

Koncepce katalogizace otevřených dat VS ČR

(září 2012)

Manažerské shrnutí

Dokument shrnuje základní body Koncepce katalogizace otevřených dat veřejné správy České republiky (dále jen Koncepce), která vznikla v období květen až září 2012 v řešitelském týmu tvořeném odborníky z akademické sféry a veřejné správy (Fakulta informatiky a statistiky Vysoké školy ekonomické v Praze, Matematicko-fyzikální fakulta Univerzity Karlovy v Praze, Ministerstvo vnitra České republiky). Koncepce vznikla v rámci plnění závazku Akčního plánu Partnerství pro otevřené vládnutí „Zpřístupnění dat a informací“ (dále jen Akční plán), který cílí na zlepšení přístupu k datům vytvářených veřejnou správou (VS). Cílem Koncepce je:

- přesně definovat pojem otevřená data VS,
- analyzovat současný stav publikace a katalogizace otevřených dat VS v ČR i ve světě,
- navrhnout a zhodnotit možné varianty budování katalogu otevřených dat VS v ČR,
- vybrat nejlepší variantu a tu detailněji rozpracovat (tj. odpovědět na otázku kdo a jak by měl katalogizovat otevřená data VS),
- zpracovat časový plán realizace vybrané varianty a odhadnout rámcové náklady realizace.

Koncepce definuje otevřená data jako data zveřejněná na internetu způsobem, který neomezuje žádné uživatele ve způsobu jejich použití a opravňuje všechny uživatele k jejich dalšímu šíření, pokud při tomto využití a šíření bude uveden autor dat a pokud i ostatní uživatelé budou mít stejná oprávnění s dále šířenými daty nakládat. Koncepce definuje detailně deset vlastností otevřených dat a dále je člení na obligatorní a fakultativní vlastnosti. Tyto vlastnosti jsou pak využity při specifikaci atributů katalogizačního záznamu. Cílem je zpřístupnit data vytvářená a sbíraná VS odborné i široké veřejnosti způsobem, který umožní opakované použití těchto dat pro různé účely a v řadě různých softwarových aplikací. Aplikace tvoří veřejnost (typicky odborná) sama, čímž mj. dochází ke snižování nákladů VS na prezentaci dat veřejnosti.

Otevřená data VS jsou ve světě běžně publikována (např. ve Velké Británii či USA). VS ČR také publikuje data, ta avšak nesplňují řadu požadavků na otevřená data (např. otevřený a strojově čitelný formát, úplnost dat či jasná specifikace podmínek užití dat). VS navíc publikuje svá data na mnoha místech (každý orgán VS na svých internetových stránkách) a je tak velmi obtížné zjistit, jaká data vlastně publikuje a kde. Řada zajímavých dat je tak pro veřejnost těžko dohledatelná a jejich potenciál zůstává nevyužit. Tyto problémy pomůže vyřešit Datový katalog otevřených dat VS. Orgány VS ČR získají možnost v Datovém katalogu vytvářet o svých datech

záznamy, popisující o jaká data se jedná a kde jsou publikována ke stažení. Veřejnost získá možnost v katalogizačních záznamech jednotným způsobem vyhledávat pomocí různých kritérií v přívětivém uživatelském rozhraní. Veřejnost i samotná VS tak získá na jednom místě přehled o datech publikovaných VS.

Koncepce navrhuje realizovat Datový katalog jako centrální systém, který bude integrován s dvěma již existujícími informačními systémy veřejné správy - Informačním systémem o informačních systémech veřejné správy (IS o ISVS) a Informačním systémem o datových prvcích (ISDP). Koncepce navrhuje Datový katalog tak, aby byl s těmito systémy provázán, a tím nedocházelo k duplicitnímu zadávání údajů v těchto třech katalogizačních informačních systémech.

Vytvářet katalogizační záznamy budou jen ty orgány VS, které data vytvářejí (tzv. původci dat). Veřejnosti nebude (z důvodů bezpečnosti, zajištění kvality záznamů a věrohodnosti) umožněno záznamy vytvářet. Obsah Datového katalogu (tj. katalogizační záznamy) bude veřejnosti plně přístupný pro prohlížení, prohledávání a stahování. Než však bude daný záznam zpřístupněn, bude muset projít redakcí.

Koncepce dále zavádí jednotlivé role uživatelů Datového katalogu. Nejdůležitějšími zúčastněnými stranami a rolemi jsou poskytovatel (orgán VS zveřejňující data), kurátor (osoba určená poskytovatelem zajišťující publikaci a údržbu dat a jejich katalogizaci), redaktor (osoba provádějící obsahovou kontrolu záznamů v Datovém katalogu) a koncový uživatel (kdokoliv, kdo Datový katalog prohledává a prohlíží). Koncepce také navrhuje proces katalogizace a strukturu katalogizačního záznamu. Proces klade velký důraz právě na zajištění kvality záznamů. Popisuje, v jakých krocích bude probíhat redakce a kdo bude opravovat chyby nalezené během redakce.

Navržená struktura katalogizačního záznamu obsahuje kromě základních atributů (jako název a popis katalogizovaných dat) i atributy pro popis souvislostí mezi jednotlivými záznamy, podmínky užití dat, zařazení záznamu do různých klasifikací (např. EUROVOC a CZ-NACE), či vazbu na příslušný ISVS a na datové prvky evidované v ISDP.

Koncepce se také věnuje otázce nutných legislativních změn souvisejících s katalogizací. Doporučuje formulovat jednotný přístup ke stanovování podmínek využití otevřených dat VS a tam, kde je to možné a vhodné, poskytnout uživatelům dat licenci k jejich užití. Aby byl licenční přístup v rámci VS sjednocen, je doporučeno vytvořit vzorové licence pro otevřená data VS. Koncepce se dále zabývá možnostmi vymezit pojem otevřená data VS v zákoně č. 106/1999 Sb., o svobodném přístupu k informacím. Pokud má být katalogizace otevřených dat povinná, bude třeba uložit orgánům veřejné správy tuto povinnost zákonem obdobně, jako je tomu v případě povinnosti předávat údaje o číselnících a datových prvcích do ISDP a povinnosti předávat údaje o spravovaných informačních systémech veřejné správy do IS o ISVS, kdy tyto povinnosti ukládá příslušným orgánům veřejné správy zákon č. 365/2000 Sb., o informačních systémech veřejné správy.

V poslední části Koncepce je popsán plán její realizace. Plán navrhuje následující hlavní etapy realizace koncepce:

- vytvoření jednoduchého (prototypového) Datového katalogu a jeho ověřovací provoz, který by měl skončit do 28.2.2013. V rámci tohoto řešení budou katalogizována publikovaná otevřená data v oblastech uvedených v Akčním plánu,
- příprava a schvalování příslušné legislativy nejpozději do 30.6.2014,
- realizace úplného Datového katalogu počítá s výběrovým řízením a dokončením jeho implementace, zprovoznění a proškolení příslušných pracovníků do 31.12.2013,
- ověřovací provoz Datového katalogu do 30.6.2014,
- zahájení plného provozu Datového katalogu od 1.7.2014.

Na základě navrženého plánu je v poslední části Koncepce odhadována pracnost a finanční náročnost vybudování a provozu Datového katalogu.

Odhady pracnosti vycházejí z počtu orgánů VS a odhadů množství jimi publikovaných otevřených dat ve třech skupinách odhadů (minimální, střední a maximální). V minimální variantě (kterou předpokládáme v prvních dvou letech provozu Datového katalogu, kdy se předpokládá katalogizace čtrnácti set katalogizačních záznamů za rok) je odhadována denní pracnost pro redaktory v objemu 2,6 člověkohodiny (tj. 0,3 FTE) a pro kurátory dat v celkovém objemu za celou veřejnou správu 8 člověkohodin (tj. 1 FTE). Ve střední variantě (kterou považujeme za cílovou do roku 2017 a pro kterou se již předpokládá založení a redakce šest a půl tisíce katalogizačních záznamů za rok) je odhadovaná průměrná denní pracnost pro redaktory 12,5 člověkohodiny (tj. 1,6 FTE) a denní pracnost pro kurátory dat za celou VS (tj. všechny orgány veřejné moci) 37,4 člověkohodiny (tj. celkem 4,7 FTE).

Odhady nákladů na pořízení, implementaci datového katalogu byly vytvářeny na základě odpovědí na žádosti o poskytnutí informací zaslané vybraným dodavatelským společnostem a na základě expertního odhadu členů řešitelského týmu tvorby Koncepce. Odhady nákladů na návrh, vytvoření (přizpůsobení), implementaci, školení a ověřování Datového katalogu se pohybují v rozsahu 5,5 až 9,8 milionu Kč (bez DPH). Vzhledem k uvedeným nákladům bude nutno realizovat veřejnou zakázku dle zák. č. 137/2006 Sb. Dále byl proveden odhad ročních provozních nákladů na údržbu, který se pohybuje ve výši 0,67 až 1,35 miliónu Kč (bez DPH) a odhad nákladů na platy a odvody zaměstnanců v rolích redaktora a kurátora dat, kdy v minimální variantě byly odhadnuty roční náklady na 0,5 miliónu Kč a ve střední variantě na částku 2,6 miliónu Kč.

Koncepce katalogizace otevřených dat veřejné správy České republiky je členěna do tří dokumentů:

- tohoto manažerského shrnutí (3 strany),
- zkrácené verze, která se zaměřuje na specifikaci Koncepce katalogizace otevřených dat veřejné správy a plán její realizace (40 stran),
- plné verze, která obsahuje vymezení vlastností otevřených dat a vhodných formátů pro jejich publikaci, podrobnou analýzu výchozího stavu v ČR, popis obdobných řešení ve světě, popis a porovnání variant řešení Datového katalogu v ČR, výběr cílové varianty a vlastní návrh Koncepce a plán její realizace s odhady pracnosti a financí (111 stran).