

Dodatok č. 1

k Metodickému pokynu k výnosu Ministerstva financií SR zo 4. marca 2014 č. 55/2014 - 173 Z. z. o štandardoch ISVS v znení opatrenia Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 11/2018 z 12. marca 2018

čl. I

Dodatok č. 1 k Metodickému pokynu k výnosu Ministerstva financií SR č. 55/2014 -173 Z. z. o štandardoch ISVS dopĺňa ustanovenia § 51 – § 53 výnosu z dôvodu novely výnosu o štandardoch a v znení opatrenia Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu (ďalej len „úrad“) č. 11/2018 z 12. marca 2018.

K § 51

Kvalita datasetu poskytovaného povinnou osobou

Klasifikácia datasetu podľa kvality je základom pre jeho efektívne strojové spracovanie, čo je jedna z najdôležitejších požiadaviek pre otvorené údaje. Nízka kvalita datasetu si vyžaduje vynaložiť značné úsilie na použitie rozličných datasetov, kým v prípade vysokej kvality datasetov je s nimi možné pracovať ako s jedným celkom.

Na stanovenie kvality datasetu sa používa modifikovaná medzinárodne akceptovaná klasifikácia “Five Star OpenData”, tzv. 5-hviezdičkové otvorené údaje, ktorej autorom je Tim Berners-Lee (viď. <http://5stardata.info/en/>). Vo Výnose je pre úplnosť doplnená úroveň 0, ktorá sa týka prístupňovania údajov výlučne v listinnej forme.

Úroveň kvality 3 je reprezentovaná štandardnými otvorenými formátmi ako XML, ODS, resp. CSV, ktoré je možné vytvoriť jednoduchým exportom dát do požadovaného formátu. Úroveň 4 a 5 je založená na otvorenom formáte RDF, ktorý predpisuje používanie URI (referencovateľných identifikátorov) na všetky entity ako dátové prvky, číselníky, referenčné údaje a podobne. Množina povinne používaných URI identifikátorov, ktoré sa pre tento dôvod registrujú v MetaIS, predstavuje úroveň 5, kým ostatné, menej dôležité údaje, ktoré nepatria medzi referenčné údaje, predstavujú úroveň 4. Výhodou úrovne 5 je, že používaním povinných URI identifikátorov vo všetkých datasetoch je možné s nimi pracovať ako s jedným celkom, čo je kľúčová požiadavka pre celkovú interoperabilitu verejných údajov. V dokumente Národnej koncepcie informatizácie verejnej správy (NKIVS), Strategická priorita Otvorené údaje, ktorý schválila Rada vlády SR pre digitalizáciu verejnej správy a jednotný digitálny trh a je dostupný na URL <http://www.informatizacia.sk/strategie-priority-erf/24190c>, je úroveň kvality 5 stanovená pre kľúčové, základné, resp. referenčné údaje verejnej správy. Realizovateľnosť tvorby datasetov úrovne 5 súvisí s vývojom nového, alebo inovovaného publikačného vybavenia ISVS v spolupráci s Centrálnou dátovou kanceláriou.

Z pohľadu formy reprezentácie údajov je základným rozdielom medzi jednotlivými úrovňami skutočnosť, že úroveň 1 je neštruktúrovaný obsah (napr. text), úroveň 2-3 je tabuľkovo orientovaná štruktúra dát a úroveň 4-5 je grafová forma údajov (reálne vzťahy, t.j. entita je spojená so všetkým, s čím súvisí). Kým prechod z úrovne 2 na úroveň 3 je často relatívne jednoduchý (napr. konverzia tabuľkových dát z formátu Excel do formátu CSV, či ODS), prechod z úrovne 3 na úroveň 4 resp. 5 predstavuje konceptuálny skok a vyžaduje si plánovanie a softvérové riešenie. Návrh 5-hviezdičkových údajov vždy prebieha v koordinácii s centrálnym orgánom (proces registrácie URI). Z tohto dôvodu sa ich tvorba hodí viac na automatické exporty.

Príklady úrovni interoperability otvorených údajov je možné nájsť v Metainformačnom systéme (MetaIS) na URL <https://metais.finance.gov.sk/publicspace> resp. na <https://wiki.finance.gov.sk/pages/viewpage.action?pageId=23986522>

K § 52 Otvorené údaje

Ustanovenie § 52 definuje podmienky na označenie údajov za otvorené, pričom otvorený údaj musí byť štruktúrovaný tak, aby bol jednoducho automatizovane spracovateľný a vzhľadom na zabezpečenie interoperability nesmie byť viazaný na nejaké uzatvorené proprietárne riešenie. Takúto podmienku spĺňajú napr. formáty XML, CSV, ODT, ale napr. formát XLS už túto podmienku nespĺňa.

Je potrebné, aby povinné osoby dohliadali na zverejňovanie otvorených dát a sprístupnenie verejne dostupných API v projektoch financovaných prostredníctvom OPII a iných verejných zdrojov.¹ Preto, ak sa prostredníctvom OPII alebo verejných zdrojov financuje inovovanie existujúceho alebo vytvorenie nového informačného systému, je potrebné dbať na to, aby obsahoval verejne dostupné aplikačné programové rozhranie a realizoval automatizovaný export zverejniteľných dát ako otvorených údajov.

K písm. a):

Požiadavka vyjadrená v tomto bode stanovuje minimálnu úroveň kvality datasetov s otvorenými údajmi. To znamená, že ak dataset nespĺňa aspoň úroveň kvality 3, nie je možné ho označiť ako dataset otvorených údajov.

Tým samozrejme nie je zabránené inému sprístupňovaniu údajov (t.j. nie vo forme otvorených údajov), ktoré túto úroveň nespĺňajú. Do katalógu v rámci portálu otvorených dát je možné zaradiť aj datasety, ktoré túto požiadavku na minimálnu úroveň kvality nespĺňa, pretože dostupnosť údajov v nižšej kvalite je lepšia než nedostupnosť týchto údajov. V tomto smere sa zároveň predpokladá, že bude následne vyvinuté úsilie na zvýšenie kvality zverejnených údajov aspoň na úroveň kvality 3 (aby mohli byť napĺňané ciele v súlade so štandardami).

¹ Úloha č. 13 uznesenia vlády SR č. 104/2017.

K písm. b):

Keďže otvorené údaje sú verejne prístupné, pri ich prevzatí používateľom nedochádza ku podpísaniu dohody o spôsobe použitia údajov, ako býva bežné pri iných spôsoboch sprístupnenia údajov. Avšak pre používateľa je dôležité poznať presné podmienky, ktorými sa pri následnom využívaní údajov musí riadiť. Vymedzenie týchto podmienok je potrebné nielen pre zvýšenie právnej istoty (pritom napr. komerčné použitie údajov spravidla nie je možné bez vysokého stupňa istoty, že daná činnosť je povolená), ale často je aj priamo legislatívne vyžadované.

Túto situáciu špecificky upravuje Autorský zákon č. 185/2015 Z.z. (ďalej len „Autorský zákon“), ktorý v ustanovení § 130 ods.1 definuje pojem databáza: *„Databázou je súbor navzájom nezávislých diel, údajov alebo iných navzájom nezávislých materiálov systematicky alebo metodicky usporiadaných a jednotlivo prístupných elektronickými alebo inými prostriedkami bez ohľadu na formu jeho vyjadrenia.“* Z uvedeného je zrejmé, že prakticky každý dataset v zmysle definície ustanovenia § 2 písm. r) výnosu o štandardoch je aj databázou v zmysle Autorského zákona.

Autorský zákon upravuje v štvrtej časti samostatne tzv. "sui generis" práva k databáze. Detailné riešenie problematiky aplikácie autorského práva v prostredí databáz presahuje rámec tohto metodického pokynu, je však možné konštatovať, že v niektorých prípadoch je podľa ustanovenia § 65 ods.1 Autorského zákona vyžadované uzavretie licenčnej zmluvy, ktorá *„obsahuje najmä spôsob použitia diela, rozsah licencie, čas, na ktorý autor licenciu udeľuje, alebo spôsob jeho určenia a odmenu alebo spôsob jej určenia, ak sa autor s nadobúdateľom nedohodol na bezodplatnom poskytnutí licencie“*.

Pri sprístupnení otvorených údajov teda nestačí iba samotné deklarovanie dodržania štandardov. Špecificky ustanovenie § 52 ods. 1 písm. b) bod 1 výnosu o štandardoch špecificky vyžaduje *„explicitné vysporiadanie právnych aspektov používania údajov“*, čo budeme pre zjednodušenie v ďalšom texte nazývať pojmom "licencia". V znení štandardu nie je použité slovo "licencia" z dôvodu, že tento termín je zaužívaným pojmom v oblasti autorského práva, a aj keď v niektorých prípadoch informácia, resp. dataset nie sú vôbec Autorským zákonom chránené (napr. podľa § 5 Autorského zákona), jednoznačné stanovenie podmienok je v týchto prípadoch nevyhnutné. Pre určitý dataset môžu zároveň existovať ďalšie právne požiadavky, ktoré je potrebné dodržať, napr. ak ide o údaje zverejňované na základe povinnosti stanovenej zákonom.

Štandardy dávajú poskytovateľovi údajov právo vybrať si licenciu podľa vlastnej vôle (samozrejme pri dodržaní všetkých aplikovateľných požiadaviek). Ak by však pre každý dataset existovala iná licencia, používateľ údajov by to mal v praxi veľmi náročné. Už len detailná právna analýza licencie býva zložitá a drahá a pri kombinácii údajov z viacerých datasetov by takto vznikala významný „licenčný zmätok“.

K písm. c):

V súlade s usmernením Európskej komisie² odporúčame použitie určitých štandardných licencií. Creative Commons je celosvetovo najširšie používanou licenčnou schémou a na Slovensku je aplikovateľná ako špecifická forma verejnej licencie podľa § 76 Autorského zákona. Na základe vyššie uvedených podmienok sú s ohľadom na charakter otvorených údajov pre datasety otvorených údajov vhodné dve konkrétne licencie:

- CC0 (Creative Commons Zero) (<https://creativecommons.org/publicdomain/zero/1.0/>) - autor sa v nej do maximálnej možnej miery vzdáva autorských práv - zjednodušene povedané, používateľ má teda právo použiť údaje poskytovateľa akýmkoľvek spôsobom, samozrejme v rámci zákonných limitov (napr. nemôže si privlastniť autorstvo, pravidlá ochrany osobných údajov zostávajú zachované, a pod.),
- CC-BY (Creative Commons Attribution) (<https://creativecommons.org/licenses/by/4.0/>) - používateľ podľa nej môže údaje použiť akýmkoľvek spôsobom (porovnateľne ako pri CC0) avšak za splnenia podmienky, že pri použití údajov vždy vhodným spôsobom uvedie ich autora (t.j. pri otvorených údajoch poskytovateľa datasetu).

Vyššie uvedené podmienky musia byť splnené kumulatívne to znamená, že na to, aby bolo možné nejaký údaj označiť ako otvorený údaj, musia byť súčasne splnené všetky potrebné náležitosti. Ak je napr. nejaký údaj dostupný bezodplatne, v požadovanej kvalite, v otvorenom formáte (aby splnil požiadavky na úroveň kvality 3) a pod akceptovanou licenciou, ale nie je dostupný prostredníctvom anonymného vzdialeného automatizovaného prístupu, nie je možné označiť ho ako otvorený údaj. Preto, ak je dataset zverejnený formou linku na externý systém, ktorý vyžaduje pre prístup k dátam prihlásenie, nejedná sa o otvorený údaj.

K písm. d):

V súlade s aktuálnym znením Výnosu o štandardoch, ktoré nadobudlo účinnosť 15. marca 2018 je štandardom pre označenie údajov ako otvorených údajov, aj potreba poskytnutia metaúdajov datasetu otvoreným spôsobom.

² Oznámenie Komisie č. 2014/C 240/01: Usmernenia o odporúčaných štandardných licenciách, súboroch údajov a o spoplatnení opakovaného použitia dokumentov (<http://eur-lex.europa.eu/legalcontent/SK/TXT/HTML/?uri=CELEX:52014XC0724%2801%29&from=EN>), kapitola 2.2. Otvorené licencie, z ktorej vyberáme: *"Prostredníctvom otvorených štandardných licencií, ako sú napríklad najnovšie licencie Creative Commons (6) (verzia 4.0), by sa mohlo povoliť opakované použitie informácií verejného sektora bez toho, aby bolo potrebné vyvíjať a aktualizovať licencie na mieru na vnútroštátnej alebo nižšej než celoštátnej úrovni. Z nich je obzvlášť zaujímavý nástroj CC0 public domain dedication (určenie na verejný prístup) (7). Ako právny nástroj, ktorým sa umožňuje vzdanie sa autorských práv a práv k databáze, pokiaľ ide o informácie verejného sektora, zaisťuje plnú flexibilitu pre opakovaných používateľov a zmierňuje komplikácie spojené so zaobchádzaním s početnými licenciami, prípadne s protichodnými ustanoveniami."*

K § 53

Poskytovanie otvorených údajov

Základnou požiadavkou pre každý dataset otvorených údajov je jeho zaevidovanie v centrálnom katalógu na portáli data.gov.sk (podľa písm. f). Cieľom je umožniť používateľom čo najjednoduchším spôsobom zistiť, ktoré datasety existujú a aké sú ich základné parametre.

Podľa písm. a) sa vyžaduje, aby bola pre každý poskytovaný dataset otvorených údajov jednoznačne uvedená jeho úroveň kvality. Označenie sa vykoná v mieste sprístupnenia datasetu alebo v katalógu otvorených údajov, v ktorom je dataset zaradený, obdobne ako je to vyžadované pri uvedení metaúdajov datasetu podľa písm. h). Ak dataset obsahuje viaceré dátové zdroje, úroveň kvality datasetu zodpovedá najvyššej úrovni kvality dátového zdroja, v ktorej je dostupný každý údaj datasetu.

Požiadavka podľa písm. b) taxatívne určuje prípustné formáty súborov obsahujúcich otvorené údaje. Jednotlivé súbory s údajmi sú samostatnými dátovými zdrojmi datasetu. Podľa znenia výnosu o štandardoch sú prípustné nasledovné formáty: CSV, JSON, XML a model RDF (serializovaný do RDF/XML alebo JSON-LD, ktorý je preferovaný). Požiadavky pre formát CSV sú ďalej upravené v § 24 písm. h) a súčasne v prílohe č. 5 výnosu o štandardoch.

Formát schémy datasetu vyžadovaný podľa písm. c) závisí na použitom formáte súboru s údajmi. Zodpovedajúce priradenie formátu súboru a schémy sú RDF – OWL a XML - XSD. Pre formáty CSV a JSON je namiesto schémy vyžadovaný textový opis štruktúry. V prípade JSON prebiehala k dátumu vydania metodického pokynu v gescii medzinárodnej štandardizačnej organizácie Internet Engineering Task Force (IETF) štandardizácia jazyka schémy (viď. <http://jsonschema.org/documentation.html>).

Požiadavka podľa písm. e) sumarizuje možné spôsoby prístupu k údajom v zmysle iných častí Výnosu o štandardoch. Aktuálne sú štandardmi špecifikované protokoly FTP/FTPS, HTTP/HTTPS a aplikačné rozhrania SOAP a REST. Po poskytnutí dotknutých rozhraní je možné údaje sprístupňovať aj inými protokolmi a rozhraniami.

Identifikáciu doby platnosti údajov podľa písm. g) bodu 1 je možné vykonať na úrovni celého datasetu alebo samostatne pre jednotlivé položky datasetu. Rozlišovanie platnosti údajov je závislé na kontexte datasetu - v niektorých prípadoch môže byť platnosť údajov definovaná časovým intervalom (napr. platnosť názvu ulice), v iných môže byť viazaná k určitému okamihu (napr. meranie teploty ovzdušia) alebo môže byť časovo neobmedzená (napr. výsledky hlasovania vo voľbách - v tomto prípade má zase naopak zmysel rozlišovať, či už údaje začali platiť alebo sú iba predčasné).

Možnosť zistiť, ktoré údaje boli po aktualizácii zmenené podľa požiadavky písm. g) bodu 2 je kľúčová pre efektívny prístup k údajom, najmä pre obmedzenie duplicitného prenosu a

spracovania údajov. Význam tejto požiadavky rastie s veľkosťou datasetu a frekvenciou jeho zmien. Identifikáciu zmenených údajov je možné realizovať rôznymi spôsobmi, napr. vytvorením samostatného dátového zdroja obsahujúceho zmeny pri každej aktualizácii, uvedením časovej značky poslednej zmeny pri každom údaji datasetu alebo uchovávaním úplnej histórie zmien údajov v datasete.

Podrobnosti týkajúce sa metaúdajov datasetu (doplnujúcich údajov k základným údajom) podľa písm. h) sú uvedené v rámci popisov tohto dokumentu k prílohe č. 9 (kapitola 10.10) výnosu o štandardoch.

Podľa aktuálneho znenia Výnosu o štandardoch, ktoré nadobudlo účinnosť 15. marca 2018, bola v ustanovení § 52 ods. 1 písm. d) doplnená požiadavka na “poskytovanie metaúdajov datasetu otvoreným spôsobom použitia podľa písmen b) a c)” na to, aby bolo možné označiť údaje ako otvorené údaje. Ku každému datasetu otvorených údajov teda musia byť k dispozícii metaúdaje poskytované otvoreným spôsobom.

Vytváranie metadát a popisovanie datasetov:

Portál otvorených údajov data.gov.sk má pri registrácii nového datasetu vyžadovať metadáta v takej podobe, aby ich vyplnením a automatickou validáciou pri odoslaní mala povinná osoba už automaticky splnenú povinnosť súvisiacu s poskytnutím potrebných metadát. Toto zároveň znižuje administratívnu náročnosť celého procesu. Pri existujúcich datasetoch je potrebné skontrolovať, či sú všetky v minulosti vyplnené metaúdaje v dostatočnom rozsahu a kvalite a v prípade potreby ich aktualizovať (častá chyba v minulosti bola chýbajúca informácia o licencií, resp. nevhodný výber licencie -- práve z tohto dôvodu Výnos o štandardoch ISVS odporúča použitie Creative Commons Zero / CC0, prípadne Creative Commons Attribution / CC-BY). Definovanie rozsahu otvorených údajov a ich povinnosti, licencií a použitia kontrolovaných slovníkov (číselníkov) je v súčasnosti realizované prostredníctvom publikačného profilu DCAT-AP-SK, ktorý je národnou špecializáciou aplikačného profilu DCAT-AP (Data Catalog Application Profile), odporúčenou programom ISA2 Európskej komisie.

<https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe>

DCAT-AP-SK je definovaný v prílohe č. 9 (od 1. júla 2018) obsahuje množstvo tried, napr. Katalóg, Dataset a Zdroj (ich špecifikácie sú dostupné v MetaIS). Tieto slúžia ako špecifikácie toho, ktoré metaúdaje je potrebné zbierať (napr. povinná trieda Dataset obsahuje povinné vlastnosti ako sú názov, popis, vydavateľ, ale aj nepovinnú ako napr. Kontaktné miesto a pod.).

Publikovanie datasetov:

Národná agentúra pre sieťové a elektronické služby (NASES) zverejnila priamo na stránke <https://data.gov.sk/> videonávody ako postupovať pri zverejnení a aktualizácii datasetu. ÚPVII odporúča postupovať podľa návodov NASES.

Metodika Open Data Ready:

Ide o súbor kritérií podstatných pre možnosť publikácie otvorených dát automaticky informačnými systémami povinnej osoby. Metodika obsahuje podrobný popis týchto kritérií, postupy ich dosahovania a postup vyhodnotenia pre plnenie týchto kritérií. Kľúčovým kritériom metodiky je vysporiadanie si autorských práv k údajom, aby ich bolo možné zverejniť ako otvorené údaje, ideálne pod licenciou Creative Commons Zero (CC0). Ďalšie informácie o Open Data Ready: http://www.informatizacia.sk/index/open_file.php?ext_dok=26035 (dokument Strategická priorita Otvorené údaje).

Metodika pre vytvorenie dátového inventára a dátového katalógu:

V dokumente “Stratégia a akčný plán sprístupnenia a používania otvorených údajov verejnej správy” (schváleným uznesením vlády SR č. 346/2017, vypracovaným na základe bodu B.11 uznesenia vlády 104/2017 a dostupným na https://www.slovensko.sk/img/CMS4/strategia_otvorene_udaje.pdf) sa píše: “Zistenie, aké údaje inštitúcia spravuje je nevyhnutné. Na úspešnú realizáciu procesu uvoľňovania otvorených údajov je nutné získanie dátového inventára a jeho pravidelného doplňovanie. Výsledkom vytvorenia dátového inventáru je tzv. "katalóg" inštitúcie. Katalóg je tiež zverejňovaný ako dataset v otvorenom formáte. Dátový inventár má formu metaúdajov k jednotlivým datasetom”.

Každá inštitúcia má mať teda prehľad o údajoch, ktoré spravuje a ktoré spravujú jej podriadené organizácie (toto zahŕňa všetky údaje, vrátane tých, ktoré z legislatívnych dôvodov nesmú byť zverejnené ako otvorené údaje). Datasetsy sa opisujú slovníkom DCAT-AP-SK (účinnosť od júna 2019) v prílohe č. 9 Výnosu. Vyplnia sa všetky údaje, ktoré je možné vyplniť a ostatné ostanú prázdne (napr. pri datasetoch, ktoré sa nezverejňujú, sa nevyplní URL, kde je dataset dostupný). Používanie DCAT-AP-SK zjednodušuje ďalšie spracovanie a zavádza spoločný jazyk, ktorým sa datasetsy popisujú.

Pri vykonávaní dátového auditu sa nad rámec DCAT-AP-SK (zoznam polí DCAT-AP-SK je v prílohe č. 9, zoznam korešpondujúcich názvov stĺpcov je dostupný v texte štandardu DCAT-AP-SK v MetaIS) pri každom datasete vyplnia nasledovné polia:

- “právne informácie”: Právne informácie o datasete, t.j. ktoré právne predpisy súvisia s datasetom, aké licencie sa viažu k údajom, či existujú špecifické právne dôvody proti zverejneniu (osobné údaje a pod.).
- “organizačné informácie”: Organizačné informácie, t.j. kto je zodpovedný za dataset, účel vzniku datasetu, spôsob ako je dataset udržiavaný, aby bol stále aktuálny, informácie o frekvencii zmien.
- “posúdenie hodnoty”: Posúdenie hodnoty, t.j. či má dataset vysokú alebo nízku prioritu, či existujú dôkazy o verejnom záujme o dataset, či existujú dôkazy o záujme iných inštitúcií verejnej správy, či existujú dôkazy o záujme medzinárodných orgánov alebo medzinárodných organizácií, či existujú dôkazy o obchodnej hodnote datasetu.
- “technické informácie”: Technické informácie, t.j. v akej forme sú údaje uchovávané, aké má príslušné programové vybavenie, funkcionality, či umožňuje publikovanie otvorených

údajov, napríklad, či má funkciu exportu do XLS a následný export z XLS do CSV, či existujú špecifické prekážky.

- “štrukturálne informácie”: Štrukturálne informácie, t.j. úplný popis všetkých dátových prvkov, vecný opis typu údaj, napríklad v tvare "meno" alebo "adresa", referenciu na externú definíciu pre údaj typu identifikátor a ak existujú, tak ohraničenia pre hodnotu typu údaj.
- “zverejnenie”: Kedy organizácia plánuje dataset zverejniť, resp. ak zverejnenie nie je možné, na základe akého zákona zverejnenie nie je možné. Ak je dataset zverejnený, uvedie sa URL na data.gov.sk.
- “organizácia”: Identifikácie organizácie zodpovednej za dataset (dôležité v prípade konsolidovania údajov; keď zodpovedná organizácia uvádza v dátovom inventári aj údaje podriadenej organizácie, uvedie jej názov, v opačnom prípade uvedie svoj názov).

S výnimkou polí “zverejnenie” a “organizácia” tieto popisy nad rámec profilu DCAT-AP-SK vychádzajú z dokumentu NASES “Stratégia a akčný plán sprístupnenia a používania otvorených údajov verejnej správy”. Inštitúcia si môže pre internú potrebu doplniť aj vlastné atribúty (napr. textové poznámky).

Dátový inventár obsahuje informácie o všetkých údajoch (nielen otvorených údajoch), ktoré inštitúcia a jej podriadené organizácie spravujú a slúži primárne pre potrebu tejto inštitúcie. Mal by byť podľa potreby priebežne aktualizovaný. Z dátového inventára sa realizujú rôzne typy reportov či exportov, napríklad:

- V kontexte otvorených údajov: exporty ako vstup pre dátový audit a taktiež exporty pre data.gov.sk (dátový katalóg),
- V kontexte informatizácie: exporty a reporty pre potreby dátovej či architektonickej kancelárie, typicky pri plánovaní či realizácii rozvoja ISVS,
- V kontexte GDPR: zdokumentovanie a doladenie potrebné pre súlad s požiadavkami na ochranu osobných údajov,
- a iné.

Keďže má dátový inventár viaceré využitia (nad rámec čisto otvorených údajov), je vhodné k nemu pristupovať dôsledne, komplexne, ale aj efektívne (typicky nerobit’ audit iba pre otvorené údaje, a potom ďalší audit pre potreby GDPR, a pod.). Ďalej, keďže dátový inventár obsahuje aj neverejné údaje, pri exportoch je nevyhnutná obozretnosť, aby boli v súlade s platnou legislatívou (napr. aby sa na katalógu otvorených údajov neobjavili utajované údaje).

Inštitúcia zverejní a primeraným spôsobom aktualizuje svoj dátový katalóg (“nultý dataset” so zoznamom datasetov) na data.gov.sk tak, že z dátového inventára najprv odstráni údaje, ktoré sú utajované, odstráni svoje interné atribúty (napr. textové poznámky) a ponechá atribúty z DCAT-AP-SK a za nimi atribúty uvedené v zozname vyššie (“právne informácie” až “organizácia”), v uvedenom poradí. V prípade 3 hviezdikového datasetu sa použijú názvy stĺpcov z poľa “Názov” prílohy č. 9 výnosu, a to vrátane diakritiky a zachovania veľkosti písmen (“Názov”, “Popis”,

“Vydavateľ”, “Dátum vydania”,...). Jednotný názov stĺpcov naprieč verejnou správou je dôležitý pre neskoršiu konsolidáciu dát a automatické spracovanie.

Výsledné poradie stĺpcov pre každý dataset je teda nasledovné:

- Názov
- Popis
- Vydavateľ
- Dátum vydania
- Správca
- URL
- Verzia
- Autor
- Téma datasetu
- URI
- Typ datasetu
- Kľúčové slová
- Distribúcia datasetu
- Kontakt
- Prispievateľ
- Periodicita aktualizácie
- Geografická príslušnosť datasetu
- Časová platnosť datasetu
- Práva
- Správnosť dát
- Zdrojový dataset
- Predchádzajúca verzia
- Poznámky k verzii
- Dokumentácia
- Nasledujúca verzia
- Jazyk
- Iný identifikátor
- Informácia o pôvode datasetu
- Súvisiaci zdroj
- Vzorka údajov
- Dátum modifikácie
- *právne informácie*
- *organizačné informácie*
- *posúdenie hodnoty*
- *technické informácie*
- *štrukturálne informácie*
- *zverejnenie*
- *organizácia*

NASES na data.gov.sk vytvorí a udržiava "meta katalóg", čiže dataset, ktorý obsahuje zoznam odkazov na katalógy jednotlivých inštitúcií, spolu s kontaktnými údajmi na zodpovednú osobu v danej inštitúcii.

Doplňujúce informácie k prílohe č. 9 Výnosu

Nasleduje znenie informácie k prílohe č. 9, účinné do 30. júna 2018. Nové znenie k DCAT-AP-SK bude zverejnené dodatočne.

Príloha č. 9 Výnosu popisuje súbor metaúdajov, ktoré je možné doplniť k poskytovanému datasetu otvorených údajov, pričom optimálne je poskytovať úplný súbor metaúdajov podľa tejto prílohy.

Termín "metaúdaje" sa v súčasnosti používa s rôznym významom v závislosti od použitia údajov. Z hľadiska Výnosu je podstatná definícia uvedená v § 2 písm. w): „metaúdajmi [sa rozumejú] štruktúrované údaje obsahujúce informácie o primárnych údajoch, pričom primárne údaje spravidla reprezentujú určitý hmotný objekt alebo nehmotný objekt; metaúdaje sú určené najmä na vyhľadávanie, katalogizáciu a využívanie primárnych údajov“.

V kontexte otvorených údajov tvoria primárne údaje obsah datasetu a metaúdaje sú používané na popis vlastností na úrovni celého datasetu (t.j. metaúdaje sa neviažu na jednotlivé údaje ktoré sú obsahom datasetu).

Ak dataset obsahuje viaceré dátové zdroje, metaúdaje podľa bodov 3, 5, 7 a 9 sú spravidla stanovené pre každý dátový zdroj samostatne.

K bodu 2 (“jednoznačný identifikátor poskytovateľa datasetu”)

Poskytovateľom datasetu je správca informačného systému verejnej správy (viď. § 2, ods.1, písm. b) Zákona o IS VS), z ktorého sú údaje sprístupňované. Ak technickú realizáciu publikácie údajov vykonáva iný prevádzkovateľ informačného systému ako jeho správca, odporúča sa uviesť v metaúdajoch túto informáciu (napr. kontaktný bod pre používateľov).

K bodu 3 (“referencia na miesto sprístupnenia datasetu a jeho schémy údajov”)

Referencia na miesto sprístupnenia datasetu a schémy údajov je vyjadrovaná vo forme Uniform Resource Locator (URL).

K bodu 4 (“informácia o poskytovaní datasetu otvoreným spôsobom použitia podľa § 52 písm. b) alebo o spôsobe vysporiadania právnych aspektov prístupu k údajom v datasete a ich používania”)

Je potrebné uviesť úplnú informáciu umožňujúcu presné vyhodnotenie možností využitia datasetu, čo spravidla znamená licenciu v zmysle Autorského zákona. Detailnejší popis je uvedený v komentári k § 52 písm. b).

K bodu 5 (“dátum prvého zverejnenia datasetu”)

Uvádza sa dátum prvého zverejnenia konkrétneho datasetu a nie dátum sprístupnenia údajov, ktoré sú obsahom datasetu. Určité údaje môžu byť napr. obsiahnuté vo viacerých datasetoch, ktoré však majú odlišné dátumy prvého zverejnenia.

K bodu 6 (“informácia o aktuálnosti datasetu podľa § 53 písm. g”)

Požiadavky podľa jednotlivých bodov § 53 písm. g) je potrebné vyjadriť samostatnými metaúdajmi.

K bodu 8 (“katalóg, v ktorom je dataset evidovaný”)

Podľa požiadavky § 53 písm. f) musí byť dataset povinne evidovaný v katalógu Portálu otvorených dát (data.gov.sk). Častou praxou je aj vytvorenie vlastného katalógu otvorených údajov určitého správcu IS VS. Vo všetkých katalógoch sa odporúča uvádzať obe takéto miesta zverejnenia, a to vo forme URL smerujúce do týchto "primárnych" katalógov.

Čl. II

Tento dodatok nadobúda účinnosť dňom zverejnenia na webovom sídle úradu.

Mgr. Martina Slabejová, PhD., v. r.
generálna riaditeľka
sekcia riadenia informatizácie