

OTEVŘENÁ DATA – NEŽÁDEJTE O NĚ, VEZMĚTE SI JE!

Ing. JIŘÍ ŠUNKA

V loňském časopisu Národní pojištění číslo 4/2016 jsem se zabýval historií a principy otevřených dat a způsobem jejich implementace v podmínkách ČSSZ. Během minulého roku se nám podařilo dopracovat celý proces publikace otevřených dat do téměř finální podoby jak po stránce provozní, tak i po stránce realizační.

Připomeňme si alespoň základní definici otevřených dat a jejich historii v ČSSZ:

Otevřená data jsou úplná, snadno dostupná, strojově čitelná, používající standardy s volně dostupnou specifikací, zpřístupněná za jasně definovaných podmínek s minimem omezení a dostupná uživatelům při vynaložení minima nákladů.

HISTORIE V PODMÍNKÁCH ČSSZ

Základním krokem bylo rozhodnutí vedení ČSSZ o zveřejňování otevřených dat na vlastních webových stránkách. Byla přijata zásadní rozhodnutí, že data budou publikována v maximální otevřenosti, bude vytvořeno vlastní řešení pro publikaci dat a kromě samotných dat bude ČSSZ zveřejňovat i jejich vizualizace ve formě interaktivních grafů pro jejich snazší interpretaci. ČSSZ svůj portál otevřených dat zveřejnila spolu s prvními datovými sadami 2. listopadu 2015 a najdete jej na adrese <https://data.cssz.cz>. Další publikace nových datových sad se uskutečnily v dubnu, listopadu a prosinci téhož roku. Současně od poloviny roku 2016 probíhají pravidelné aktualizace zveřejněných datových sad podle publikačního plánu a stanovené periodicity.

V současné době je publikováno celkem 63 datových sad z různých oblastí – důchodového, sociálního a nemocenského pojištění, lékařské posudkové služby, OSVČ, správního řízení, statistických ročenek, personalistiky a dalších.

ČSSZ nezůstala pouze u publikace samotných dat, ale připravila také interaktivní vizualizace vybraných datových sad v podobě grafů (sloupcových, spojnicových, koláčových) nebo mapových vizualizací na úrovni států, krajů a okresů. U většiny vizualizací lze interaktivně měnit parametry výběru dat pro zvolený graf (například rok, druh důchodu, pohlaví...) a získané výstupy uložit formou tabulky ve formátu CSV pro další zpracování nebo jako obrázek grafu ve formátu PNG. Celkem je publikováno 75 vizualizací.

PORTÁL OTEVŘENÁ DATA ČSSZ

Portál Otevřená data ČSSZ je přístupný komukoliv na adrese <https://data.cssz.cz>. Na úvodní stránce se nacházejí základní informace o možnostech portálu, podmínkách užití a jeho registraci v Národním katalogu otevřených dat. V pravé části je umístěn panel novinek, který obsahuje průběžné časově seřazené informace o provedených publikacích nových datových sad, o aktualizacích a úpravách portálu. V dolní části





jsou odkazy na Nejčastější dotazy, Prohlášení o přístupnosti, Technickou pomoc a webové stránky provozovatele portálu ČSSZ.

Nejdůležitější záložkou v menu portálu je Katalog otevřených dat. Po jeho výběru se zobrazí formulář, který v horní části obsahuje seznam klíčových slov, podle nichž se dají v seznamu vyhledávat datové sady. Zcela vpravo nahoře je umístěno pole pro zadání fulltextového vyhledávání, které umožní vybrat záznamy, obsahující text uvedený v názvu nebo popisu datové sady. Hlavní část (dvousloupcová tabulka) obsahuje v prvním sloupci abecedně setříděný seznam datových sad a ve druhém seznam vizualizací příslušné datové sady s ikonou informující o typu vizualizace.

Kliknutím na požadovanou datovou sadu se zobrazí stránka s jejím katalogizačním záznamem, který v první části obsahuje její název, popis, periodicitu aktualizace, související geografické území (tj. zda jsou data kumulována na úrovni státu, kraje nebo okresu), mezinárodní klasifikaci oblasti dat podle EuroVoc, odkaz na dokumentaci vybrané datové sady a časový interval publikovaných dat. Druhá část – Datové zdroje – v jednotlivých řádcích uvádí odkazy na soubory, které obsahují publikovaná data ke stažení v uvedeném formátu (typicky RDF/Trig a CSV). Kromě toho obsahuje každý řádek i odkaz na technický popis struktury dat (typicky RDF/Trig a JSON). V poslední části je uveden seznam vizualizací.

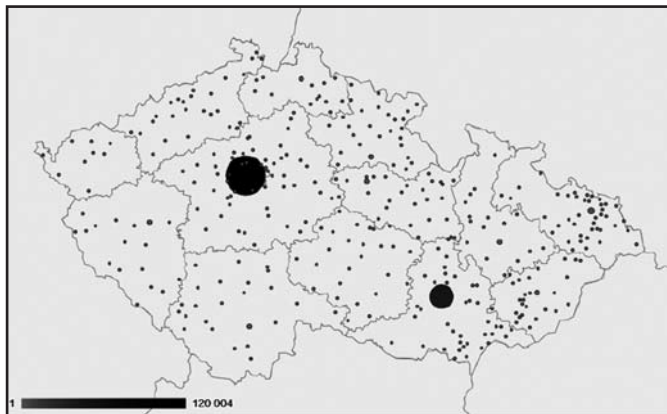
Interaktivní vizualizace umožňují ve většině případů volbu výběrových podmínek pro prezentaci dat. Po nastavení požadovaného filtru (výběrem z hodnot v rozbalovacím seznamu) se po stisknutí tlačítka Zobrazit graf vykreslí data pro zadanou podmínku. Pomocí tlačítka Zobrazit tabulku se zobrazí data odpovídající výběrové podmínce. Poslední tlačítko Export umožňuje exportovat vybraná data ve formátu CSV nebo uložit aktuálně zobrazený graf ve formátu PNG nebo PDF.

DOSAVADNÍ ZKUŠENOSTI A POZNATKY

Zkušenosti a poznatky získané z ročního provozování portálu otevřených dat ČSSZ lze shrnout do následujících bodů:

- a) Poskytovaná data ČSSZ mají specifický charakter vyplývající ze základních činností, které ČSSZ zajišťuje. Určitě nemají tak široké využití jako jiná publikovaná otevřená data, např. jízdní řády, informace o vybavenosti obcí a veřejné registry, což do jisté míry ovlivňuje i okruh typických uživatelů. Jde převážně o datové analytiku, novináře, členy akademické obce, výzkumníky v sociální a demografické oblasti, statistiky apod.
- b) Podle měření přístupů na stránky portálu můžeme konstatovat, že i přes specifčnost dat se průměrný počet přístupů pohybuje v rozmezí 2 500 až 3 000 měsíčně a to z celé republiky (viz obr. 1).
- c) Pro zajímavost lze uvést, že v jednotkách až desítkách evidujeme přístupy z celého světa (viz obr. 2).
- d) Zájem uživatelů je rovněž o samotné datové soubory, zaznamenáváme cca 200 až 300 aktivit stažení měsíčně. Zajímavostí je, že poměrně vyrovnaný je zájem jak o data ve formátu RDF (usnadňuje propojování dat z různých zdrojů), tak i o data ve formátu CSV.

Obrázek 1 – Přístupy na portál z ČR



Obrázek 2 – Přístupy na portál celosvětově



- e) V každodenní praxi se ukázala jako nezbytná potřeba vypracování a naprogramování důkladné evidence, která výraznou měrou zjednoduší proces publikace otevřených dat, jeho vyhodnocování a poskytuje statistické přehledy o publikovaných datech.
- f) Zaznamenali jsme několik dotazů a požadavků na publikovaná data. Jejich počet je výrazně nižší, než bychom předpokládali, což může svědčit o dostatečné srozumitelnosti publikovaných dat, nebo o jisté neatraktivnosti těchto dat pro běžného uživatele. Proto, abychom od uživatelů získali zpětnou vazbu o využití našich dat, jsme na portál umístili výzvu ke spolupráci.
- g) Ve spolupráci s odborem komunikace poskytujeme informace veřejnosti rovněž prostřednictvím sociálních sítí – Twitter a Facebook. I zde zaznamenáváme příznivé reakce jejich uživatelů.





VYUŽITÍ OTEVŘENÝCH DAT ČSSZ

Nejjednodušší formou použití otevřených dat ČSSZ je přímý přístup na portál. Technicky náročnější je práce s propojenými daty ve formátu RDF. V obecné rovině použitý formát RDF usnadňuje propojování dat z různých zdrojů, a tím i analýzu možných souvislostí mezi jevy, které tato data popisují. Jako příklad lze uvést možnost propojení dat o počtu a struktuře důchodců v jednotlivých okresech a letech, které zveřejňuje ČSSZ, s daty o počtu obyvatel a jejich věkové struktuře v okresech publikovanými Českým statistickým úřadem. Výsledkem takového propojení je možnost výpočtu ukazatelů typu počet důchodců na počet obyvatel v okrese, počet invalidních důchodců ve věkových pásmech na počet obyvatel apod., které lze použít pro demografické studie nebo pro rozhodování o aktivitách orgánů veřejné moci v sociální oblasti. Protože jsou data zveřejněna v časových řadách, lze zpracovávat prognózy a odhady dalšího vývoje jako jeden z podkladů pro řízení společnosti. Zajisté lze najít další nepřehledné množství nápadů, jak a s kterými daty (a to nejen ČSSZ, ale i jiných institucí a orgánů státní správy) propojení realizovat a jak takto získané informace využít.

Vzhledem ke specifitě dat ČSSZ bude jejich využití asi nižší než například využití dat o jízdních řádech nebo z volně dostupných celostátních registrů, ale i tak portál otevřených dat ČSSZ zaznamenal od svého spuštění do konce března 2017 již kolem 30 tisíc návštěv.

BUDOUCNOST

Otevřená data jsou vnímána jako jeden z nástrojů podporujících myšlenku tzv. otevřeného vládnutí. Otevřené vládnutí představuje snahu o posílení transparentnosti, demokracie a spolupráce mezi politiky, veřejnou správou, zástupci podnikatelské sféry a občany. Otevřené vládnutí je i jedním z témat strategie nazvané Digitální agenda pro Evropu. Aby se občané a další subjekty mohli efektivně podílet na politické diskusi, potřebují relevantní data a informace. Zvyšování dostupnosti otevřených dat tak pomáhá naplnit tento předpoklad pro vyšší zapojení občanů a dalších subjektů do politické diskuse.

ČSSZ nezustane pouze u dat zveřejněných v uplynulém roce, ale připravuje další publikační plán zahrnující nové datové sady a další rozšiřování vizualizace již zveřejněných otevřených datových sad, aby usnadnila přístup laické veřejnosti k těmto datům. Samozřejmostí pro příští období je zajištění průběžné aktualizace již publikovaných datových sad.

Nakonec bych se chtěl vrátit k výzvě obsažené v titulku tohoto článku. Mnohá data ve veřejné správě a samosprávě jsou většinou výsledkem činnosti hrazené z veřejných prostředků, a proto by měla být rovněž veřejně k dispozici. Publikace otevřených dat plně respektuje a podporuje tuto myšlenku. Tlak veřejnosti by měl směřovat k tomu, aby bylo tímto způsobem publikováno maximum dat a informací (při současném dodržení všech zákonných požadavků na jejich ochranu) využitelných jak pro účinnou kontrolu správní moci, tak i pro vytváření nových aplikací a řešení zjednodušujících každodenní život. Snažme se všichni o to, abychom mohli v oblasti otevřených dat každému jenom doporučit: **„Nežádejte o ně, vezměte si je!“** ■

Autor článku je vedoucím oddělení metodiky správy dat ČSSZ.