



Vzorové publikační plány pro publikaci otevřených dat veřejné správy

Ing. Dušan Chlapek, Ph.D.,
Mgr. Martin Nečaský, Ph.D.,
Mgr. Michal Kubáň, M. A.,
Mgr. Tomáš Kroupa,
Mgr. Jiří Kárník,

Ministerstvo vnitra



Obsah prezentace

- 1. Otevřená data a současná praxe**
2. Vzorové publikační plány
3. Postup validace publikačních plánů

Současná praxe otevřených dat

- Existuje řada OVM, které publikují datové sady v otevřené podobě.
 - např. *MF ČR* nebo *ČOI*
 - např. *seznam smluv* nebo *výsledky kontrol*
- Není sjednocen seznam datových sad, které by měly být publikovány všemi.
- Pokud více OVM publikuje stejný typ datové sady (např. *seznam smluv*), každá opakovaně vymýšlí jiný formát (CSV, XLS(X), JSON, XML, ...) a ad-hoc datové schéma.

Současná praxe otevřených dat

- Celkově je tedy současná situace pro potenciální uživatele datových sad i pro samotné OVM matoucí a nepřehledná.
- OVM nevědí, jaké datové sady mají publikovat a jak je mají publikovat.
- Uživatelé dat nevědí, kde mají datové sady hledat a je velmi obtížné až nemožné je hromadně zpracovávat
 - Různé pojmenování datových sad, různé formáty a datová schémata.



Obsah prezentace

1. Otevřená data a současná praxe
- 2. Vzorové publikační plány**
3. Postup validace publikačních plánů

Vzorový publikační plán

- Vymezuje doporučené datové sady k publikaci v podobě otevřených dat a určuje podobu publikace.
- V čem vzorový publikační plán pomůže?
 - Stanoví doporučované datové sady pro všechny OVM v ČR.
 - Ujednotí způsob publikace těchto sad napříč OVM (formáty, schémata, umístění v lokálním datovém katalogu/NKOD).
 - Usnadní práci jednotlivých OVM při otevírání jejich dat. OVM nemusí
 - analyzovat, jaké datové sady bude publikovat,
 - navrhovat datové formáty a schémata,
 - určovat, jakým způsobem budou datové sady publikovány.
- Je vzorový publikační plán povinný?
 - Zatím se bude jednat o doporučení MV ČR, avšak s možným dopadem do legislativy v budoucnosti.

Typy vzorových publikačních plánů

- Různé typy OVM spravují různé datové sady a mají různé (technické, personální, ...) možnosti.
- Proto budou pro různé typy OVM vytvořeny různé vzorové publikační plány:
 - ústřední orgány – ministerstva
 - ústřední orgány – ostatní
 - krajský úřad
 - obce s rozšířenou působností
 - obecní úřad - ostatní



Obsah vzorového publikačního plánu

- Seznam doporučených datových sad k otevření a datových oblastí, ze kterých je doporučeno datové sady vybírat.
 - včetně doporučeného stanovení priorit a vzorového harmonogramu publikace
- Pro každou datovou sadu popisuje
 - datové schéma, včetně referenčních údajů
 - vzor katalogizačního záznamu
 - typická rizika a přínosy spojené s publikací a doporučené kroky k jejich zmírnění
 - doporučení pro zajištění kvality
- Doporučení pro komunikaci
- Doplněno o příklady konkrétních publikačních plánů



Obce a kraje – specifika přístupu

- Napříč všemi obcemi lze identifikovat řadu stejných datových sad, které obce spravují a lze je doporučit k plošné publikaci.
 - Podobná situace je u krajů.
- Při sestavování vzorového publikačního plánu postupujeme vzhledem k jejich možnostem a kapacitám.
 - Pečlivě vybíráme seznam datových sad s maximálním potenciálem.
 - Definujeme datová schémata s co nejjednodušším formátem

Centrální orgány – specifika přístupu

- Napříč centrálními orgány existuje menší množství stejných datových sad, které lze doporučit k publikaci.
- Plány proto zahrnují řadu datových oblastí (např. doporučovaných G8 Open Data Charter), ve kterých mají centrální orgány primárně identifikovat svoje datové sady.
 - Snahou je naplnit oblasti pomocí datového fondu centrálních orgánů.
 - Plány doporučují i konkrétním centrálním orgánům publikovat konkrétní datové sady.
- Klademe důraz na publikaci referenčních číselníků v otevřené podobě.

Oblasti datových sad





Příklady doporučených datových sad

- Seznam smluv / objednávek / faktur
- Nabídka nepotřebného majetku
- Telefonní seznam
- Seznam poradních orgánů
- Volná pracovní místa
- Katalog knihovny
- Pořádané akce
- ...



Doporučené datové schéma

- Přínosy otevřených dat se projeví, pokud vznikají SW aplikace, které s daty pracují, nebo pokud s daty pracují datoví analytici a publikují svá zjištění.
- V případě typu datové sady, která plošně existuje napříč různými OVM (např. *seznam smluv* či *výsledky kontrol*) se nevyplatí pracovat pouze s daty jedné OVM.
- Je nutné umět pracovat s daty všech OVM, které tento typ datové sady publikují.
- Pokud ale každý OVM vytváří vlastní ad-hoc datové schéma pro datovou sadu, je plošnému využití prakticky zamezeno.
 - Není však na druhou stranu účelné pro každý takový typ datové sady vytvářet nový registr (jako např. *registr smluv*).
 - V řadě případů postačí ujednotit datové schéma (a správně katalogizovat).



Doporučené datové schéma

- Cílem je pro každý takový typ datové sady stanovit doporučené datové schéma, které je snadno rozšiřitelné.
- Co znamená doporučené datové schéma?
 - Bude zdokumentováno v sekci otevřených dat na stránkách MV ČR.
 - Stanovuje podobu základních a napříč různými OVM se vyskytujících datových položek.
- Co znamená rozšiřitelné?
 - Každý si může přidat svoje datové položky.
 - Stejně položky přidané více různými OVM budou podkladem pro aktualizaci doporučeného datového schématu.

Doporučené datové schéma

- Pro tabulková data (XLS(X), ODS, CSV) popisuje sloupce tabulky
 - pouze lidsky čitelná dokumentace
 - 2* + 3*
- Pro XML data popisuje požadované XML elementy a atributy
 - strojově čitelné schéma (XML Schema) a lidsky čitelná dokumentace
 - 3*
- Pro JSON data popisuje požadované JSON atributy
 - strojově čitelné schéma (JSON Schema) a lidsky čitelná dokumentace
 - 3*
- Pro RDF data popisuje požadované třídy a predikáty
 - strojově čitelná ontologie (RDF Schema) a lidsky čitelná dokumentace
 - 4* + 5*

Vzorové katalogizační záznamy

- Pro každou doporučovanou datovou sadu bude vytvořen i předvyplněný vzorový katalogizační záznam.
- Výhody
 - Usnadní a urychlí proces katalogizace datových sad.
 - Kurátoři nemusí vymýšlet položky záznamu jako jsou název, popis, klíčová slova, EUROVOC klasifikace, atd.
 - Zpřehlední datové katalogy (lokální i národní) pro jejich uživatele.
 - Stejně věci budou ve všech katalozích vypadat stejně.
- Vzorové katalogizační záznamy budou vytvořeny v lidsky čitelné podobě i ve strojových formátech NKOD-XML a DCAT-AP.
 - NKOD-XML je datový formát pro zasílání zpráv do NKOD prostřednictvím datových schránek.
 - DCAT-AP je standard EU.



Typické přínosy otevřených dat

- (P1) posílení transparentnosti,
- (P2) podpora ekonomického růstu,
- (P3) zlepšení služeb veřejné správy,
- (P4) podpora opětovného použití dat,
- (P5) zlepšení vnímání veřejné správy veřejností,
- (P6) zlepšení procesů a dat veřejné správy,
- (P7) zamezení chyb vzniklých při práci s daty
- (P8) snížení počtu dotazů dle zák. č. 106/1999 Sb.
- (P9) pořádek ve vlastních datech
- (P10) možnost transparentně informovat veřejnost o subjektech, které neplní své zákonné povinnosti
- (P11) zvýšení hodnoty dat
- (P12) budování dobrého jména

Typická rizika otevřených dat

- (R1) Zveřejnění dat v rozporu se zákonem
- (R2) Riziko porušení ochrany obchodního tajemství
- (R3) Riziko porušení ochrany osobních údajů
- (R4) Zveřejnění nevhodných dat či informací
- (R5) Dezinterpretace dat
- (R6) Absence konzumentů dat
- (R7) Překrývání dat

Stupně otevřenosti a datové formáty



Zdroj: HAUSENBLAS, Michael. 5 star Open Data. Dostupné z: <http://5stardata.info/>

Otevřená data – stupně otevřenosti

Stupeň: ★

- Data poskytována pod otevřenou licencí či podmínkami užití umožňujícími jejich další užití
- Data poskytována ve strojově čitelném formátu

Formáty: strojově čitelné PDF či DOC(X) či ODT, HTML

- Výhody:
 - Jednoduchost a relativně nízká pracnost (data není třeba transformovat, zaměření pouze na právní otevřenost)
 - Uživatelé vědí, že mohou data dále zpracovávat
- Nevýhody:
 - Data může být obtížné využít – např. potřeba vytěžování tabulkových dat z PDF dokumentů (příklad: tabulky s údaji v ročenkách a výročních zprávách)

Otevřená data – stupně otevřenosti

Stupeň: ★★

- Data poskytována pod otevřenou licencí či podmínkami užití umožňujícími jejich další užití
- Data poskytována v libovolném formátu

Formáty: XLS(X), ODS

- Výhody:
 - Relativně jednoduché, pokud jsou podkladová data již dostupná ve formátu typu MS Excel, nebo pokud je lze takového formátu jednoduše uložit
 - Data jsou ve formátu, který je snáze strojově zpracovatelný
- Nevýhody:
 - Pokud neexistují volně dostupné nástroje pro práci se zvolenými formáty, je uživatel nucen pořizovat odpovídající sw nástroje

Otevřená data – stupně otevřenosti

Stupeň: ★★ ★

- Data poskytována pod otevřenou licencí či podmínkami užití umožňujícími jejich další užití
- Data poskytována ve strojově čitelném formátu
- Formát dat je otevřený:
 - Specifikace formátu je volně dostupná
 - Lze využívat zdarma, další využití formátu není omezeno
 - Formát nezávislý na platformě, resp. lze vytvořit nezávislé implementace pro různé platformy

Formáty: CSV, XML, JSON

- Výhody:
 - Uživatelé nejsou nuceni používat aplikace určitého výrobce, aby s daty mohli pracovat
- Nevýhody:
 - Může být nutné data do otevřeného strojově čitelného formátu transformovat

Otevřená data – stupně otevřenosti

Stupeň: ★★★★★

- Data poskytována pod otevřenou licencí či podmínkami užití umožňujícími jejich další užití
- Data poskytována ve strojově čitelném formátu
- Formát dat je otevřený
- Jako identifikátory objektů jsou použity URI (Unified Resource Identifier)

Formáty: RDF (Resource Description Framework)

- Výhody:
 - Objekty jsou jednoznačně identifikovány způsobem, který umožňuje se na ně odkazovat obdobně jako na HTML stránky
 - Lze kombinovat s jinými datovými sadami na stupních 4 a 5 hvězdiček
- Nevýhody:
 - Příprava dat vyžaduje více času a úsilí – definice schémat pro tvorbu URI a přiřazení URI identifikátorů objektům
 - Ne všichni v současné době disponují znalostmi pro publikaci a zpracování dat v této podobě

Otevřená data – stupně otevřenosti

Stupeň: ★★★★★

- Data poskytována pod otevřenou licencí či podmínkami užití umožňujícími jejich další užití
- Data poskytována ve strojově čitelném formátu
- Formát dat je otevřený
- Jako identifikátory objektů jsou použity URI
- Data jsou pomocí odkazů propojena na jiná související data

Formáty: RDF (Resource Description Framework)



Obsah prezentace

1. Otevřená data a současná praxe
2. Vzorové publikační plány
3. **Postup validace publikačních plánů**



Validace vzorových publikačních plánů

- Po vytvoření první verze vzorových publikačních plánů proběhne jejich validace vybranými validátory.
- Pro každý druh OVM, pro který vytváříme specifický publikační plán, vybíráme několik validátorů z konkrétních OVM.
- Vybraní validátoři budou připomínkovat vzorové publikační plány, zejména
 - sestavení doporučených datových sad a jejich doporučených datových schémat
 - přínosy a rizika spojená s publikací doporučených datových sad
- Na základě připomínek bude vytvořena druhá verze, která bude vystavena na stránkách MV ČR



Postup validace

- Výběr validátorů (do 30.4.2015)
- Poskytnutí první verze vzorových publikačních plánů (30.4.2015)
- Workshopy s validátory (červen až srpen 2015)
- Zapracování připomínek a vytvoření druhé verze publikačních plánů (30.10.2015)