



OTEVŘENÁ DATA ČSSZ

Ing. JIŘÍ ŠUNKA

Publikace otevřených dat se stává významnou aktivitou stále většího počtu států a jejich vlád. Celosvětově roste počet oblastí nabízejících sady otevřených dat. Data se dnes stále více dostávají do centra pozornosti, ať už se jedná o otevřená data nebo data velkých objemů (tzv. big data). V případě otevřených dat jde o základní předpoklad pro vytváření hodnoty závislé na schopnosti sdílení těchto dat třetími stranami, nikoliv o jejich objem, jako je tomu u dat velkých objemů.

HISTORIE OTEVŘENÝCH DAT

Pojem Open Data (otevřená data) se poprvé objevil v roce 1995 v dokumentu Americké vědecké agentury, která se zabývala popisem geofyzikálních a environmentálních údajů. Zdůrazňovala se v něm mimo jiné podpora úplné a otevřené výměny vědeckých informací mezi různými zeměmi jako hlavní předpoklad pro analýzu a pochopení globálních jevů.

Již dlouho předtím, než byla otevřená data definována, vědci vnímali výhodu otevřenosti a sdílení dat v oblasti vědeckých objevů a informací. V prosinci 2007 uspořádali internetoví propagátoři v Kalifornii setkání, jehož cílem bylo před nastávající prezidentskou kampaní definovat pojem *otevřené veřejné údaje*. Základní myšlenkou bylo, že veškeré veřejné údaje jsou společným majetkem, hrazeným ze společných prostředků, proto musí být veřejně a svobodně dostupné.

V roce 2007 to znělo jako sen. Ale výsledek časem překročil veškerá očekávání. Přibližně o rok později nastoupil do Bílého domu prezident Barack Obama a podepsal tři prezidentská memoranda. Dvě z nich se týkala otevřeného vládnutí, pro které je otevřeno dat jedním z hlavních pilířů. Otevřenost je koncept společný pro počítačové programy s otevřeným zdrojovým kódem (tzv. open source), otevřenou vládu i otevřená data. Je to filozofie činnosti, praxe a způsobu dosahování stanovených cílů.

Dnes jsou veřejná data ve většině zemí definována zákonem, který se dotýká veřejných a soukromých subjektů zapojených do veřejné služby. Dostupnost mnoha údajů je stále ve vývoji: většina vlastníků dat uděluje prioritu takovým datovým souborům, které lze nejsnadněji otevřít (technicky, právně i politicky). Údaje, které jsou vnímány jako citlivé, nebo ty s vyšším potenciálem sociálního a společenského vlivu zůstávají do značné míry mimo rámec otevřených dat. Myšlenka otevřených dat si našla cestu do společnosti. Nicméně ambice otevřených dat se postupně mění: cílem už není snaha změnit svět prostřednictvím dat, ale skromněji zkusit zmodernizovat veřejnou moc.

CO JSOU OTEVŘENÁ DATA

A co že otevřená data jsou? Jednoduše řečeno, otevřená data jsou úplná, snadno dostupná, strojově čitelná, používající standardy s volně dostupnou specifikací, zpřístupněna za jasně definovaných podmínek s minimem omezení a dostupná uživatelům při vynaložení minima nákladů.

Podle této definice se otevřeným datům přiřazují tzv. stupně otevřenosti, které říkají, do jaké míry splňují výše uvedené parametry. Stupně otevřenosti se označují jednou až pěti hvězdičkami, od nejméně otevřených dat až po maximálně otevřená, splňující bezvýtku výše uvedené požadavky.

Cílem zveřejňování otevřených dat je zejména umožnit jejich opakované použití a zajistit k nim svobodný a neomezený přístup. Z toho vyplývá možnost zpracování těchto dat k libovolným účelům. A v tom se skrývá a postupně odkrývá jejich obrovský potenciál. Vytvářejí se tak podmínky k tvorbě nezávislých programů, které tato data zpracovávají, k vypracování studií a k hledání souvislostí, které by za jiných okolností znamenaly obrovské úsilí. Tím je umožněn i vznik nových pracovních příležitostí a podporován trh práce. Rovněž otevřenost a dostupnost tohoto typu dat je současně velkým přínosem pro zvyšování transparentnosti státní správy.

Primárním úkolem poskytovatele otevřených dat je data zveřejnit, publikovat, nikoliv se zabývat aplikacemi pro jejich využití. To je úlohou okolního prostředí, protože to s velkou pravděpodobností přijde na možnosti využití otevřených dat, které jejich poskytovatele ani nenapadnou. Představíme-li si na jedné straně velký úřad, kde se touto problematikou zabývá několik osob, a na druhé straně internetovou komunitu, je výše uvedený závěr nasnadě. Navíc má poskytovatel dat tendenci uvažovat o svých datech v kontextu své problematiky, a tak lze jen souhlasit s konstatováním Rufuse Pollocka, zakladatele nadace Open Knowledge, že „nejzajímavější využití vašich dat za vás vymyslí někdo jiný“.


ZAČÁTKY V ČSSZ

Na přípravě a realizaci procesu publikace otevřených dat se vedle ČSSZ podíleli i pracovníci Fakulty informatiky a statistiky Vysoké školy ekonomické v Praze (FIS VŠE). ČSSZ se v roce 2014 podílela jako partner na projektu aplikovaného výzkumu *Publikace dat statistických ročenek ve standardu otevřených dat (TD020121)*, který řešila FIS VŠE ve spolupráci se společností KOMIX, s.r.o., a který byl podpořen Technologickou agenturou České republiky. V rámci tohoto projektu byl navržen a na datech statistických ročenek z oblasti důchodového pojištění za roky 2008–2012 pilotně ověřen způsob publikace těchto dat ve formátu RDF s využitím principů propojených dat. Ty umožňují vzájemně propojovat data v prostředí webu a dávat je tak do souvislostí. Publikování otevřených dat v prostředí internetu s využitím principů propojených dat je metodicky a technologicky inovativní přístup, který zatím využívají jenom technicky nejvyspělejší organizace.

Logickým pokračováním bylo pak rozhodnutí vedení ČSSZ o zveřejňování otevřených dat na vlastních webových stránkách. V souvislosti s tímto úkolem byly připraveny odpovídající organizační a technické podmínky. V rámci přípravy bylo přijato zásadní rozhodnutí, že data budou publikována v maximální otevřenosti, že bude vytvořeno vlastní řešení pro publikaci dat a že kromě samotných dat bude ČSSZ zveřejňovat i jejich vizualizace ve formě interaktivních grafů pro jejich snazší interpretaci pro běžného netechnického uživatele.

Výsledkem všech těchto aktivit bylo, že Česká správa sociálního zabezpečení zveřejnila 2. listopadu 2015 na adrese <https://data.csz.cz> portál pro publikaci a vizuali-





zaci otevřených dat. Otevřená data představují jeden ze způsobů, kterým se orgány veřejné správy v různých zemích světa snaží usnadnit využití svých dat (World Wide Web Foundation, 2015). ČSSZ se tak připojila k této mezinárodní iniciativě.

ZÁKONNÉ ODŮVODNĚNÍ

ČSSZ při své činnosti plně respektuje a dodržuje veškerá ustanovení platné legislativy, a to i v případě otevřených dat. V § 4b zákona č. 106/1999 Sb., o svobodném přístupu k informacím, je stanoveno, že *„...informace poskytovaná zveřejněním se poskytuje ve všech formátech a jazycích, ve kterých byla vytvořena; při zveřejnění takové informace v elektronické podobě musí být jeden z těchto formátů otevřený a, je-li to možné, též strojově čitelný. Je-li to možné a vhodné, zveřejní povinný subjekt spolu s informací též metadata, která se k ní vztahují. Formát i metadata by měly co nejvíce splňovat otevřené formální normy...“*. Když 10. září 2015 vstoupila v účinnost tato novela zákona, ČSSZ již byla na tuto změnu připravena.

PŘÍPRAVA OTEVŘENÝCH DAT

Poskytování informací zveřejněním na portálu otevřených dat je náročným procesem, který má své jednotlivé etapy. Před spuštěním vlastního portálu musely být zabezpečeny následující úkony:

- Vypracování návrhu vnitřní organizační směrnice upravující oblast otevřených dat.
- Zajištění podmínek technického charakteru pro realizaci procesu publikace otevřených dat v prostředí ČSSZ jak pro vytváření datových sad, tak pro jejich publikaci a vizualizaci.
- Zajištění bezpečnosti technického řešení i otevřených dat.
- Vypracování publikačního plánu pro zveřejňování otevřených dat na základě vyhodnocení přínosů a rizik.
- Vypracování licenčních podmínek (jinak též podmínek užití) otevřených dat.
- Vytvoření prostředí internetových stránek s otevřenými daty ČSSZ, které obsahují datové sady pro statistické ročenky let 2008 až 2014.
- Implementace lokálního katalogu otevřených dat nebo jeho ekvivalentu na webových stránkách ČSSZ.
- Zajištění evidence v Národním katalogu otevřených dat (dále jen NKOD), který provozuje Ministerstvo vnitra. Tato závěrečná fáze byla zajištěna po zveřejnění otevřených dat na webových stránkách ČSSZ.

Tento proces si vyžádal asi 6 měsíců.

Před publikací otevřených dat proběhla detailní analýza dat a informací, které ČSSZ spravuje, včetně informací dostupných v současné době prostřednictvím webových stránek ČSSZ. Na základě této analýzy byly určeny potenciální datové sady k publikaci, u kterých byly vyhodnoceny přínosy, možná rizika a odhadovaná náročnost jejich publikace. Tím bylo zajištěno, že byla zvolena pouze data, jejichž publikace je v souladu s platnou legislativou, a že byly vhodným způsobem nastaveny priority jejich publikace s ohledem na očekávané přínosy a náročnost samotného zveřejnění dat.

CO ČSSZ ZVEŘEJŇUJE

V současné době je publikováno osmnáct datových sad z oblasti důchodových statistik, z nichž za nejzajímavější lze označit například:

- Měsíční výše důchodů
- Počet nově přiznaných důchodů v České republice dle měsíční výše důchodu
- Počet nově přiznaných důchodů v České republice dle věkové kategorie
- Počet nově přiznaných invalidních důchodů v České republice
- Počet vyplacených důchodů v České republice dle měsíční výše důchodu
- Počet vyplacených důchodů v České republice dle věkové kategorie
- Počet vyplácených invalidních důchodů v České republice dle skupin diagnóz
- Počet zaniklých důchodů v České republice
- Průměrná délka pobírání starobního důchodu
- Průměrná výše důchodů v Kč u nově přiznaných důchodů v České republice
- Průměrný věk u nově přiznaných důchodů v České republice
- Přehled o počtu důchodců podle území, pohlaví, průměrné výše důchodu, průměrného věku a podle druhu důchodu
- Výdaje na důchody
- Statistika žádostí dle zák. 106/1999 Sb., o svobodném přístupu k informacím

Řada z těchto ukazatelů byla přístupná i před publikací ve formátu otevřených dat, nicméně nebyla k dispozici ve strojově čitelném formátu. Velkým přínosem je možnost pracovat s publikovanými daty v časových řadách (ve většině případů od roku 2008 do roku 2014).

Díky analýze, která byla provedena před publikací dat, bylo zajištěno, že byla zvolena pouze data, jejichž zveřejnění je v souladu s platnou legislativou. ČSSZ důsledně chrání osobní údaje svých klientů a publikovaná data představují statistické údaje, tj. agregovaná data. Proto zde nenajdete žádné informace o konkrétních osobách a organizacích.

Velkým přínosem je, že 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 krajů a okresů. U většiny vizualizací lze interaktivně měnit parametry pro výběr dat pro zvolený graf (například rok, druh důchodu, pohlaví ...) a získané výstupy uložit v podobě tabulky ve formátu CSV pro další zpracování nebo jako obrázek grafu ve formátu PNG.

Z publikovaných vizualizací uvádíme například:

- měsíční výše důchodů
- počty nově přiznaných invalidních důchodů v České republice
- průměrnou délku pobírání starobního důchodu u jednotlivých letech
- průměrný věk u nově přiznaných důchodů v České republice
- výdaje na důchody v ČR
- počet důchodců v krajích a okresech ČR

CÍLOVÁ SKUPINA

Charakter zveřejněných dat vyplývá ze základních činností, které ČSSZ zajišťuje. Tím je do jisté míry určen i okruh typických uživatelů. Kromě běžného uživatele, který





projeví individuální zájem o statistické informace z oblasti důchodového a nemocenského pojištění, jsou data určena pro každého, kdo se chce hlouběji zabývat zkoumáním závislostí a vztahů mezi daty poskytovanými různými orgány. Jde o datové analytiku, novináře, členy akademické obce, výzkumníky v sociální a demografické oblasti, statistiky apod. Ti všichni mají možnost svobodného a volného přístupu k publikovaným datům a mohou je zpracovávat dle svého uvážení.

VYUŽITÍ OTEVŘENÝCH DAT

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, věkové struktuře v okresech, které zveřejňuje Český statistický úřad. Výsledkem takového propojení je možnost výpočtu ukazatelů typu počet důchodců na počet obyvatel okresu, 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.

Jak je uvedeno v předchozích částech článku, zajisté se najde mezi uživateli další nepřeborné množství nápadů, jak a s kterými daty propojení realizovat a jak takto získané informace využít.

Vzhledem ke specifice 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 29. února 2016 již více než 25 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 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.

Proto ČSSZ nezůstane pouze u dat zveřejněných v listopadu 2015, ale má zpracovaný podrobný publikační plán zahrnující další datové sady z oblasti lékařské posudkové služby, nemocenského pojištění, evidence osob samostatně výdělečně činných, správního řízení a plánuje i další rozšiřování vizualizace již zveřejněných otevřených datových sad, aby veřejnosti přístup k publikovaným datům usnadnila. ■

Autor článku je vedoucím Oddělení metodiky správy dat ČSSZ.

Zdroje:

<http://www.paristechreview.com/2013/03/29/brief-history-open-data/>