležité aj role kurátora dát, správca katalógu otvorených údajov, IT špecialisty. kurátor dát zodpovedá za:

- analýzy súborov údajov predchádzajúce ich zverejneniu;
- zaistenie, že otvorené súbory údajov budú publikované v súlade s platnými právnymi predpismi SR a štandardmi publikácie a katalogizácie otvorených údajov);
- určenie prípadnej nutnosti transformácie súborov údajov, napr. anonymizáciu, agregáciu alebo inú zmenu štruktúry súborov údajov;
- správu a kontrolu kvality súborov údajov (t.j. kontrola správnosti, úplnosti, konzistentnosti) pred ich publikáciou a katalogizáciou;
- spoluprácu pri procese zverejnenia súboru údajov s koordinátorom otvárania údajov a IT špecialistom;
- odovzdanie súborov údajov
- návrh obsahu katalogizačných záznamov (metaúdajov) správcovi katalógu otvorených údajov.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

Kurátor dát je povinný poskytnúť IT špecialistovi a Správcovi katalógu otvorených údajov potrebnú súčinnosť pri:

- transformácii súboru údajov a kontrole;
- určeníu prácnosti a prínosov plynúcich z publikácie otvorených súborov údajov;
- vytváraniu a aktualizáciu katalogizačných záznamov.

Správca katalógu otvorených údajov zodpovedá za:

- formálnu kontrolu a správu katalogizačných záznamov (metaúdajov);
- vkladanie katalogizačných záznamov a príp. aj súborov údajov do centrálného katalógu otvorených údajov;
- správu katalogizačných záznamov v centrálnom katalógu otvorených údajov.

V prípade, že bude v organizácii implementovaný aj lokálny katalóg otvorených údajov, Správca katalógu otvorených údajov ďalej zodpovedá za:

- zabezpečenie výberu softvérového nástroja pre lokálny katalóg otvorených údajov;
- zabezpečenie implementácie lokálneho katalógu otvorených údajov;
- registráciu lokálneho katalógu v centrálnom katalógu otvorených údajov;
- archiváciu katalogizačných záznamov a údajov uložených v lokálnom katalógu otvorených údajov.

IT špecialista zodpovedá za:

- poskytovanie technickej podpory a súčinnosti kurátorovi dát pri plnení povinností v oblasti otvorených údajov, najmä pri analýze predchádzajúcej zverejneniu údajov a príprave súborov údajov na otvorenie;
- prípadné zaistenie inštalácie a prevádzky lokálneho katalógu otvorených údajov.

5.1 Identifikácia súborov údajov na otvorenie

Vypracovanie plánu zverejňovania otvorených údajov začína analýzou zdrojov údajov spravovaných organizáciou a identifikáciou potenciálnych súborov údajov obsiahnutých v zdrojoch údajov. Túto úlohu je možné preskočiť, keď sa organizácia už rozhodla, že sa budú otvárať len konkrétne súbory údajov a je jasné, kde, kto a ako súbory údajov v organizácii spravuje. V opačnom prípade odporúčame postupovať podľa nasledujúcich krokov:

1. Analýza organizačnej štruktúry organizácie, jej predpisy a agendy – Cieľom je identifikovať činnosti jednotlivých organizačných jednotiek, ktoré súvisia so zberom, tvorbou alebo správou niektorých údajov. Identifikované súbory údajov je potrebné zaznamenať do zoznamu súborov údajov.

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

2. Analýza výročných správ organizácie a ďalších verejných dokumentov (vrátane webových portálov organizácie), ktoré informujú o činnosti a výsledkoch organizácie – Cieľom je identifikovať v dokumentoch tabuľky a grafy, ktoré informujú o potenciálnych súboroch údajov. Nové identifikované súbory údajov je potrebné zaznamenať do zoznamu súborov údajov.
3. Identifikácia informačných systémov v organizácii – Cieľom je identifikovať potenciálne súbory údajov spravované systémami. Identifikované súbory údajov v informačných systémoch je potrebné zaznamenať do zoznamu súborov údajov.
4. Analýza žiadostí o informácie odoslaných organizácii verejnosťou – Cieľom je identifikovať súbory údajov, ktoré sú pre verejnosť zaujímavé, a porovnať ich s predtým identifikovanými súbormi údajov. Nové identifikované súbory údajov je potrebné zaznamenať do zoznamu súborov údajov.
5. Vytvorenie mapy, ktorá zobrazuje organizačné jednotky a súbory údajov, za ktoré sú zodpovedné. Mapa by mala byť graficky znázornená.

Pre každý identifikovaný súbor údajov je potrebné zaznamenať nasledujúce informácie:

- názov a popis,
- zodpovedná organizačná jednotka,
- kontaktná osoba (pre konzultácie o súbore údajov),
- aktuálny formát súboru údajov (uložená v relačnom alebo inom databázovom systéme; uložená ako tabuľkový súbor alebo súbory, vo formáte XLS(X), ODS atď.; formát XML; formát CSV; proprietárny tabuľkový formát; iba v neštruktúrovanej alebo pološtruktúrovanej textovej forme) a stručný popis.

Mapa dostupných zdrojov údajov, súborov údajov a zodpovedných organizačných jednotiek slúži ako vstup pre identifikáciu vhodných súborov údajov na otvorenie. Cieľom tejto úlohy je rozhodnúť, ktoré z kandidátskych súborov údajov by sa mali zverejniť ako otvorené údaje a identifikovať súbory údajov, ktoré nemožno otvoriť.

Otvorený súbor údajov je akýkoľvek typ údajov, ktoré sa sprístupňujú verejnosti pre rôzne ciele, ako je ekonomický rast, lepšie vládne služby, zvýšená transparentnosť a lepšia dôvera a postoje k orgánom verejného sektora. Rozpočtové údaje, údaje o nákupoch a verejných zákazkách môžu pomôcť zvýšiť transparentnosť, zatiaľ čo údaje o výkonnosti verejného sektora môžu pomôcť vybudovať dôveru. Údaje o doprave, informácie o verejných službách a údaje o životných udalostiach sú príkladmi typov údajov, ktoré by sa mali sprístupniť verejnosti. Určité štúdie, ako napríklad index otvorených údajov (*Open Data Index*) a barometer otvorených údajov (*Open Data Barometer*), môžu poskytnúť návod, ktoré súbory údajov sú bežne požadované. Pri zvažovaní súboru údajov na zverejnenie, ak je už publikovaná index, je potrebné skontrolovať, či nejde o duplikát, a ak je už zverejnená len časť súboru údajov, môže byť

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

možné rozdeliť súbor údajov a zverejniť primárne údaje s odkazmi na externý zdroj v zázname katalógu alebo uplatnením princípov prepojených údajov.

Koordinátor otvárania údajov nepozná vecný obsah jednotlivých súborov údajov, a preto nie je často schopný kvalifikovane rozhodnúť, aké súbory údajov by mali byť zahrnuté do publikačného plánu. Vecnú znalosť a zodpovednosť za súbory údajov nesú kurátori dát. Tí by teda mali aj rozhodovať o umiestnení súborov údajov do publikačného plánu. Kurátor dát navrhuje také súbory údajov, za ktoré v rámci svojej bežnej pracovnej činnosti nesie vecnú zodpovednosť (tzn., vytvára ich, koordinuje ich zber a pod.). Postup by preto mal byť:

1. Od koordinátora otvorených údajov dostane kurátor dát zoznam odporúčaných súborov údajov na otvorenie.
 2. Kurátor dát vyberie v zozname súbory údajov, ktoré spadajú do jeho kompetencie.
 3. Kurátor dát z nich vyberie tie súbory údajov, ktoré je možné zaradiť do publikačného plánu a zverejniť v podobe otvorených údajov (aj keď v tom vidia riziká).
 4. Kurátor dát analyzuje, či nespravuje ďalšie súbory údajov, ktoré by bolo možné publikovať v podobe otvorených údajov a pritom nie sú v zozname odporúčaných súborov údajov uvedené. Novoidentifikované súbory údajov umiestni na zoznam.
 5. Kurátor dát pre každý vybraný súbor údajov doplní nasledujúce atribúty:
 - názov súboru údajov,
 - popis súboru údajov,
 - Kurátor dát.
 6. Kurátor dát odovzdá zoznam súborov údajov koordinátorovi otvárania údajov a pomáha mu s analýzou prínosov a rizík otvorenia jednotlivých súborov údajov, pokiaľ o to koordinátor otvárania údajov požiada.

Pri návrhu súborov údajov môže kurátor dát požiadať o spoluprácu IT špecialistu. IT špecialista by mal byť schopný popísať, aké informačné systémy a databázy sú pre agendu spravovanú kurátorom dát prevádzkované a aké údaje sú v nich uchovávané. Túto informáciu by mal kurátor dát zvážiť a analyzovať, či nejde o ďalšie súbory údajov, ktoré by bolo možné otvoriť. Pri vytváraní zoznamu odporúčame postupovať nasledovne:

- Koordinátor otvárania údajov odovzdá jednotlivým kurátorom dát zoznam odporúčaných súborov údajov, aby vybrali súbory údajov, ktoré spadajú do ich kompetencií a je možné ich zverejniť v podobe otvorených údajov, a prípadne zoznam rozšírili o ďalšie súbory údajov v ich kompetencii.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- Koordinátor otvárania údajov konsoliduje súbory údajov vybrané kurátormi dát do jedného zoznamu.
- Koordinátor otvárania údajov kontroluje, že každý súbor údajov v zozname je doplnený o nasledujúce atribúty. Pokiaľ nie sú, komunikuje s príslušnými kurátormi dát o ich doplnení:
 - názov súboru údajov,
 - popis súboru údajov,
 - Kurátor dát.

5.1.1 Stanovenie cieľovej úrovne otvorenosti

1. Určite cieľovú úroveň otvorenosti podľa 1 – 5* schémy pre každý súbor údajov. Minimálna úroveň je 3*.
 - Úroveň 2* je možné určiť v špeciálnych prípadoch, keď súbor údajov existuje len vo forme neštruktúrovaných dokumentov a organizácia ich nemôže previesť do štruktúrovanej podoby.
2. Vyberte si jeden alebo viac formátov údajov na publikovanie každého súboru údajov na základe určenej úrovne otvorenosti:
 - Pre 2* úroveň odporúčame formáty tabuľkových editorov (ODS, XLSX atď.) alebo HTML. V prípade textových dokumentov (napr. zmluvy o verejnom obstarávaní) je možné zvoliť formáty textových editorov (ODT, DOC(X) atď.).
 - Pre 3* úroveň odporúčame CSV, XML alebo JSON.
 - Pre 4* úroveň je možné zvoliť RDF, čo je tiež odporúčaný formát pre 5* úroveň. Je tiež možné zvoliť CSV, XML alebo JSON. Kľúčovou charakteristikou úrovne 4* v porovnaní s úrovňou 3* je, že entity v súbore údajov sú identifikované pomocou URL, takže je možné na ne transparentne odkazovať z iných súborov údajov. Odporúčame zaznamenávať adresy URL vo formátoch CSV, XML a JSON nasledujúcim spôsobom:
 - Pre CSV odporúčame pridať nový stĺpec pre adresy URL entít. Stĺpec by mal byť umiestnený vedľa už existujúceho stĺpca pre identifikátory entity.
 - V prípade XML odporúčame zaznamenávať adresy URL pomocou rozšírenia dokumentov HTML a XML s názvom RDFa (zdroj atribútov XML).
 - Pre JSON odporúčame zaznamenávať adresy URL pomocou rozšírenia JSON s názvom JSON-LD (konštrukcia @id).
 - Pre 5* úroveň, t. j. prepojené otvorené údaje reprezentované v modeli RDF, odporúčame TTL. Kľúčovou charakteristikou úrovne 5* v porovnaní s úrovňou 4* je, že neposkytujeme iba identifikátory entít vo forme adres URL, ale spájame ich

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

aj s adresami URL iných súvisiacich entít v iných súboroch údajov. Je tiež potrebné zabezpečiť, aby každá URL našich zverejnených entít bola dereferencovateľná, t. j. aby klientska aplikácia dostala strojovo čitateľnú reprezentáciu entity v modeli RDF, keď pristúpi na URL.

3. Určite cieľovú periodicitu zverejňovania aktualizácií súboru údajov.
 - Aktualizácie sa nemusia zverejňovať tak často, ako sa objavujú zmeny zdrojových údajov. Napríklad, aj keď sa zdrojové údaje menia každú hodinu, môžeme určiť, že aktualizácie budú zverejňované týždenne. Inými slovami, cieľová periodicitu je jeden týždeň. Čím je cieľová periodicitu bližšie k frekvencii zmien, tým vyššia je kvalita súboru údajov. Na druhej strane vyššia cieľová periodicitu môže viesť k vyšším nákladom.
 - Je možné reflektovať spätnú väzbu od verejnosti, aby sme identifikovali očakávanú cieľovú periodicitu alebo overili, či je určená periodicitu dostatočná pre potenciálnych používateľov súboru údajov.
4. Rozhodnite, ako bude súbor údajov zverejnený. V zásade by mal byť zverejnený súbor údajov pre každý zvolený dátový formát. Mal by obsahovať položky, ktoré sú platné v čase zverejnenia súboru údajov.
 - Ak je súbor údajov príliš veľký, možno ho rozdeliť na viac súborov údajov.
5. Rozhodnite, či budú zverejnené historické verzie súboru údajov. Ak áno, rozhodnite, koľko verzií späť do histórie bude zverejnených.
6. Ak je to technicky možné, odporúča sa zverejniť zoznam zmien vykonaných v súbore údajov medzi aktuálnou a predchádzajúcou verziou súboru údajov.
7. Ak je to technicky a finančne možné, odporúča sa publikovať nielen súbory údajov, ale aj API, ktoré umožňuje priamy prístup ku konkrétnym položkám (entitám) datasetu:
 - Pre 3* a 4* by API malo byť službou REST.
 - Pre 5* by API malo byť koncovým bodom SPARQL.

Formáty súborov sú štandardizované spôsoby kódovania informácií na použitie počítačmi. Formát popisuje, ako je súbor reprezentovaný pri ukladaní. Súbory končia príponami (napríklad: .mp3, .html, .doc), ktoré sú zvyčajne skrátenými verziami názvu formátu.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

Tabuľka 3: Bežné formáty súborov otvorených údajov

Formát	Popis
.xml (<i>Extensible Markup Language</i>)	Jazyk na popis textových údajov pomocou značiek. XML môžu čítať ľudia aj stroje a je založené na otvorených a bezplatných štandardoch.
.csv (<i>Comma Separated Values</i>)	Ľudsky čitateľný formát používaný na ukladanie tabuľkových údajov ako textu, ktorému rozumie mnoho rôznych programov, a často sa používa ako spôsob interakcie s informáciami v softvérových aplikáciách (výmena údajov).
.json (<i>JavaScript Object Notation</i>)	Ďalší ľudsky čitateľný formát na výmenu údajov, ktorý ukladá informácie o objektoch v „pároch“ názvu atribútu a hodnoty.
.pdf (<i>Portable Document Format</i>)	Veľmi bežný a otvorený (od roku 2008) formát na prezentáciu dokumentov nezávislý od aplikácie alebo zariadení používaných na prístup k nemu. Súbor .pdf obsahuje všetky informácie potrebné na to, aby ho čítačka pdf mohla zobraziť. PDF môže alebo nemusí byť strojovo čitateľné.
.rdf (<i>Resource Description Framework</i>)	Štandardizovaný rámec pre konceptuálne modelovanie sémantických (človekom čitateľných) informácií takmer o čomkoľvek.
.kml (<i>Keyhole Markup Language</i>)	Špeciálna verzia alebo rozšírenie XML používaného špeciálne pre geografické informácie a prepojenie súradníc na mape s informáciami, ako je konkrétne miesto alebo atribút. Pôvodne vyvinutý spoločnosťou Google pre aplikáciu Google Earth.
.doc(x) (<i>Open Office XML</i>)	Rozšírenie XML na zobrazovanie a úpravu dokumentov Microsoft Word. Dá sa prečítať niektorými inými programami.
.xls(x) (<i>Open Office XMP Workbook</i>)	Rozšírenie XML na zobrazovanie dokumentov Microsoft Excel, ktoré môžu čítať niektoré iné programy.
.shp, .shx, .dbf (<i>Shapefile</i>)	Shapefiles sú skupiny súborov používané na popis geopriestorových prvkov z hľadiska bodov, čiar a oblastí (polygónov). Často prichádzajú v komprimovanom priečinku (.zip) a nie sú čitateľné pre ľudí, pokiaľ nepracujú s programom geografického informačného systému (GIS).

5.1.2 Definícia plánu zverejňovania otvorených údajov

Plán zverejňovania otvorených údajov predstavuje konečný výstup procesu prípravy zverejňovania otvorených údajov. Informácie o potenciálnych otvorených súboroch údajov sú zostavené do konečného katalógu kandidátskych súborov údajov na otvorenie. Mala by sa vykonať prioritizácia súborov údajov a podľa stanovených priorít by sa mal stanoviť plán zverejňovania otvorených údajov. Napokon plán zverejňovania otvorených údajov by mal obsahovať popis úloh zainteresovaných strán zapojených do zverejňovania otvorených údajov a zodpovednosti týchto rolí by mali byť stanovené a zdokumentované v pláne zverejňovania otvorených údajov. Definícia plánu

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

zverejňovania otvorených údajov by mala byť koordinovaná s definíciou plánu manažmentu benefitov, ako aj plánom na zmiernenie rizík. Preto by sa definícia týchto plánov nemala vykonávať oddelene od seba, ale ich definícia by sa mala zosúladiť a harmonizovať.

Ak nie je možné zverejniť všetky kandidátske súbory údajov naraz, zverejnenie by malo byť rozdelené do niekoľkých vydaní podľa stanovených priorít. Stanovenie priority súborov údajov na otvorenie môže zahŕňať nasledujúce kroky:

1. Určite, aké kľúčové atribúty súboru údajov sa použijú ako kritériá priority.
2. Určite váhy kritérií priorít.
3. Odstráňte súbory údajov, ktoré nemožno publikovať ako otvorené údaje, z kandidátskych katalógov údajov.
4. Vypočítajte priority na základe kritérií priorít a ich váhy. Zoradte súbory údajov kandidátov podľa vypočítaných priorít.
5. Presuňte súbory údajov nahor alebo nadol v zozname, ak máte pocit, že vypočítané skóre neodráža ciele zverejnenia otvorených údajov alebo skutočnú hodnotu súborov údajov. Prípadne môžete upraviť váhy kritérií priority a prepočítať skóre.

Pri stanovovaní priorít, ktoré súbory údajov sa majú zverejniť, je dôležité zvážiť aktuálny dopyt po údajoch, aktuálne formáty, dostupnosť dokumentácie schémy, očakávané prínosy, odhadované riziká a odhadované úsilie pri publikovaní súborov údajov. Na určenie dopytu je možné vykonať prieskum alebo analyzovať frekvenciu žiadostí o informácie, aby ste zistili, či existuje vysoký dopyt po určitých súboroch údajov. Pokiaľ ide o súčasné formáty, mali by sme zvážiť náklady spojené s transformáciou údajov do strojovo čitateľného formátu, pretože publikovanie už strojovo čitateľných súborov údajov bude rýchlejšie a menej nákladné. Dostupnosť dokumentácie schémy je tiež niečo, čo treba zvážiť, pretože môže uľahčiť opätovné použitie súborov údajov. Okrem toho by sa pri výbere súborov údajov mali brať do úvahy aj možné výhody, ktoré môžu poskytnúť otvorené údaje, ako sú lepšie verejné služby a lepšia transparentnosť. Je tiež dôležité zvážiť možné riziká, ktoré so sebou prináša zverejňovanie otvorených údajov, ako je riziko narušenia súkromia a riziko zneužitia údajov. Nakoniec je dôležité zvážiť odhadované úsilie pri publikovaní určitých súborov údajov, pretože by sa mali brať do úvahy súbory údajov, ktoré môžu mať vysokú hodnotu, ale vyžadujú si značné úsilie na zverejnenie.

Pri zostavení harmonogramu odporúčame postupovať nasledovne:

- Kurátori dát odhadnú prácnosť otvorenia súborov údajov, ktoré vybrali na otvorenie.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- Odporúčame pásmový odhad prácnosti, napr. nízka prácnosť (do 1 pracovného dňa), stredná prácnosť (do 1 pracovného týždňa), vysoká prácnosť (viac ako 1 pracovný týždeň)
- Koordinátor otvárania údajov zotriedi zoznam súborov údajov podľa ich priority:
 - Prioritu určí s ohľadom na prácnosť a na výsledky analýzy prínosov a rizík otvorenia súboru údajov;
 - Najvyššiu prioritu majú súbory údajov s nízkou prácnosťou, najvyššími prínosmi a najnižšími rizikami;
 - Koordinátor otvárania údajov môže s vedením komunikovať, aké sú preferované prínosy a najobávanejšie riziká.
- Koordinátor otvárania priradí s ohľadom na priority dátum, v ktorom bude súbor údajov otvorený.
 - Postačí, keď si zvolia 2 dátumy (napr. za 1 mesiac a za 6 mesiacov), alebo uvedie všeobecnejšiu informáciu, že k zverejneniu dôjde neskôr (u tých súborov údajov, u ktorých nie je možné v súčasnosti určiť presný dátum).

Výsledný harmonogram publikácie vybraných súborov údajov vytvára publikačný plán, ktorý bude postúpený vedeniu poskytovateľa údajov na schválenie. Publikačný plán ešte nie je v tejto chvíli kompletný, pretože neobsahuje pre súbory údajov technické parametre (stupeň otvorenosti a dátový formát), ktoré však nie sú pre vedenie poskytovateľa údajov dôležité. Koordinátor otvárania údajov by tiež mal periodicky vyhodnocovať plnenie publikačného plánu (odporúčame raz ročne). Pri vyhodnotení by sa mal venovať predovšetkým nasledujúcemu:

- Podarilo sa otvoriť všetky súbory údajov podľa predpísaného harmonogramu?
- Ak sa niektoré súbory údajov otvoriť nepodarilo, prečo?
- Podarilo sa otvoriť niektoré súbory údajov nad rámec predpísaného harmonogramu?
- Existujú používatelia už otvorených súborov údajov?
- Vznikli nad otvorenými súbormi údajov nejaké aplikácie?
- Aké boli náklady na otvorenie súborov údajov?

Vyhodnotenie je možné zdokumentovať vo výročnej správe.

5.2 Príprava súborov údajov na publikovanie

Ak je potrebné zaznamenať metaúdaje o pôvode súboru údajov, organizácia by mala použiť W3C slovník PROV Data Model, najmä PROV Ontology (PROV-O), na zachytenie informácií týkajúcich sa procesu vytvárania údajov, krokov v rámci procesu

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

a kto je za kroky zodpovedný. Na uchovanie týchto informácií sa odporúča, aby organizácia používala svoj interný katalóg údajov, ktorý obsahuje používateľské rozhranie na úpravu katalógových záznamov podľa schémy DCAT alebo VoID a má možnosť exportovať tieto katalógové záznamy ako LOD (5* údaje) vo formátoch TTL a JSON-LD. Okrem toho by mali používať DCAT aj na opis metaúdajov miestneho katalógu verejných údajov organizácie. Okrem toho DCAT a VoID poskytujú konštrukcie na rozlíšenie medzi metaúdajmi pre súbor údajov ako celok a dodatočnými metaúdajmi pre konkrétne súbory údajov, známe ako distribúcie, ktoré obsahujú údaje v súbore údajov.

Pri použití DCAT sú katalógové záznamy (ale nie samotné súbory údajov) pripravené na publikovanie ako 5* otvorené údaje. Inými slovami, katalógové záznamy sú pripravené na publikovanie ako Linked Open Data.

Správna aplikácia slovníkov DCAT a VoID je pomerne zložitý problém. Odporúča sa postupovať podľa podrobných špecifikácií tých slovníkov publikovaných W3C:

- DCAT [[Odkaz](#)]
- VoID [[Odkaz](#)]

Taktiež je potrebné navrhnuť URL adresy každého katalógového záznamu a ďalších súvisiacich entít. Tieto adresy URL sa používajú na identifikáciu týchto entít podľa princípov Linked Open Data. Tiež sa odporúča, aby organizácia prevádzkovala svoj vlastný katalóg údajov s katalógovými záznamami vlastných súborov údajov. Ak v krajine organizácie existuje centrálny katalóg údajov, je potrebné zverejňovať katalógové záznamy aj v tomto katalógu.

- Ak centrálny katalóg podporuje DCAT (a VoID) a automatizovaný import katalógových záznamov, odporúčame využiť túto funkčnosť na automatické kopírovanie katalógových záznamov z katalógu údajov organizácie do centrálného.
- Ak centrálny katalóg nepodporuje DCAT (a VoID) alebo automatizovaný import katalógových záznamov, odporúčame organizácii požiadať o túto funkčnosť správcu centrálného katalógu.

5.2.1 Opis súborov údajov

Vytvorenie katalógového záznamu je nevyhnutným krokom v publikačnom procese organizácie. Záznam by mal byť pripravený a uložený v internom katalógu údajov organizácie v súlade s navrhnutou schémou katalógových záznamov. Príprava katalógového záznamu zahŕňa vyplnenie hodnôt tých atribútov známych v čase návrhu a nebude zahŕňať tie atribúty, ktoré budú známe len v čase publikovania, pretože tieto môžu byť automaticky vyplnené, ak sa na automatizáciu procesu publikovania použije softvérový nástroj.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

Okrem toho sa pri práci s 3* a 4* údajmi odporúča zostaviť dátovú schému pre súbor údajov, pretože to pomáha vývojárom aplikácií a ostatným používateľom pracovať so súborom údajov efektívnejšie a efektívnejšie. Pomáha tiež znížiť možnosť nesprávnej interpretácie súboru údajov. Schéma údajov by mala zahŕňať štruktúru a sémantiku položiek súboru údajov. Spôsob zverejnenia dátovej schémy závisí od formátu údajov zvoleného na publikovanie súboru údajov. Odporúčame nasledovné:

1. Na základe zvoleného dátového formátu určite jazyk na vyjadrenie dátových schém.
 - Pre formát CSV zvolte Metadata Vocabulary for Tabular Data.
 - Pre formát XML zvolte DTD alebo XML Schema.
 - Pre formát JSON zvolte Schema JSON.
2. Vytvorte schému údajov pre každý súbor údajov a formát, v ktorom súbor údajov publikujete. Pokúste sa vysvetliť čo najviac sémantiky formou opisov a komentárov. Použitie správnych konštrukcií zvoleného jazyka schémy na špecifikáciu popisov a komentárov.
3. Nezávisle od zvoleného jazyka schém špecifikujte dátové typy primitívnych hodnôt pomocou primitívnych typov jazyka XML Schema.
4. Zverejnite každú schému údajov ako samostatný súbor údajov. Na prepojenie použite vhodné konštrukcie daného jazyka.
5. Ak existuje lokálny katalóg údajov organizácie a poskytuje webovú stránku pre každý súbor údajov, umiestnite na webovú stránku odkaz na schému údajov pre súbor údajov.

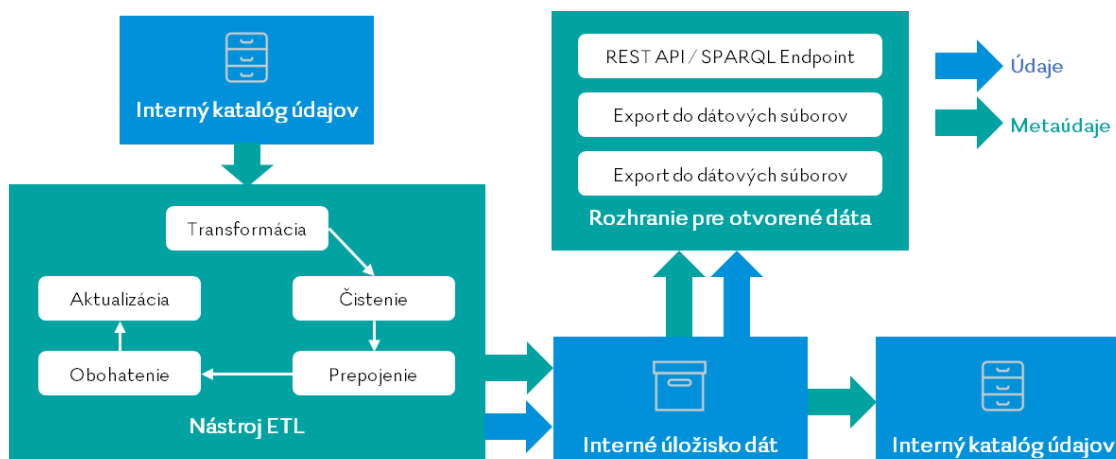
5.2.2 Výber a implementácia softvérových nástrojov

Pre publikovanie súborov údajov sa odporúča zvoliť niekoľko softvérových nástrojov, ktoré implementujú komponenty referenčnej architektúry znázornenej na obrázku nižšie. Tieto softvérové nástroje uľahčia a zautomatizujú proces publikovania súborov údajov.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

Obrázok 1: Referenčná architektúra softvérových nástrojov na publikovanie otvorených údajov



Komponenty referenčnej architektúry sú nasledovné:

- **Interný katalóg údajov:** Komponent na zaznamenávanie všetkých identifikovaných súborov údajov a napĺňanie ich katalógových záznamov. Nie je to verejný katalóg údajov. Obsahuje záznamy o súboroch údajov, ktoré budú zverejnené niekedy v budúcnosti a tiež o tých, kde si organizácia ešte nie je istá, či budú zverejnené. Môže byť implementovaný vo forme tabuľkového súboru a upravovaný pomocou tabuľkového editora. Odporúčame však zvoliť nástroj na katalogizáciu súborov údajov, ktorý podporuje DCAT.
- **Nástroj ETL** (Extract-Transform-Load): Softvérový nástroj, ktorý umožňuje konfigurovať takzvané ETL procedúry. Procedúra ETL pripraví súbor údajov uvedený v internom katalógu údajov na zverejnenie. Časť E (Extract) zhromažďuje údaje z pripravených zdrojov údajov. Časť T (Transformácia) implementuje súbor transformácií, vrátane konverzie dátových formátov, čistenie údajov, anonymizácia, prepojenie a obohatenie. Časť L (Load) obohatí pripravený súbor údajov o metaúdaje (katalógový záznam) a načíta súbor údajov do interného úložiska údajov.
- **Interné úložisko údajov:** Priestor, kde sa ukladajú pripravené súbory údajov. Môže to byť lokálny súborový systém organizácie, vhodný databázový server alebo kombinácia oboch. Závisí to od zvolených dátových formátov a spôsobu sprístupnenia datasetov koncovým používateľom.
- **Rozhranie pre otvorené údaje:** Sprístupňuje publikované súbory údajov koncovým používateľom. Konektory katalógov údajov exportujú záznamy katalógu do vybraných katalógov údajov. Jedným z nich je verejný katalóg údajov organizácie, ak ho organizácia má. Export do súborov údajov prevezme súbory

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

údajov z interného úložiska údajov a ponúkne ich koncovým používateľom na hromadné sťahovanie. REST API / SPARQL Endpoint sú služby, ktoré poskytujú vývojárom aplikácií priamy programový prístup k súborom údajov.

- **Katalóg údajov organizácie:** Webová stránka alebo súbor štruktúrovaných webových stránok.

5.2.3 Definícia prístupu k publikovaniu súboru údajov

Konkrétne kroky potrebné na prípravu a zverejnenie môžu závisieť od povahy súboru údajov. Vo všeobecnosti by sa však malo analyzovať, ako sa budú vykonávať nasledujúce činnosti a kroky:

- transformácia a čistenie súborov údajov,
- anonymizácia súborov údajov,
- aktualizácia súborov údajov.

V prípade prepojených údajov by mal byť definovaný prístup k integrácii údajov a prepojenie na iné súbory údajov.

- Prístup k čisteniu údajov** - Čistenie údajov zahŕňa niekoľko základných prístupov určených na zlepšenie kvality a spoľahlivosti súboru údajov. Po prvé, audit údajov používa štatistické metódy na nájdenie a vyhodnotenie abnormálnych alebo poškodených údajov. Ďalej validácia údajov testuje konzistentnosť údajov a kontroluje prítomnosť obmedzení integrity. Nesprávne údaje sa potom spracúvajú prostredníctvom procesu úpravy údajov, ktorý zahŕňa úpravu alebo vymazanie takýchto údajov. Transformácia údajov to ďalej spresňuje úpravou údajov tak, aby sa zmestili do bežného štandardizovaného formátu, aby sa zabezpečila jednotnosť. V prípade neúplných súborov údajov doplní imputácia údajov chýbajúce hodnoty na základe stanovených údajov z podobných súborov údajov alebo iných relevantných zdrojov údajov. Na uľahčenie analýzy nepretržitých nespracovaných údajov ich diskretizácia údajov transformuje na konečný súbor intervalov. Nasleduje deduplikácia údajov, ktorá zahŕňa odstránenie duplicitných alebo opakovaných záznamov. Pre väčšiu hĺbku a kontext, obohatenie údajov zlepšuje existujúce údaje integráciou informácií z dodatočných zdrojov údajov. Zovšeobecnenie údajov je tiež možnosťou, keď sa špecifické údaje nahrádzajú širšími kategóriami na účely zabezpečenia alebo počítačovej analýzy. Napokon, pri použití viacerých zdrojov údajov ich harmonizácia presúva na spoločnú platformu, ktorá zabezpečuje kompatibilitu a porovnateľnosť. Tieto kumulatívne procesy zabezpečujú vyššiu kvalitu, hodnotnejšiu a spoľahlivejšiu súbor údajov.
- Prístup k anonymizácii údajov** - Anonymizácia údajov zahŕňa rôzne metodiky, ktoré pomáhajú zachovať dôvernú citlivých informácií. Patrí medzi ne zovšeobecňovanie, kde sa podrobné údaje transformujú do širších kategórií, a pseudonymizácia, pri ktorej sa identifikovateľné údajové polia nahrádzajú fiktívnymi identifikátormi. Môžu sa použiť aj techniky ako výmena údajov alebo permutácia, kde sa údaje medzi záznamami

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

vymieňajú, aby sa pomohli chrániť jednotlivé identity. Porucha údajov mierne modifikuje pôvodné údaje pri zachovaní podobných štatistických vlastností. Ďalšou technikou je pridávanie šumu, kde sa pridáva náhodný šum, aby sa zamaskovali skutočné jednotlivé položky. Agregácia sa často používa na kombinovanie údajov takým spôsobom, že jednotlivé položky sú nerozoznateľné. Maskovanie údajov, metóda primárne používaná pri testovaní databáz, zahmlieva skutočné údaje demonštratívny, ale pozmenenými údajmi. Syntetické súbory údajov sú úplne novo vygenerované, ale stále si zachovávajú štatistickú podobnosť s originálom. Využíva sa aj princíp k-anonymity, kedy je uvoľňovanie údajov riadené tak, aby každá entita bola na nerozoznanie od minimálne k-1 iných v rámci súboru údajov. Každý z týchto prístupov chráni identitu jednotlivcov a zároveň umožňuje vykonávať dôležité analýzy údajov.

- c) **Prístup k integrácii údajov:** Integrácia údajov zahŕňa niekoľko kľúčových prístupov na konsolidáciu údajov z rôznych zdrojov do jedného jednotného zobrazenia. Manuálna integrácia, hoci je časovo náročná a náchylná na chyby, je jednou z metód, ktorá zahŕňa manuálne zadávanie údajov z rôznych zdrojov do jednej databázy. Skladovanie údajov pomocou operácií Extrahovať, Transformovať, Načítať (ETL) je bežnejšou technikou, kde sa údaje extrahujú z viacerých zdrojov, transformujú sa do spoločného formátu a načítajú sa do dátového skladu. Pre jednotný prístup k heterogénnym zdrojom údajov možno použiť federáciu údajov. Middlevérové nástroje ponúkajú iný prístup, ktorý umožňuje používateľom pristupovať a manipulovať s údajmi z rôznych zdrojov, ako keby boli umiestnené na jednom mieste. Podobný koncept, virtualizácia údajov, poskytuje jednotné používateľské rozhranie, ktoré predstavuje údaje z viacerých zdrojov ako súčasť jedného jednotného zobrazenia. Šírenie údajov zahŕňa v reálnom čase alebo plánovanú replikáciu a synchronizáciu údajov z rôznych zdrojov. Master Data Management (MDM) poskytuje konzistentné zaobchádzanie s kľúčovými údajmi a ponúka jediný referenčný bod pre kľúčové informácie. Prepojenie záznamov sa používa na nájdenie záznamov, ktoré odkazujú na rovnakú entitu v rôznych súboroch údajov. Mapovanie a zlučovanie údajov sú tiež bežné postupy, ktoré zahŕňajú identifikáciu toho, ako sa údaje z jedného modelu mapujú na iný, a následne kombinovanie viacerých súborov údajov do jedného, čo si často vyžaduje deduplikáciu na odstránenie nadbytočných údajov. Vo všetkých týchto prístupoch zostáva primárnym cieľom dátovej integrácie jednotný, konzistentný pohľad na údaje naprieč všetkými zdrojmi.
- d) **Prístup k aktualizácii údajov:** Aktualizácia údajov zahŕňa rôzne prístupy v závislosti od systémových požiadaviek, objemu údajov, prevádzkových potrieb a kritickosti údajov. Dávková aktualizácia je jednou z bežných metód, ktorá zahŕňa zhromažďovanie zmien údajov a následnú aktualizáciu celej databázy alebo súboru údajov v čase mimo špičky, čo je obzvlášť výhodné pre veľké súbory údajov. Aktualizácia online alebo v reálnom čase umožňuje okamžité vykonanie aktualizácií, keď dôjde k zmenám. Hoci táto metóda ponúka najaktuálnejšie informácie, môže ovplyvniť výkon systému. Na druhej strane, prírastková aktualizácia upravuje iba prvky údajov, ktoré boli zmenené alebo pridané, čím šetrí systémové prostriedky. Plánovaná aktualizácia nastavuje konzistentné aktualizácie v určených intervaloch, napríklad denne alebo týždenne, čím poskytuje pravidelné a predvídateľné obnovovanie údajov. Prístup známy ako aktívna aktualizácia údajov používa spúšťače na automatické spustenie aktualizácií, keď sa zistia

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

špecifikované zmeny údajov. Delta aktualizácia zahŕňa počiatočné úplné zálohovanie údajov s pravidelnými zálohami iba zmien od poslednej úplnej zálohy. Automatizované nástroje a skripty sa široko používajú na aktualizáciu údajov a ponúkajú presnosť a konzistentnosť prostredníctvom plánovaných operácií. Manuálna aktualizácia, aj keď je časovo náročná, umožňuje zmeny a aktualizácie údajov riadené osobou. Aktualizácie riadené používateľom umožňujú koncovým používateľom upravovať ich vlastné údaje, čím sa zvyšuje rýchlosť procesu a potenciálne sa zvyšuje presnosť údajov. A napokon, rozhrania API na aktualizáciu údajov sa bežne používajú v moderných aplikáciách a databázach na programovú aktualizáciu údajov. Kombinácia týchto prístupov umožňuje udržiavať údaje včasné a presné.

5.3 Publikácia súborov údajov

Iniciatívy otvorených údajov, ako napríklad Európska štúdia o zrelosti otvorených údajov, sa zameriavajú na sprístupnenie súborov údajov s vysokou hodnotou (*High-Value Datasets*, HVD). Orgány verejného sektora môžu mať obmedzené zdroje, takže si musia uprednostniť, ktoré súbory údajov zverejnia. HVD ponúkajú dôležité výhody pre spoločnosť, životné prostredie a ekonomiku. Môžu byť použité na vytváranie služieb s pridanou hodnotou, aplikácií a nových pracovných miest. Smernica EÚ o otvorených údajoch identifikuje kritériá na určovanie HVD, ako je ich potenciál vytvárať sociálno-ekonomické alebo environmentálne výhody, prinášať výhody mnohým používateľom (najmä MSP), generovať príjmy a môžu byť kombinované s inými súbormi údajov. Smernica uvádza aj tematické kategórie pre HVD:

- Geopriestorové údaje: PSČ, národné a miestne mapy;
- Pozorovanie Zeme a životné prostredie: spotreba energie a satelitné snímky;
- Meteorologické údaje: údaje in situ z prístrojov a predpovede počasia;
- Štatistika: demografické a ekonomické ukazovatele;
- Spoločnosti a vlastníctvo spoločností: obchodné registre a registračné identifikátory;
- Mobilita: dopravné značky a vnútrozemské vodné cesty.

V praxi existuje veľké množstvo literatúry, ktoré sa podrobnejšie venujú metodológii publikácie údajov (tabuľka nižšie). Tieto publikácie však majú spoločné niekoľko obecných princípov, týkajúcich sa uverejňovaniu otvorených údajov:

1. Licencovanie otvorených údajov,
2. Poskytovanie komplexných metaúdajov,
3. Geolokácia údajov,
4. Spolupráca s komunitou používateľov,

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

5. Meranie využitia údajov.

Tabuľka 4: Dostupné metodiky publikovania otvorených údajov

Názov	Referencia
Best Practices for Publishing Linked Data	HYLAND, B. et al. 2014. [Odkaz]
Government Data Openness and ReUse	ÁLVAREZ, M. 2014. [Odkaz]
Guide for disclosure of public data	Difi. 2013. [Odkaz]
Open Data Handbook	Open Knowledge Foundation. 2012. [Odkaz]
Open Data Handbook	Flemish government. 2014. [Odkaz]
Methodological Guidelines for Publishing Linked Data	Corcho, O. 2011. [Odkaz]

5.3.1 Licencovanie otvorených údajov

Všetky údaje zverejnené ako otvorené údaje by mali byť spojené s licenciou otvorených údajov, aby bolo jasné, za akých podmienok možno údaje opätovne použiť. V európskej smernici o otvorených údajoch z roku 2019 sa uvádza, že „akékoľvek licencie na opakované použitie informácií verejného sektora by mali v každom prípade klásť čo najmenej obmedzení na opakované použitie, napríklad obmedzenia na uvedenie zdroja“.

Verejné licencie umožňujú zhotoviteľovi databázy alebo inému nosičovi práv šíriť ich predmet pod verejnou licenciou bez nutnosti uzatvárať štandardnú licenčnú zmluvu alebo platiť licenčné poplatky. Tieto licencie môžu byť napríklad *Creative Commons* (CC), *Open Data License* (ODbL) alebo *Open Data Commons Attribution License* (ODC-BY). Verejní licencionári musia však vykonávať všetky potrebné práva, vrátane autorského práva a práva na osobné údaje, na aplikovanie verejnej licencie. Ako alternatívu je možné vytvoriť vlastnú verejnú licenciu, avšak táto možnosť nie je odporúčaná, pretože neprináša hlavnú výhodu masovo využívania univerzálne známych licencií. Technický rámec otvorených údajov z roku 2015 odporúča používanie licencie *Creative Commons Attribution 4.0 International* (CC BY 4.0). V rámci licencie CC BY 4.0 môže používateľ:

- Zdieľať – kopírovať a redistribuovať materiál na akomkoľvek médiu alebo formáte,
- Prispôbovať – upravovať, transformovať a používať materiáli na akýkoľvek účel, dokonca aj komerčne.

CC BY 4.0 License Attribution Statement – Podľa licencie CC BY 4.0 musia používatelia uviesť zdroj informácií vo svojom produkte alebo aplikácii.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

CC0 1.0 Universal (CC0 1.0) – Ďalšou možnosťou licencovania pre otvorené údaje je CC0, ktorá je pre používateľov ešte menej obmedzujúca. Na rozdiel od licencií CC, ktoré umožňujú držiteľom autorských práv vybrať si zo širokej škály povolení pri zachovaní ich autorských práv, poskytuje CC0 ešte ďalšiu možnosť – možnosť odmietnuť ochranu autorských práv a databáz a exkluzívne práva automaticky udeľované tvorcom – alternatívu „žiadne práva vyhradené“ k licenciám. Označením CC0 *public domain* sa autor vzdáva autorských práv a umožňuje opätovným používateľom používať obsah aj na komerčné účely.

Sprievodca licenciami

Sprievodca licenciami poskytuje popis dostupných licencií. Poskytuje tiež prehľad o tom, ako používať licencie ako opätovný vydavateľ/distribútor otvorených údajov a ako kombinovať viaceré licencie. Sprievodca je dostupný na stránkach Európskej komisie na Oficiálnom portáli pre údaje o Európe ([Odkaz](#)).

Odporúčania pre vydavateľa otvorených údajov:

- Uistite sa, že vykonávate všetky práva k súborom údajov, ktoré chcete otvoriť. Ak nevykonávate všetky práva, vysporiadajte si ich osobitne (napr. získajte potrebnú licenciu umožňujúcu ďalšie šírenie aj pod verejnou licenciou, uzavrite dodatok k predchádzajúcej zmluve a pod.). Nezabudnite, že okrem databázových práv môžu existovať aj iné práva, ktoré môžu brániť šíreniu údajov, napríklad ochrana osobných údajov.
- Rozhodnite sa, ako budú môcť tretie osoby používať súbor údajov (napr. či bude zahrnuté aj komerčné použitie). Ak umožníte tretím osobám robiť aj zmenu, úpravu alebo modifikáciu databázy, nijakým spôsobom tým nezasiahnu do pôvodnej databázy. Pôvodná databáza ani práva k nej nebudú nijako dotknuté.
- Vyberte si vhodnú verejnú licenciu a ak obsahuje viac verzií a typov, tak aj konkrétny typ. Nezabudnite, že keď chcete skutočne otvorené údaje, zvolená verejná licencia by nemala obsahovať prvok *No Derivatives*, ktorá neumožňuje meniť, upravovať alebo modifikovať súbor údajov. Tiež nie je vhodný prvok *Non-Commercial*, ktorý neumožňuje obchodné využitie databázy. Staršia verzia licencií CC sa nemohla aplikovať na údaje a súbory údajov chránené osobitným právom, preto je nutné používať aktuálnu verziu 4.0.
- Pri každom šírení súboru údajov dôsledne uvádzajte zvolenú verejnú licenciu vrátane jej časovej verzie (napr. Creative Commons BY-SA, 4.0)¹.

¹ Zdroj: Prečo a ako licencovať otvorené údaje. [Odkaz](#).

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

5.3.2 Poskytovanie komplexných metaúdajov

Metaúdaje sú nevyhnutné pre efektívne zverejňovanie otvorených údajov. Poskytujú spôsob, ako objaviť a pochopiť súbory otvorených údajov, pretože poskytujú dodatočný kontext a informácie o obsahu údajov. Metaúdaje napríklad zahŕňajú informácie, ako sú tvorcovia údajov, dátum vytvorenia a geografická oblasť, ktorú údaje pokrývajú. Metaúdaje umožňujú používateľom ľahko nájsť a pochopiť údaje, pretože zhŕňajú alebo opisujú obsah údajov a ďalšie súvisiace informácie. Okrem toho sú metaúdaje kľúčové pre kurátorov otvorených údajov, aby mohli mať súbory údajov aktuálne a usporiadané, čo všetkým uľahčuje vyhľadávanie a používanie údajov.

Metaúdaje poskytujú dodatočné informácie, ktoré pomáhajú používateľom údajov lepšie pochopiť význam údajov, ich štruktúru a objasniť ďalšie problémy, ako sú práva a licenčné podmienky, organizáciu, ktorá údaje vygenerovala, kvalita údajov, metódy prístupu k údajom a plán aktualizácie súborov údajov.

Poskytnutie komplexnejších a presnejších metaúdajov zlepšuje vyhľadateľnosť a použiteľnosť údajov. Analytická správa Európskeho portálu údajov² zdôrazňuje potrebu metaúdajov v troch dimenziách: relevancia (sú to údaje, ktoré potrebujem?); použiteľnosť (môžem to použiť v praxi?) a kvalitu (aké dobré sú údaje a aké ľahké bude ich použitie?). Správa ODI³ tiež zdôrazňuje, že jedným z hlavných problémov zdôrazňovaných v našich zisteniach sú zlé a nekonzistentné metaúdaje.

Odporúčania pre vydavateľa otvorených údajov:

- Poskytnite metaúdaje kompatibilné s DCAT-AP.
- Použite zmysluplné konvencie pomenovania, ktorým porozumie používateľ, ktorý nie je oboznámený so súborom údajov.
- Poskytnite komplexný popis súboru údajov.
- Poskytnite úplný zoznam kľúčových slov vrátane synonym, ktoré môže používateľ vyhľadať.
- Zahrňte informácie o pôvode súboru údajov, t. j. pôvod údajov a akékoľvek zmeny, ktoré boli vykonané.
- Zlepšite metaúdaje na základe hodnotenia kvality metaúdajov.

5.3.3 Geolokácia údajov

Geopriestorové údaje sú jedným zo základných komponentov pri zverejňovaní otvorených údajov. Poskytnutím informácií o lokácii môžu používatelia efektívnejšie

² EDP. 2020. The future of open data portals. [Odkaz](#).

³ ODI. 2019. Discovering the future of the London Datastore. [Odkaz](#).

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

lokalizovať, identifikovať a geopriestorovo analyzovať zdroje údajov. To umožňuje používateľom kombinovať rôzne súbory údajov na základe geografickej oblasti a vytvárať zmysluplné spojenia. Geopriestorové údaje môžu používateľom pomôcť rýchlo preskúmať konkrétnu oblasť, filtrovať oblasti záujmu a lepšie vizualizovať vzťahy. Výsledkom je, že zverejňovanie otvorených údajov, ktoré zahŕňajú geopriestorové informácie, poskytuje koncovému používateľovi väčšiu hodnotu a podporuje široké prijatie iniciatív otvorených údajov.

Geopriestorové údaje boli v európskej smernici o otvorených údajoch zdôraznené ako jedna z tematických kategórií súborov údajov s vysokou hodnotou. Množstvo údajov môže zahŕňať geopriestorové informácie, ako sú súradnice GPS, adresy atď., ktoré možno použiť na prekrytie údajov z viacerých zdrojov a tiež na filtrovanie údajov na miesto, ktoré je pre používateľa relevantné. Je dôležité používať štandardizované geopriestorové údaje, aby sa uľahčila interoperabilita údajov.

- Odporúčania pre vydavateľa otvorených údajov
- Ak údaje zahŕňajú konkrétnu geopriestorovú polohu, uveďte zemepisnú šírku/dĺžku.
- Ak používate premietané súradnice (x/y), zahrňte použitú priestorovú projekciu.

5.3.4 Spolupráca s komunitou používateľov

Spolupráca s komunitou počas procesu zverejňovania otvorených údajov je rozhodujúca pre zvýšenie sociálnej a ekonomickej hodnoty údajov. Pomáha to zabezpečiť, aby zainteresované strany mohli porozumieť údajom, identifikovať, ako by mohli byť prínosom pre ich každodenný život, a merať ich úspešnosť. Komunity môžu pomôcť pri hľadaní využitia údajov, stanovovaní výkonnostných cieľov a sledovaní pokroku, odhaľovaní oblastí na zlepšenie a ponúkaní premyslenej a konštruktívnej kritiky. Prostredníctvom zmysluplnej spolupráce môže publikovanie otvorených údajov pomôcť posilniť prepojenia medzi vládou a občanmi. To zase pomáha budovať dôveru, podporuje občiansku angažovanosť a načrtáva plán na dosiahnutie spravodlivého rozvoja komunity.

Zapojenie používateľov v každej fáze zdieľania údajov je nevyhnutné na podporu opätovného používania údajov a v konečnom dôsledku na stimuláciu hospodárskeho, sociálneho, environmentálneho a politického vplyvu. Najúčinnjším prístupom je prístup založený na problémoch alebo výzvach: práca s používateľmi údajov a dotknutými ľuďmi s cieľom identifikovať problém, ktorý môžu údaje pomôcť vyriešiť, a potom zvýšiť prístup k údajom potrebným na jeho vyriešenie. Zameranie sa na výzvy namiesto jednoduchého vytvárania inventárov údajov s väčšou pravdepodobnosťou prinesie opätovné použitie.

Odporúčania pre vydavateľa otvorených údajov:

- Uveďte kontaktné údaje pre každý zverejnený súbor údajov. Kontaktné údaje by mali zahŕňať všeobecnú e-mailovú adresu spojenú s rolou, nie jednotlivca.

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- Zapojte sa do diskusií o žiadostiach o údaje.
- Zapojte sa do komentárov k súborom údajov.
- Vypočujte si spätnú väzbu týkajúcu sa súborov údajov a ak je to vhodné, konajte na základe tejto spätnej väzby
- Propagujte otvorené údaje na všetkých svojich komunikačných kanáloch vrátane sociálnych médií, bulletinov, udalostí atď.
- Spoluorganizujte udalosti otvorených údajov založené na problémoch/výzvach s používateľmi a ďalšími vlastníkmi.

5.3.5 Meranie využitia údajov

Správne meranie využívania zverejnených otvorených údajov je nevyhnutné na úspešné využitie údajov pre zmysluplné aplikácie. Prostredníctvom merania dokážeme pochopiť mieru využívania a zapojenie používateľov, čo pomáha pri vytváraní stratégie akejkoľvek iniciatívy založenej na údajoch. Okrem toho, pochopenie vzorov používania tiež pomáha pri identifikácii prípadov optimálneho použitia, potenciálnych rizík a oblastí zlepšenia, ktoré sú nevyhnutné pre vládne agentúry aj podniky pri rozhodovaní na základe údajov. Správne meranie a analýza údajov o používaní umožňuje vlastníkom údajov a vydavateľom lepšie pochopiť fázy zapojenia používateľov, čo vedie k zvýšeniu transparentnosti, lepšiemu rozhodovaniu a zlepšeniu verejných služieb.

Existuje mnoho motivácií na meranie otvorených údajov, napríklad pre udržanie kvality údajov a podpory; odôvodnenie investíc; porovnanie výkonnosti medzi krajinami a inštitúciami. Ak je cieľom publikovania otvorených údajov pomôcť dosiahnuť pozitívny environmentálny, ekonomický, politický a sociálny dopad, je dôležité, aby existoval jasný spôsob, ako merať, ako sa údaje používajú a aký majú skutočný vplyv. Meranie využívania údajov pomôže vydavateľom zistiť, či sa dosahujú ciele stanovené v pláne publikovania otvorených údajov ich organizácie. Priebežná kontrola používania údajov môže vydavateľom pomôcť pochopiť problémy, ktorým používatelia čelia, a potenciálne zlepšiť údaje, aby sa dali ľahšie používať.

Definovanie metrick na meranie využitia údajov môže byť zložité, najmä ak to, čo sa meria, môže byť subjektívne, napríklad kvalita údajov. Existuje však množstvo ukazovateľov, ktoré možno automatizovať a dôsledne kontrolovať na meranie prístupu k údajom, ako sú zobrazenia súboru údajov a sťahovanie v katalógu údajov. Ak vydavateľ publikuje údaje prostredníctvom rozhrania API, počet žiadostí o rozhranie API možno merať, aby sme lepšie pochopili používanie.

Odporúčania pre vydavateľa otvorených údajov:

- Priebežne kontrolujte a porovnávajte metriky prístupu k údajom.
- Ak sú údaje publikované prostredníctvom rozhraní API, použite analýzu API na meranie požiadaviek API. Tieto informácie o prístupe k údajom priebežne kontrolujte a porovnávajte.

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- Definujte kvantitatívne metriky špecifické pre danú doménu na meranie používania konkrétnych súborov údajov vašej organizácie.
- Definujte kvalitatívne metódy, aby ste pochopili, ako sa údaje vašej organizácie používajú a aký majú vplyv.

Použitá literatúra:

Government of South Australia, 2014. Open Data Process Guide.
EDP. 2018. Open Data Goldbook for Data Managers and Data Holders.
DGA. 2020. Open Data Process Quick Guide.
UK Biobank. 2023. Data Access Guide.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

6 Osvedčené postupy pre zdieľanie údajov verejného sektora

6.1 Implementácia celoorganizačnej stratégie

Koordinovaná stratégia je pre pokrok v organizáciách verejného sektora nevyhnutná. Centralizovaná stratégia spája viaceré organizácie a ich rôzne snahy. To znamená, že rozhodovacie orgány majú celkový prehľad a môžu sledovať, usmerňovať a kontrolovať pokrok. Tento koordinovaný prístup umožňuje štátu rýchlo identifikovať, posúdiť a reagovať na príležitosti aj riziká spojené s otvorenými údajmi a informáciami verejného sektora. Táto stratégia pôsobí ako nevyhnutný prvý krok pre efektívne rozhodovanie a zlepšovanie verejných služieb, pretože umožňuje informovanú analýzu príčin a dôsledkov v danej oblasti politiky. Centrálné plánovanie a hodnotenie sú nevyhnutné aj pre tvorcov politík a úradníkov, pretože si vyžadujú dohľad nad organizáciami vykonávajúcimi stratégiu, čo im umožňuje preskúmať pokrok v porovnaní s dohodnutými cieľmi a podať správu zainteresovaným stranám. Centrálné plánovanie a hodnotenie je tiež nevyhnutné pre lepšie strategické riadenie verejného sektora. Centralizované stratégie pomáhajú rôznym organizáciám pochopiť, čo je potrebné, zodpovedajúco plánovať a merať pokrok.

Existencia kontrolovateľného plánu zverejňovania údajov umožňuje vedúcim predstaviteľom verejného sektora identifikovať nové príležitosti a zmierniť prípadné riziká spojené s otvorenými údajmi a informáciami verejného sektora. V konečnom dôsledku je aktívna koordinovaná stratégia nevyhnutná na to, aby sa verejný sektor posunul vpred a vytvoril základ pre rozhodovacie orgány.

Odporúčanie:

- Zverte zodpovednosť za vypracovanie plánu jednému zamestnancovi alebo oddeleniu.
- Zvolajte stretnutie alebo sériu stretnutí medzi zainteresovanými stranami – výrobcami údajov, používateľmi údajov atď.
- Vypracujte plán prostredníctvom iteratívnych procesov predtým, než budete hľadať podporu na vysokej úrovni.

6.1.1 Príklad najlepšej praxe: Belgicko

Príklad z Belgicka poukazuje na to, ako flámska vláda podporuje flámske úrady a miestne orgány, aby otvárali údaje a vytvárali aplikácie prostredníctvom financovania alebo spolufinancovania hodnotných inovatívnych iniciatív otvorených údajov. Bolo doručených 24 návrhov, z ktorých bolo vybraných desať. Prípadové štúdie a prezentácia popisujú prístup a očakávané výsledky tejto iniciatívy.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

Digitálna agenda pre Európu je iniciatívou stratégie Európa 2020 a zameriava sa na úlohu IKT pri plnení ambícií Európy do roku 2020. V rámci tohto programu sa čoraz väčší počet európskych vlád zaviazal sprístupniť údaje z verejný sektor, vzhľadom na jeho potenciál prinášať spoločnosti ekonomickú a sociálnu hodnotu. Jednou z týchto iniciatív je Open Data Program zriadený vo Flámsku, regióne v Belgicku, ktorého cieľom je zabezpečiť občanom a podnikom prístup k vládnym údajom a informáciám s minimálnymi obmedzeniami. Len Belgicku to môže priniesť odhadovanú ekonomickú hodnotu 875 – 900 miliónov eur. S cieľom využiť tento potenciál sa flámska vláda dohodla na stratégii otvorených údajov, ktorá je štandardne založená na princípe otvorených údajov.

Vláda rozhodla o prístupe zhora nadol pri implementácii stratégie otvorených údajov. Bola vypracovaná koncepcná poznámka, ktorú podpísali všetci ministri. Po koncepcnej poznámke nasledoval akčný plán na roky 2012 a 2013, v ktorom sa podrobne uvádzali potrebné kroky na napredovanie programu. Efektívna implementácia tohto akčného plánu však do značnej miery závisela od informovanosti zainteresovaných strán na všetkých úrovniach národnej, regionálnej a miestnej samosprávy. V roku 2013 bola okrem mnohých iných iniciatív zverejnená výzva na predkladanie návrhov, v ktorej sa žiadali správy a miestne orgány vo Flámsku, aby predložili návrhy zamerané na začlenenie otvorených údajov do ich programu vytvorením a implementáciou otvorených údajov v ich organizácii a/alebo miestnych konštituentoch. Z dvadsiatich štyroch návrhov bolo vybraných desať projektov, ktoré budú efektívne podporené rozpočtom takmer pol milióna eur.

6.2 Podpora crowdsourcingu

Crowd sourcing sa vo verejnom sektore stáva čoraz populárnejším ako prostriedok na zvýšenie kvality a množstva dostupných strojovo čitateľných údajov. Zapojením komunity do procesu čistenia a formátovania údajov môžu vlastníci údajov využívať celý rad vonkajších vplyvov, ktoré uľahčujú pokrok oveľa efektívnejším a efektívnejším spôsobom. Podobne možno pomocou inovatívnych techník, vrátane gamifikácia, zvýšiť vlastníctvo a angažovanosť v projekte, čím sa podporí vyššia úroveň účasti a nadšenia. Platformy ako GitHub umožňujú používateľom rýchlo a jednoducho ponúkať opravy existujúcich údajov, aj keď tieto musia byť nakoniec schválené zdrojom údajov. Tento osvedčený postup zahŕňa efektívny spôsob, akým môžu verejné inštitúcie jednoduchšie zdieľať verejné údaje bez ohľadu na ich zdroje a dostupnosť rozpočtu.

Odporúčanie:

- Najprv identifikujte presnú potrebu a potom vyhľadajte skupiny schopné podporiť riešenie tejto potreby prostredníctvom crowdsourcingu.
- Myslite na crowdsourcing ako na ďalší nástroj na vytváranie/zlepšovanie súborov údajov a zamyslite sa nad fázami vášho projektu zberu údajov a nad tým, kam by sa crowdsourcing najlepšie hodil.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- Zapojte zainteresované strany, ktoré by mohli mať úžitok z bezplatného zdroja určitých súborov údajov, a nechajte ich, aby poskytli finančné prostriedky na udržanie úsilia o crowdsourcing.
- Úlohy musia byť malé, aby ich dobrovoľníci zvládli s obmedzeným časom.
- Ak je to možné, použite gamifikačný prístup, to znamená, že hraním hry používateľia vykonávajú užitočnú úlohu.

6.2.1 Príklad najlepšej praxe: Česká republika

Nadácia OSF v spolupráci s Digitálnou a informačnou agentúrou ČR vyhlasuje každoročne súťaž *Společně otvíráme data*, ktorá oceňuje najlepšie aplikácie postavené na otvorených údajoch. Zmyslom súťaže je podporiť vznik nových služieb pre verejnosť a zviditeľniť spoločenský a ekonomický potenciál otvorených údajov a open source technológií. Ocenenia získavajú aplikácie, ktoré prispievajú k väčšiemu zapojeniu verejnosti do verejného diania, pomáhajú ľuďom v každodennom živote, vnášajú transparentnosť do fungovania verejnej správy, kombinujú využitie otvorených údajov a open source alebo upozorňujú na významný spoločenský problém.

Súťažiteľ môžu záujemcovia z radov verejnosti, neziskových organizácií, verejných inštitúcií, študentov aj firiem. Každý víťaz dostane finančnú odmenu na ďalší rozvoj aplikácie. Súťaží sa v niekoľkých kategóriách, v každej z nich vyhráva 1 aplikácia. Myšlienka súťaže je, aby do roku 2021 v ČR fungovala svojbytná komunita presadzujúca verejnú správu, ktorá:

- je otvorená, publikuje otvorené údaje a umožňuje, aby s jej digitálnou stopou mohol pracovať ktokoľvek
- je zodpovedná, poskytuje občanom informácie o svojom fungovaní a rozhodovacích procesoch, a tí ju vďaka digitalizácii môžu kontrolovať odkiaľkoľvek
- je interaktívna, a preto podporuje ako občiansku participáciu, tak čo najjednoduchšiu komunikáciu v rámci vzťahu štát a občan (nástroje ako e-voľby, možnosť vybaviť si veci z domova, získať prístup k informáciám o vlastnej osobe)

Súťaž pomáha zosilniť tlak zvonku na systémovú zmenu, zvýšiť počet občanov, občianskych iniciatív, neziskových organizácií a biznisu aktívne využívajúcich digitálnu stopu štátu na watchdogové účely a vývoj služieb pre verejnosť.

6.3 Vytvorenie rámca kvality údajov

Pokiaľ ide o otvorené údaje, tieto úvahy je potrebné brať veľmi vážne, aby sa zabezpečila najvyššia možná kvalita pôvodne poskytnutých súborov údajov a presnosť rozhodnutí z nich odvodených. Preto by organizácie prostredníctvom postupov riadenia kvality údajov mali zabezpečiť platnosť otvorených údajov, ktoré poskytujú, a zabezpečiť, aby boli údaje aktuálne, dôveryhodné a najvyššej kvality. Konkrétne by riadenie kvality údajov

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

malo zahŕňať procesy, ako je správa údajov, stanovenie noriem údajov a presadzovanie prísnych metrík kvality údajov. Prostredníctvom štruktúr správy údajov môžu spoločnosti zabezpečiť presnosť údajov prostredníctvom pravidelne uplatňovaných procesov validácie, pokiaľ je to možné, založených na znalostiach špecifických pre danú oblasť odvodených odborníkmi na danú problematiku. Prijatie dátových štandardov znižuje bremeno spracovania mnohých rôznych dátových formátov, keďže sa zvyšuje kvalita údajov a ich spracovateľnosť, a zároveň uľahčuje zabezpečenie presnosti údajov v rámci organizačných procesov. Nakoniec je dôležité merať kvalitu údajov a monitorovať ich v priebehu času, pričom sa zabezpečí, že údaje budú mať správne kvalitatívne charakteristiky, ktoré budú slúžiť na presné rozhodnutia. Preto by subjekty okrem tradičných metrík presnosti, použiteľnosti a zrozumiteľnosti mali pri hodnotení kvality otvorených údajov zväžiť aj také opatrenia, ako je otvorenosť, včasnosť, prvenstvo, konzistentnosť, dostupnosť, úplnosť, zhoda, dôveryhodnosť, spracovateľnosť a relevantnosť.

Odporúčanie:

Na zvýšenie kvality údajov sú potrebné efektívne riadiace štruktúry a na zabezpečenie dodržiavania konvencií, noriem a štandardov možno použiť rôzne systémy a nástroje. Okrem toho by sa mali implementovať základné slovníky, CSV on the Web Recommendations, DCAT-AP, W3C Data Quality Vocabulary (DQV), RDF Data Cube validator a UnifiedViews, pretože sú dôležité pre procesy publikovania údajov a integrácie.

Z hľadiska organizácie je úloha *Chief Information Officer* (CIO) nevyhnutná pri vytváraní rámca správy údajov, ktorý integruje správu architektúry údajov, správu metaúdajov a správu hlavných údajov (MDM). Tento rámec má spájať alebo ovplyvňovať celú organizáciu, pretože údaje sú kľúčovým aktívom, ktoré prispieva k celkovej prevádzke a finančnej výkonnosti. Implementácia tohto rámca by mala dodržiavať normy ISO 8000; to znamená zabezpečiť, aby používanie slovnej zásoby bolo konzistentné, sémantické kódovanie bolo presné a úplné a aby sa zachoval pôvod a presnosť údajov. Tento rámec správy údajov môže organizácii výrazne prospieť z hľadiska efektívnosti, presnosti a úspory nákladov.

1. Definujte metriky kvality údajov: Identifikujte, ktoré merania by sa mali vykonať na posúdenie kvality súborov otvorených údajov. Napríklad presnosť údajov, úplnosť, relevantnosť, konzistencia atď.
2. Vytvorte proces hodnotenia kvality: Vypracujte proces hodnotenia kvality údajov pomocou metrík určených v kroku 1. Uistite sa, že proces zdokumentujete, aby ho mohli použiť aj iní.
3. Vyškolte personál: Vyškolte zamestnancov v procese hodnotenia kvality údajov a metrikách stanovených v kroku 1, aby mohli konzistentne a presne hodnotiť kvalitu súborov otvorených údajov.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

4. Automatizácia procesov: Hľadajte príležitosti na automatizáciu čo najväčšieho množstva procesu hodnotenia kvality údajov. Napríklad použite softvér alebo automatizované systémy na generovanie vizualizácií alebo detekciu odľahlých hodnôt.
5. Zverejňovanie výsledkov: Zverejňujte výsledky hodnotenia kvality údajov, aby verejnosť mohla pochopiť kvalitu údajov.
6. Monitorujte a revidujte: Monitorujte proces hodnotenia kvality údajov a výsledky v priebehu času a hľadajte príležitosti na zlepšenie procesu a výsledkov.

6.3.1 Príklad najlepšej praxe: Spojené kráľovstvo

Vláda potrebuje pri porozumení, dokumentovaní a zlepšovaní kvality svojich údajov zaujať štruktúrovanejší prístup. Vyžaduje si to stratégiu založenú na dôkazoch, ktorá musí zahŕňať proaktívnu a cieľenú prácu v oblasti kvality údajov, ktorá vedie k neustálemu zlepšovaniu. S cieľom zabezpečiť, aby štátni zamestnanci pochopili dôležitosť kvality údajov a cítili sa pohodlne pri aktívnej identifikácii a riešení problémov, vytvorili v spojení s kráľovstvom rámec kvality údajov (*Data Quality Framework*), ktorý im ponúka súbor zásad, ktoré pomôžu efektívne spravovať údaje a zosúladiť prístup s cieľom dosiahnuť vysokokvalitné údaje potrebné na dosiahnutie lepších výsledkov pre verejnosť.

Prvá časť rámca poskytuje organizáciám a jednotlivcom štruktúru, aby mohli uvažovať v rámci:

- Princípy kvality údajov na podporu organizácií pri vytváraní kultúry kvality údajov
- Sprievodca životným cyklom údajov, ktorý pomôže organizáciám identifikovať a zmierniť potenciálne problémy s kvalitou údajov vo všetkých fázach
- Dimenzie kvality údajov, na základe ktorých je možné vykonávať pravidelné hodnotenia kvality údajov

Druhá časť rámca poskytuje návod na praktické nástroje a techniky, ktoré možno použiť na hodnotenie, komunikáciu a zlepšenie kvality údajov:

- Akčné plány kvality údajov, ktoré sa používajú na identifikáciu praktických krokov na posúdenie kvality údajov a na cieľené zlepšenia
- Analýza základných príčin, aby sa zabezpečilo, že práca s kvalitou údajov rieši problémy pri zdroji
- Usmernenie pre metaúdaje na podporu lepšieho využívania metaúdajov na komunikáciu a interpretáciu kvality
- Komunikácia usmernení o kvalite vrátane navrhovaných prístupov na jasné sprostredkovanie kvality používateľom

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- Úvod do modelov zrelosti údajov pre tých, ktorí chcú zaujať holistický prístup k hodnoteniu a zlepšovaniu kvality údajov

Kvalita údajov je nevyhnutná pre efektívne rozhodovanie založené na dôkazoch. Nízka alebo neznáma kvalita údajov môže mať škodlivé následky, ako sú slabšie dôkazy, nedôvera a zlé výsledky, a v konečnom dôsledku môžu viesť k tomu, že organizácie budú menej efektívne. Aby sa zabezpečilo, že údaje budú vhodné na daný účel, je potrebné vytvoriť „kultúru údajov“, čo znamená, že organizácie by mali riešiť problémy pri zdroji, zaviazat' sa k nepretržitému monitorovaniu a podávaniu správ, zamerať sa na zlepšenia, ktoré prinesú najväčšiu hodnotu, a povzbudiť jednotlivcov, aby boli pri získavaní údajov proaktívni. manažment kvality. Tento rámec využíva najlepšie medzinárodné a priemyselné postupy a ponúka princípy, postupy a nástroje zamerané na dosiahnutie kvality údajov vhodnej na daný účel. Na pomoc pri vytváraní silnej kultúry kvality údajov rámec poskytuje usmernenia vo forme princípov, ktoré vysvetľujú najlepšie postupy a postoje, ktoré najviac pomôžu zabezpečiť, aby údaje vyhovovali účelu. Tie princípy sú:

- Zamerajte sa na kvalitu údajov
- Poznajte svojich používateľov a ich potreby
- Posúďte kvalitu počas celého životného cyklu údajov
- Komunikujte kvalitu údajov jasne a efektívne
- Očakávajte zmeny ovplyvňujúce kvalitu údajov

6.4 Publikovanie prehľadu spravovaných údajov

Aby sa zabezpečilo, že komunita používateľov vie o súboroch údajov, ktoré by organizácia mohla potenciálne sprístupniť ako otvorené údaje, organizácia by mala zvážiť zverejnenie prehľadu súborov údajov, ktoré spravuje. Používateľskej komunite by to poskytlo úplnejší obraz o súboroch údajov dostupných na potenciálne zverejnenie, aby sa verejnému sektoru mohla poskytnúť užitočná spätná väzba. To by malo tiež zlepšiť efektívnosť zhromažďovania spätnej väzby od komunity, pretože bude mať prístup k potrebným znalostiam, aby mohla poskytnúť zmysluplné informácie o budúcich vydaniach. Navyše to verejnému sektoru umožní presne stanoviť priority na základe tejto spätnej väzby. Tento osvedčený postup zverejňovania prehľadu súborov údajov spravovaných organizáciou by mal umožniť komplexnejšiu a zmysluplnejšiu výmenu medzi komunitou a verejným sektorom, podporovať otvorenosť a zvyšovať efektívnosť iniciatív otvorených údajov.

Odporúčanie:

Prax kategorizácie otvorenosti údajov poskytuje užívateľsky prívetivý a transparentný indexovací systém na prezentáciu rôznych úrovní otvárania údajov. Mať zavedený systém, ktorý jasne definuje úroveň dostupnosti každého súboru údajov, je kľúčom k zvýšeniu dôvery a angažovanosti. Technické požiadavky na túto prax sú minimálne;

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

všetko, čo je potrebné, je jednoduchá zverejnená tabuľka a kontaktné miesto, na ktoré možno poslať spätnú väzbu. Okrem toho je možné prehľad tejto praxe jednoducho zahrnúť do existujúceho katalógu otvorených údajov pomocou funkcie „prístupových práv“ v rámci DCAT-AP. Vďaka kategorizácii otvorenosti budú používatelia schopní rozlišovať medzi legálne obmedzenými a neverejnými súbormi údajov a efektívnejšie využívať a opätovne používať otvorené údaje.

- Vytvorte štandardizovaný systém zberu a správy údajov: Prvým krokom k implementácii stratégie publikovania údajov je vytvorenie jednotného systému na zber a správu údajov. To zahŕňa vytvorenie procesu na rozhodovanie o tom, ktoré údaje je dôležité zachytiť, definovanie metrik a štandardov na zber údajových bodov a vývoj protokolov na zabezpečenie presnosti údajov.
- Stanovte správu a vlastníctvo údajov: Je tiež dôležité určiť tím alebo jednotlivca, ktorý bude dohliadať na proces správy a vlastníctva údajov. Táto osoba by mala byť zodpovedná za nastavenie postupov na umožnenie prístupu k údajom, zaistenie bezpečnosti a kvality údajov a vytvorenie politík pre prístup k údajom a ich správu.
- Vytvorte stratégiu zverejňovania údajov: Po zavedení systému zberu a správy údajov by sa mala vytvoriť stratégia zverejňovania údajov. Môže to zahŕňať vytvorenie online portálu, kde môžu používatelia pristupovať k údajom, alebo publikovanie súborov údajov v otvorených dátových formátoch. Okrem toho je nevyhnutné vytvoriť protokoly a usmernenia na sprístupnenie údajov.
- Monitorovanie a audit údajov: Po zavedení procesu zverejňovania údajov je dôležité zabezpečiť, aby boli údaje presné a aktuálne. To môže zahŕňať nastavenie procesu auditu, ktorý pravidelne kontroluje presnosť údajov. Okrem toho je dôležité pravidelne prehodnocovať postupy zabezpečenia údajov, aby ste sa uistili, že sú aktuálne a že nedochádza k neoprávnenému prístupu k údajom.
- Vyhodnoťte a aktualizujte svoju stratégiu publikovania údajov: Postupom času je dôležité pravidelne vyhodnocovať a aktualizovať stratégiu publikovania údajov. To zahŕňa monitorovanie množstva údajov, ktoré sa zdieľajú, nastavenie smerníc pre prístup k údajom a vývoj nových spôsobov, ako sprístupniť údaje a zvýšiť ich hodnotu.

6.4.1 Príklad najlepšej praxe: Fínsko

Mestá nepretržite zhromažďujú a produkujú veľké množstvo údajov, ktoré sa využívajú v procesoch plánovania, výroby a rozhodovania. Využíva sa aj pri hodnotení činnosti samosprávy. Tradične sa tieto údaje používajú iba na interné účely mesta, ale od roku 2011 tieto údaje otvárajú metropolitné mestá Helsinky.

Služba *Helsinki Region Infoshare* (HRI) www.hri.fi/en/ je služba otvorených údajov, ktorá poskytuje verejné súbory údajov týkajúce sa Helsinskej metropolitnej oblasti (Helsinki, Espoo, Vantaa a Kauniainen) ako otvorené údaje, ktoré môže ktokoľvek voľne používať.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

HRI financujú mestá Helsinki, Espoo, Vantaa a Kauniainen. HRI tiež podporil fínsky inovačný fond Sitra a fínske ministerstvo financií. HRI bol vytvorený a vyvinutý v úzkej spolupráci s Forum Virium Helsinki. Službu koordinuje City of Helsinki Urban Facts.

Koncept služby Helsinki Region Infoshare sa začal na začiatku roka 2010 a plánovanie sa začalo v júni 2010. Prvá verzia webovej služby bola spustená online v marci 2011. Po vydaní projekt pokročil podľa testovania a učenia sa používateľov. Do roku 2012 webová služba ponúkala širokú škálu údajov o regionálnej oblasti Helsínk a jej lokalitách. V priebehu roka 2013 bolo cieľom sprístupniť viac údajov a zaviesť webovú službu HRI ako súčasť bežnej prevádzky obcí. Tento cieľ bol dosiahnutý začiatkom roka 2014; služba HRI sa stala súčasťou bežnej mestskej prevádzky.

Otvorené údaje môžu využívať občania, podniky, univerzity, akadémie, výskumné pracoviská a samospráva. HRI poskytuje široké spektrum otvorených údajov týkajúcich sa napr. životné podmienky, ekonomika a blahobyť, zamestnanosť a doprava. Veľká časť údajov je založená na GIS. Otvorené údaje sú užitočnou surovinou pri výskumných a vývojových aktivitách, rozhodovaní, vizualizácii, dátovej žurnalistike a pri vývoji aplikácií. HRI sa zameriava na rozvoj štyroch základných oblastí: produkcia údajov, prístup k údajom, zdieľanie údajov a využitie údajov. Hlavnou operatívnou činnosťou je podpora producentov informácií pri otváraní ich údajov a zvýšenie ich využitia využitím viackanálovej komunikácie (napr. sociálne médiá, blogy a pod.). Kritériá HRI na zverejňovanie údajov sú:

- Súbor údajov musí vytvoriť a udržiavať verejná agentúra (napr. oddelenie).
- Súbor údajov sa musí týkať regiónu Helsinki alebo niektorého jeho okresu.
- Obsah súboru údajov musí byť verejný a zverejnenie údajov nesmie porušovať žiadne zákony, nariadenia alebo dohody.
- Súbor údajov by mal byť dostupný pre kohokoľvek, aby ho mohol voľne používať.
- Súbor údajov by mal byť dostupný na internete pre kohokoľvek.
- Súbor údajov by mal byť uvoľnený s otvorenou licenciou (povolením), ktorá tiež umožňuje komerčné opätovné použitie údajov. HRI dôrazne odporúča používať medzinárodnú licenciu Creative Commons BY 4.0.
- Údaje musia byť v strojovo čitateľnom a štruktúrovanom formáte, napríklad ako súbor CSV alebo dostupné prostredníctvom otvoreného aplikačného programovacieho rozhrania (REST API), ktoré vracia údaje vo formáte XML alebo JSON. Súbor údajov je možné zverejniť aj napr. ako súbor programu Excel alebo tabuľku Dokumentov Google.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

6.5 Monitorovanie a meranie opätovného použitia otvorených údajov

Dnes sú ľahko dostupné sofistikovanejšie metódy, ktoré nám umožňujú posúdiť potenciálne výhody a nevýhody publikovania údajov a informácií. Celosystémový integrovaný prístup nám umožní robiť tie najlepšie rozhodnutia, pokiaľ ide o uvoľnenie potenciálne citlivých alebo cenných informácií. Pri rozhodovaní založenom na dôkazoch môžu organizácie nájsť rovnováhu medzi stratou príjmu a celkovým prínosom pre organizáciu. To môže zahŕňať vyššiu efektivitu, zmeny v spôsobe poskytovania služieb alebo zvýšenú transparentnosť. Model využívajúci pokročilé metriky možno použiť na získanie podpory od zainteresovaných strán, ako aj na pokrok v plánoch na bezpečné zverejňovanie a speňaženie údajov. V konečnom dôsledku to zníži ľudské úsilie a náklady, zvýši príležitosti na analytické pochopenie a poskytne dostupnejšie údaje pre ciele verejného a súkromného sektora. Takéto stratégie, ak sú dobre navrhnuté a implementované, zjednodušujú správu údajov a informácií a podporujú prostredie inovácií pre organizácie všetkých veľkostí.

Odporúčanie:

Aby bolo možné efektívne posúdiť hodnotu vytvorenú činnosťami verejného sektora, federálne orgány by mali zaviesť stratégiu otvorených údajov. To nielenže uľahčí meranie hodnoty na makroúrovni, ale uľahčí aj posúdenie prínosov a nákladov spojených s miestnou implementáciou smernice o informáciách verejného sektora. Pomocou riadenia podľa cieľov, analýzy údajov a prediktívneho modelovania na meranie účinku zverejňovania údajov a informácií na federálnej úrovni budú orgány môcť lepšie monitorovať pokrok v užitočnosti merania hodnoty. Je tiež dôležité poukázať na to, že tento proces by sa nemal brať na ľahkú váhu, pretože niektorým jednotlivým oddeleniam môžu vzniknúť značné náklady len s výhodami na vyššej úrovni, čo sťažuje dlhodobú užitočnosť meradla hodnoty.

- Rozvinúť jasné pochopenie typov údajov, ktoré sú určené na opätovné použitie, a potenciálnych výhod opätovného použitia.
- Určte zodpovedného jednotlivca alebo tím, ktorý bude priebežne monitorovať metriky opätovného použitia.
- Stanovte základné línie na meranie výkonnostných kritérií.
- Sledujte používanie otvorených množín údajov v priebehu času.
- Vyhodnoťte počet aplikácií a služieb, ktoré sú vytvorené pomocou otvorených množín údajov.
- Zmerajte celkové zvýšenie spotreby údajov a dostupnosti.
- Analyzujte spätnú väzbu od používateľov, aby ste pochopili hodnotu a vplyv otvorených množín údajov.
- Posúďte potenciálne speňaženie opätovného použitia otvorených údajov.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- Monitorujte vnímanie iniciatív otvorených údajov zo strany externých médií.
- Poskytnite používateľom jednoduchou navigačnú platformu na navrhovanie aplikácií alebo služieb, ktoré by mohli byť vyvinuté s otvorenými množinami údajov.

6.5.1 Príklad najlepšej praxe: Estónsko

V Estónsku je veľký záujem o pochopenie úrovne opätovného použitia otvorených údajov. Väčšina verejných orgánov požaduje a opakovane používa údaje od iných verejných orgánov pomocou bezpečnej národnej infraštruktúry výmeny údajov X-tee – technológia vyvinutá spoločne Estónskom, Fínskom a Islandom prostredníctvom MTÜ Nordic Institute for Interoperability Solutions. X-tee umožňuje sledovať presný počet žiadostí o údaje prichádzajúce od každej organizácie.

Služby e-Estonia sú postavené na dvoch kľúčových stavebných blokoch – systéme e-ID a platforme na výmenu údajov X-tee. Občania majú digitálny podpis pomocou e-ID kariet, čo sa používa na autentifikáciu identity a tvorí základ poskytovania digitálnych služieb v oblasti zdravotníctva, spravodlivosti atď.

Estónski občania, e-rezidenti a organizácie môžu využívať alebo poskytovať viac ako 2860 digitálnych služieb prostredníctvom vrstvy X-tee na výmenu údajov podporovanej eGovernmentom a cez X-tee sa mesačne odošle viac ako 150 miliónov žiadostí. Väčšina transakcií sa uskutočňuje medzi organizáciami verejného sektora. Štatistiky sa zhromažďujú a uverejňujú ako otvorené údaje s oneskorením 10 dní od skutočného času vykonania transakcie.

Použitá literatúra:

BABA, Y. 2015. Crowdsourcing Data Understanding: A Case Study using Open Government Data.

CRUSOE, J. 2020. Towards a Framework for Open Data Publishers: A Comparison Study between Sweden and Belgium.

ECIS. 2020. How to Implement an Open Data Strategy? Analyzing Organizational Change Processes to Enable Value Creation by Revealing Data.

Helsinki Region Infoshare. https://hri.fi/en_gb/

EDP. Helsinki Region Infoshare. <https://data.europa.eu/sites/default/files/use-cases/finland - helsinki regional infoshare.pdf>

Open Data Watch. 2022. Better measurement and monitoring of data for development.

OSF. 2018. Data pro lepší Česko.

RIA. 2023. Data exchange layer X-tee

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

UK. 2020. The Government Data Quality Framework.

UK. 2020b. Data Quality and Methodology: Private registered providers social housing stock and rents in England statistics.

WALRAVENS, N., BALLON, P., Van COMPERNOLLE, M., & BORGHYS, K. 2021. Data Ownership and Open Data: The Potential for Data-Driven Policy Making. In The Data Shake.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

7 Zahraničný prístup k archivácii údajov verejného sektora – eARD Projekt (SE)

V papierovom prostredí sa „aktívne“ dokumenty uchovávajú v kancelárii a „neaktívne“ dokumenty sa presúvajú do „blízkeho archívu“, samostatnej miestnosti v blízkosti, umožňujúcej jednoduchý prístup. Keď záznamy už nie sú pravidelne potrebné, sú opäť presunuté buď do centrálného archívu (často umiestneného v suteréne), alebo na archívny orgán. To znamená, že v praxi je rozdiel medzi zaobchádzaním s „aktívnymi“, „poloaktívnymi“ a „archívnymi“ záznamami. Digitálne záznamy nezaberajú kancelársky priestor, ale nie je jasné, ako dlhodobo archivovať uzavreté digitálne dokumenty a ako riešenie sa predkladá „stredný archív“.

Jedna z definícií „stredného archívu“ je výsledkom projektu e-Archive and e-Diarium (eARD), kolaboratívneho projektu spojeného s rozvojom elektronickej verejnej správy, ktorý ovplyvnil archívne postupy vo Švédsku. Je to preto, že jeho výsledky – spoločné špecifikácie pre prenos informácií medzi akýmikoľvek a všetkými systémami informačných technológií, zlepšili schopnosť prenosu digitálnych záznamov z obchodného systému do elektronického archívu.

Iniciatíva pre projekt pochádza od Delegácie pre eGovernment (spojenie 16 verejných organizácií vrátane VINNOVA), ktorá označila elektronicкую archiváciu a elektronicкую registráciu záznamov za jeden z niekoľkých strategických faktorov úspechu pre vytvorenie efektívnej elektronickej verejnej správy. VINNOVA je švédska vládna agentúra pre inovačné systémy Švédska. Spolu s Asociáciou miestnych samospráv a regiónov (SALAR) a Asociáciou miestnych samospráv okresu Štokholm (KSL) veľmi aktívne vytvárajú regulačné dokumenty o otvorených údajoch.

Švédsky národný archív (Riksarkivet) v mene delegácie pre elektronicкую verejnú správu vykonal predbežnú štúdiu pre projekt e-Archive a e-Diarium na jeseň 2010 v spolupráci s niekoľkými verejnými inštitúciami a KSL. Výsledkom štúdie bolo, že:

- vláda poverila Národný archív vedením projektu zameraného na zdieľané služby pre služby e-Archive a e-Diarium pre vládne inštitúcie,
- vláda poverila Národný archív celkovou zodpovednosťou za rozvoj a správu spoločných špecifikácií (FGS), ktoré budú tvoriť základ pre spoločné služby e-Archive a e-Diarium,
- vláda iniciovala audit s cieľom objasniť oficiálnu úlohu Národného archívu a iných archívnych inštitúcií v súvislosti so službami e-Archive a e-Diarium.

Predbežná štúdia bola prijatá a v októbri 2012 bol projekt eARD označený za jeden z prioritných projektov delegácie elektronickej verejnej správy. Ciele projektu boli stanovené takto:

- vyvinúť spoločné špecifikácie (FGS) pre vládne inštitúcie na prenos digitálnych záznamov medzi systémami správy záznamov a do elektronického archívu;

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

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

- pripraviť návrhy FGS pre e-Archive a e-Diarium v spolupráci s národnými, regionálnymi a miestnymi vládnymi inštitúciami.
- testovanie a zabezpečenie kvality spoločných špecifikácií, najmä prostredníctvom pilotných projektov v existujúcich prevádzkových podnikových systémoch a elektronických archívoch; a
- vytvorenie organizácie, ktorá bude riadiť a ďalej rozvíjať FGS.

Vízia, ktorá poháňala projekt, bola, že by malo byť pomocou spoločných špecifikácií uľahčené získať, opätovne použiť a preniesť informácie uchovávané verejnými orgánmi do archívu bez ohľadu na to, kde a ako sú informácie uložené.

Spoločné špecifikácie (*förvaltningsgemensamma specifikationer*, FGS), sú vyvíjané tak, aby uľahčili vyhľadávanie a získavanie informácií pre vládne inštitúcie a občanov. Spoločná špecifikácia je štruktúrovaný popis funkčných a technických požiadaviek, ktoré spĺňajú potreby celej štátnej správy alebo jej časti. Špecifikácia poskytuje návod pri tvorbe predpisov, špecifikácií pre obstarávanie systému a pri písaní zmlúv.

Špecifikácie majú za cieľ zabezpečiť požadovanú funkčnosť a interoperabilitu v rámci organizácií verejnej správy, ako aj pri jednaní s občanmi a podnikmi. Sú potrebné ako základ pre zavádzanie riešení založených na službách a prevádzkovej podpore podnikania a sú predpokladom pre vytvorenie celoštátneho zdieľania informácií a dlhodobého poskytovania informácií. Správa spoločných špecifikácií navyše zjednodušuje vývoj, obstarávanie a nasadzovanie jednotných riešení – môžu znížiť náklady a vytvoriť podmienky, ktoré uľahčia vyhľadávanie a opätovné použitie informácií. Pre ich vytvorenie, musia byť identifikované požiadavky na povinné metaúdaje, aby boli splnené základné požiadavky všetkých dotknutých organizácií.

eARD projekt priniesol niekoľko FGS, ktoré vo všeobecnosti pozostávali z technických špecifikácií štruktúry a obsahu metaúdajov alebo prenosového formátu. Výsledkom projektu eARD bol tiež zoznam konceptov s celkovo 109 termínmi, medzi nimi aj „stredný archív“. Šesť z týchto pojmov údajne potrebovalo ďalšie skúmanie zo strany Národného archívu: „archívny systém“, „e-archív“, „e-archivácia“, „typ informácií“, „systém na uchovávanie“ a „systém na dlhú dobu“. Hlavný prínos eARD je vytvorenie špecifikácie, ktorá podnietila ďalšie inovácie technologických metód na riešenie archivácie a interoperability otvorených údajov. Významným projektom, ktorý priamo čerpal z výstupov projektu eARD bol projekt E-ARK.

E-ARK bol nadnárodný výskumný Big Data projekt, ktorý zlepšil metódy a technológie digitálnej archivácie s cieľom dosiahnuť konzistentnosť v celoeurópskom meradle.

E-ARK, ktorý riešil celý rad problémov spojených s technológiami, systémami a praxou nezávislého uchovávania záznamov, prospel rozvoju medzinárodne dostupných archívov prostredníctvom i. poskytovania technických špecifikácií a nástrojov; ii. rozvoja integrovanej infraštruktúry archivácie; iii. demonštrácie lepšej dostupnosti; iv. prístupu a používania a dôsledná analýzy metaúdajov.

Error! Reference

Koncept publikovania a archivácie údajov verejnej správy vo forme otvorených údajov

V období od 1. februára 2014 do 31. januára 2017 bol E-ARK spolufinancovaný Európskou komisiou v rámci programu na podporu politiky IKT (PSP) v rámci rámcového programu pre konkurencieschopnosť a inovácie. Projekt E-ARK poskytol:

- Špecifikácie pre:
 - spoločnú špecifikáciu informačných balíkov, príjem, archiváciu, prístup
 - relačnú databázu, elektronickú evidenciu a archiváciu geoúdajov
- Vzorové nástroje na:
 - kontrolu, či špecifikácie fungujú
 - demonštrovanie, že špecifikácie fungujú
- Pilotné projekty:
 - implementácia špecifikácie a nástrojov do prevádzkového prostredia
- Centrum znalostí:
 - E-ARK Maturity Model a Nástroj na hodnotenie
 - Všeobecný model E-ARK pre vizualizáciu

Cieľom projektu od roku 2016 do roku 2017 bolo navrhnuť výsledky pre E-ARK, aby sa stal stavebným blokom CEF (*Connecting Europe Facility*). Stavebný blok je balík technických špecifikácií a/alebo služieb a/alebo vzorového softvéru, ktorý možno opätovne použiť v projektoch akejkolvek oblasti politiky. Konkrétne:

- Technické špecifikácie stavebného bloku sú čo najotvorenejšie a orientované na trh.
- Služby stavebného bloku musia byť dobre definované (t. j. zdokumentované, so zmluvami SLA, školením, helpdeskom atď.)
- Softvér stavebného bloku musí mať dostatočnú zrelosť (t. j. úspešne otestovaný pre cezhraničné transakcie.)

Zdroje:

CCSDS. 2012. *Reference Model for an Open Archival Information System (OAIS)*. public.ccsds.org/publications/archive/650x0m2.pdf

E-ARK Project. 2014. *Deliverable D3.1 E-ARK Report on Available Best Practices*. <http://e-ark-project.com/resources/projectdeliverables/6-d31-e-ark-report-on-available-best-practices>

Riksarkivet. 2014. *The e-archive and e-Diarium project, eARD*.

Riksarkivet. 2013. *Delprojekt 1 (DP1) inom e-arkiv och e-diarium (Begreppsdefinitioner) Ordlista*

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