

MINISTERSTVO FINANČÍ
Praha 1, Letenská 15

V Praze dne 12. prosince 2014
Č.j.: MF – 69 949/2014/4703-2

S M Ě R N I C E č. 6/2014
ministra financí

Věc: **Publikace a katalogizace otevřených dat Ministerstva financí**

Čl. 1

Předmět úpravy a závaznost

1. Účelem směrnice je upravit jednotný postup při publikaci a katalogizaci otevřených dat Ministerstva financí (dále také jen „ministerstvo“) ve vztahu k veřejnosti.
2. Směrnice je závazná pro zaměstnance ministerstva, generálního ředitele Generálního ředitelství cel a generálního ředitele Generálního finančního ředitelství. Generální ředitel Generálního finančního ředitelství a generální ředitel Generálního ředitelství cel zajistí provedení této směrnice v orgánech finanční a celní správy.

Čl. 2

Pojmy

1. Za **otevřená data** se považují data, která splňují následující podmínky:
 - a) úplnosti; úplnost je zabezpečena publikováním v maximálním možném rozsahu,
 - b) snadné dostupnosti, data jsou dostupná na internetu a dohledatelná běžnými informačními a komunikačními technologiemi (ICT),
 - c) strojové čitelnosti; data jsou uvedena ve formátu, který je strukturovaný takovým způsobem, že pomocí běžné softwarové aplikace z nich lze získat požadované údaje,
 - d) používají standardy s volně dostupnou specifikací (otevřené standardy); data jsou ve formátu, který je volně (bezplatně) dostupný pro libovolné použití nebo do takového formátu převoditelný volně (bezplatně) dostupnou softwarovou aplikací,
 - e) zpřístupnění za jasně definovaných podmínek užití dat s minimem omezení (podmínky užití); podmínky užití dat jsou jasně a zřetelně definovány a publikovány,
 - f) dostupnosti uživatelům při vynaložení minima možných nákladů na jejich získání.
2. **Datová sada** je množina souvisejících dat, vztahující se k jednomu věcnému tématu. Datová sada je tvořena jedním nebo více datovými soubory, tedy strukturovanými seznamy položek.
3. **Otevřená datová sada** je sada, která je zpřístupněna prostřednictvím webových stránek a splňuje definici otevřených dat.

4. **Transformace dat** je změna obsahu a struktury datové sady, např. anonymizace, agregace či jiná úprava původních dat.

5. **Katalogizační záznam** tvoří strukturovaná data o určité otevřené datové sadě (metadata) v katalogu otevřených dat; struktura katalogizačního záznamu je uvedena v příloze této směrnice a představuje doporučený minimální rozsah atributů pro popis datových sad.

6. **Katalog otevřených dat** je softwarová aplikace umožňující vytváření, úpravu, správu a případné rušení katalogizačních záznamů. Tyto záznamy prezentuje v plném rozsahu uživatelům sítě Internet, včetně funkcionality vyhledávání a získání přehledového seznamu všech katalogizačních záznamů.

Čl. 3

Přehled rolí a jejich obsazení

Pro práci s datovými sadami jsou rozlišovány role uvedené v následující tabulce:

Role	Obsazení
Poskytovatel	Ministerstvo financí
Kurátor	Určený zaměstnanec věcně příslušného útvaru ministerstva, který zodpovídá za zveřejnění konkrétní datové sady ve formě otevřených dat.
Správce katalogu otevřených dat	Odbor 33 - Řízení a provoz ICT resortu
Analytický útvar	Odbor 47 - Centrální harmonizační jednotka

Čl. 4

Práva a povinnosti rolí

1. Poskytovatelem datových sad je ministerstvo, které zodpovídá za:
 - a) identifikaci datových sad k publikaci a jejich publikaci ve formě otevřených dat,
 - b) identifikaci potenciálních rizik plynoucích z publikace otevřených dat,
 - c) kontrolu souladu publikace otevřených dat s právními předpisy ČR,
 - d) určování podmínek, za jakých jsou otevřená data publikována, a to zejména právních, ekonomických, časových a technických.

2. Kurátor je jmenován ministrem, role se váže na konkrétní pracovní pozici. Kurátor datové sady zodpovídá zejména za:

- a) zajištění analýzy datové sady předcházející zveřejnění dat,
- b) koordinaci procesu zveřejnění datové sady dle postupu navrženého Správcem katalogu otevřených dat,
- c) zajištění, že otevřená datová sada je publikována v souladu s právními předpisy ČR,
- d) poskytování informací o Otevřené datové sadě veřejnosti v návaznosti na dotazy; poskytování návrhu aktualizace katalogizačního záznamu Správci katalogu otevřených dat, vyplyne-li z dotazů veřejnosti zásadní nejasnost.

3. Správce katalogu otevřených dat zodpovídá za:

- a) zajištění implementace a provozu katalogu otevřených dat,
- b) správu a formální kontrolu katalogizačních záznamů,
- c) stanovení procesního a technického postupu publikace datových sad,
- d) poskytování technické podpory a součinnosti Kurátorovi při plnění povinností plynoucích z této směrnice, zejména při analýze předcházející zveřejnění dat a jejich průběžné kontrole.

4. Analytický útvar zodpovídá za:

- a) sběr podnětů, přípravu a předkládání návrhu datových sad pro zveřejnění,
- b) zpracování výroční zprávy o zveřejňování otevřených dat,
- c) poskytování metodické podpory Kurátorovi při plnění povinností plynoucích z této směrnice,
- d) předložení návrhu neúplných datových záznamů Správci katalogu otevřených dat v návaznosti na rozhodnutí ministra o publikaci Otevřené datové sady.

Čl. 5

Postup publikace a katalogizace otevřených dat

1. Analytický útvar předkládá ministrovi návrh datových sad k publikaci. Návrh datových sad konzultuje Analytický útvar s navrhovaným Kurátorem. Tento návrh zahrnuje:

- a) zevrubný popis datové sady,
- b) návrh Kurátora,
- c) předběžný odhad pracnosti a spojených právních rizik,
- d) doporučení ohledně rozsahu dat a periodicity publikace,
- e) identifikaci skupiny potenciálních uživatelů datové sady (nepovinně),
- f) doporučení termínu první publikace datové sady,

g) stanovisko navrženého Kurátora.

2. O publikaci Otevřené datové sady rozhoduje ministr, který vymezí datovou sadu, očekávaný termín publikace a Kurátora. Kurátor dále odpovídá za koordinaci procesu publikace této datové sady jako otevřené.

3. Rozhodnutí podle odstavce 2 se nepovažuje za rozhodnutí ministra ve smyslu směrnice č. 2/2005 ministra financí, o vnitřních předpisech Ministerstva financí, ve znění dodatků.

4. Proces publikace zahrnuje přípravu datové sady, katalogizaci a samotné zveřejnění datové sady. Proces zveřejnění je opakován ve stanovené periodicitě.

5. Příprava datové sady pro zveřejnění ve formě otevřených dat obsahuje analýzu datové sady a stanovení procesního a technického postupu publikace, a to.

a) analýzu datové sady zahrnující povinnosti:

aa) stanovit detailní rozsah publikovaných dat jako maximálně možný, který neklade nepřiměřené náklady na proces zveřejnění a zároveň není v rozporu s právními předpisy ČR,

ab) ověřit kvalitu dat, zejména jejich úplnost, konzistenci a správnost; dále popsat zjištěné nedostatky včetně jejich četnosti a závažnosti,

ac) určit případnou nutnost transformace dat, zejména s ohledem na řešení problémů identifikovaných v bodech aa) a ab),

ad) určit periodicitu publikace otevřených datových sad, jako nejkratší možnou, která neklade nepřiměřené ekonomické náklady a je zároveň uživatelsky přínosná,

ae) detailně popsat data, zejména jejich věcný obsah, a rozsah;

af) konzultovat zvolený rozsah, periodicitu a transformaci dat se zástupci potenciálních uživatelů, pokud jsou tito zástupci identifikováni v návrhu na zveřejnění datové sady, nebo pokud tito zástupci kontaktovali ministerstvo v návaznosti na zveřejnění neúplného katalogizačního záznamu; zapracovat jejich případné připomínky v případech, ve kterých je to právně a ekonomicky přípustné a nepředstavuje nepřiměřená rizika pro poskytovatele,

ag) stanovit podmínky užití datové sady s využitím otevřené licence nebo jako volné dílo,

ah) konzultovat výstupy analýzy datové sady s Analytickým odborem a zapracovat jeho eventuální připomínky,

ai) navrhnout katalogizační záznam a příložené dokumenty ve struktuře dle přílohy této směrnice (s výjimkou bodů označených jako technické), které zejména vycházejí z předcházejících bodů analýzy datové sady.

- b) stanovení procesního a technického postupu publikace:
 - ba) určit formáty zveřejněných datových sad splňující požadavky přílohy,
 - bb) zpracovat návrh, implementaci a testování postupů pro převod dat do cílového formátu,
 - bc) vyhotovit návrh optimálního technického zajištění samotné publikace,
 - bd) stanovit odhad prvotních nákladů nastavení procesu publikace a nákladů jedné další aktualizace dat zahrnující zejména náklady na straně kurátora a správce katalogu otevřených dat,
 - be) vytvořit dokument popisující navržený postup publikace, včetně nákladů, evidence a poskytnutí tohoto dokumentu Analytickému útvaru a kurátorovi,
 - bf) vyhotovit a aktualizovat katalogizační záznam na základě návrhu Kurátora; doplnit body označené jako technické.

6. Výstupem z přípravy datové sady pro zveřejnění ve formě otevřených dat může být závěr, že datovou sadu není možné zveřejnit v jakémkoliv rozsahu a periodicitě z důvodů právních, ekonomických či technických. V takovém případě kurátor předkládá ministrovi zprávu popisující tyto důvody a obsahující stanoviska Kurátora, Analytického odboru, příp. Správce katalogu otevřených dat. Tato zpráva je zveřejněna v katalogu otevřených dat spolu s neúplným katalogizačním záznamem. Tímto je proces publikace ukončen.

7. Zveřejnění datové sady se řídí navrženým postupem a zahrnuje:

- a) transformaci datové sady za předpokladu, že je možné ji provádět a převod datové sady do cílového formátu,
- b) kontrolu kvality dat, zahrnující kontrolu správnosti, úplnosti a konzistenci v návaznosti na možné změny,
- c) publikaci datové sady včetně uveřejnění odkazu v katalogu otevřených dat,
- d) v případě opakované publikace ověření správnosti a klasifikace katalogizačního záznamu, případné opravy katalogizačního záznamu.

Čl. 6

Katalog otevřených dat

1. Pro zajištění snadného přístupu k otevřeným datům ministerstva je na stránce data.mfcr.cz zpřístupněn katalog otevřených dat.

2. Katalog obsahuje katalogizační záznamy veškerých Otevřených datových sad zveřejněných dle této směrnice.

3. Katalog obsahuje rovněž neúplné katalogizační záznamy k datovým sadám, o jejichž publikaci rozhodl ministr. Tyto záznamy obsahují přinejmenším datum rozhodnutí ministra ve smyslu čl. 5 odst. 2, identifikaci Kurátora, očekávaný termín zveřejnění, název a popis datové sady.

Čl. 7

Závěrečná ustanovení

1. Kontrolu plnění této směrnice vykonává odbor 47 - Centrální harmonizační jednotka (Analytický útvar).
2. Analytický útvar na základě kontroly prováděné podle odstavce 1 předkládá ministrovi jednou ročně výroční zprávu o zveřejňování otevřených dat. Účelem této zprávy je podat popis zejména stavu zveřejňování jednotlivých datových sad a dále informaci o četnosti využití datových sad třetími subjekty.
3. Nedílnou součástí této směrnice je příloha – Specifikace publikace otevřených dat.
4. Tato směrnice nabývá účinnosti dnem 1. ledna 2015.

Ing. Andrej Babiš, v.z.
ministr financí

PŘÍLOHA Č. 1: SPECIFIKACE PUBLIKACE OTEVŘENÝCH DAT (dále jen „Specifikace“)

Formát dat

- a) Datové sady jsou zveřejněny nejméně v jednom z následujících formátů: XLSX, CSV, JSON, XML. Pro datové sady s dostatečně jednoduchou strukturou jsou preferovány formáty XSLX, CSV.
- b) Data jsou poskytována ve formě ucelených souborů. Zároveň mohou být poskytována i formou REST webové služby.
- c) Data jsou zveřejňována primárně v režimu volných děl či v režimu otevřených licencí, které umožňují nejširší možnosti využití zpracovaných datových sad.
- d) Datové soubory mohou být vhodným způsobem rozděleny, aby usnadňovaly praktické využití. Například mohou být periodicky zveřejňovány verze datasetu při zachování konvence v struktuře URL.

Struktura katalogizačního záznamu

Název datové sady	Výstižné pojmenování datové sady.
Popis	Text, který jasně uvádí, o jaká data se jedná a co popisují. Pokud je to vhodné, obsahuje bodový seznam datových položek s popisem. Pokud je to pro správné pochopení nutné, je třeba uvést i o čem data nevytvírají.
Periodicita publikace datové sady	Denně/Týdně/Měsíčně/Čtvrtletně/Pololetně/Ročně/Opakovaně s jinou periodicitou/Jednorázově.
Aktualizace datové sady	Datum, kdy naposledy došlo ke změně vlastních dat datové sady (např. k přidání datového souboru). Ve formátu: dd.mm.rrrr.
Datum rozhodnutí o publikaci datové sady	Datum, kdy ministr rozhodl o publikaci datové sady
Poskytovatel dat	Ministerstvo financí České republiky.
Kurátor	Určený zaměstnanec příslušného útvaru ministerstva, který zodpovídá za správnost a aktuálnost věcného obsahu datové sady a za realizaci kroků v rámci procesu publikace příslušné datové sady dle stanovené kompetence, resp. datových sad určených k publikaci ve formě otevřených datových sad. K dispozici je minimálně jméno a emailový kontakt.
Odpovědný útvar	Číslo a název útvaru ministerstva, který datovou sadu vytvořil. Typicky útvar ke kterému je příslušný kurátor.
Související geografické území	Kódy NUTS. Pokud se data vztahují např. k celé České republice, použije se NUTS pro ČR. Pokud se data vztahují k pouze k určitému kraji, použije se kód daného kraje.
Klasifikace	Každá datová sada je zařazena jedním nebo více identifikátory dle klasifikace EUROVOC (http://eurovoc.europa.eu). Měly by být zvoleny takové identifikátory, které nejlépe vystihují, o čem otevřená datová sada vypovídá. Pro každý přiřazený identifikátor se uvádí: 1) Název 2) Odkaz (odkaz na výše uvedený web)
Podmínky užití	Stanovení režimu užití datové sady odkazem na vhodnou otevřenou licenci nebo jako volné dílo.
Schéma popisující strukturu datového souboru (technický bod)	Odkaz na schéma tam, kde to bude mít smysl, tj. zejména u dat, která budou v XML. Pokud není schéma, bude kolonka prázdná.

Dokumentace datového souboru	Odkaz na soubor s dokumentací. Dokumentace typicky obsahuje výstupy analýzy datové sady – detailní popis včetně známých nedostatků a metodiku získávání.
-------------------------------------	--

Soubory ke stažení (technický bod)			
Název souboru dat	Formát	Velikost	Datum publikace
Výstižný název souboru dat, který bude současně sloužit jako odkaz ke stažení datového souboru.	Formát souboru. Je doporučeno používat MIME Type.	Velikost souboru.	Datum, kdy byl soubor publikován jako otevřená data na webových stránkách ministerstva.

Konvence URL entit katalogu ministerstva

Každý datový soubor publikovaný v rámci Otevřené datové sady má vlastní stabilní URL. Ta jej jednoznačně určuje, zároveň dodržuje konvenci, díky které lze datový soubor identifikovat, včetně náležitosti ke konkrétní datové sadě, případně datu publikace.

- Jako základ URL bude užit vzor `http://data.{domena2}.{domena1}/resource`, kde `{domena1}` je doména prvního řádu a `{domena2}` je doména druhého řádu organizace. Toto URL budeme dále nazývat *základní URL entit* a budeme jej na dalších řádcích označovat pomocí `{base-URL}`.
 - `http://data.mfcr.cz/resource`
- URL katalogu doporučujeme ve tvaru `{base-URL}/catalog`
 - `http://data.mfcr.cz/resource/catalog`
- URL datové sady doporučujeme ve tvaru `{base-URL}/dataset/{id-datové-sady}`, kde `{id-datové-sady}` je libovolný řetězec, který lze syntakticky umístit do URL a který jednoznačně identifikuje datovou sadu. Doporučujeme jej vytvořit z názvu datové sady.
 - Např. pro datovou sadu s kontrolami provedenými Ministerstvem financí by URL mohlo být `http://data.mfcr.cz/resource/dataset/kontroly`
- URL souboru v rámci datové sady ve tvaru `{base-URL}/dataset/{id-datové-sady}/{id-distribuce}`, kde `{id-distribuce}` je libovolný řetězec, který lze syntakticky umístit do URL a který jednoznačně identifikuje distribuci v rámci datové sady.
 - Např. pro distribuci datové sady s kontrolami provedenými Ministerstvem financí k 31. 3. 2014 reprezentované ve formátu CSV by URL mohlo být `http://data.mfcr.cz/resource/dataset/kontroly/2014-03-13.CSV`