

## **Otevřená data Ministerstva financí**

### **Praktický návod k jejich publikaci**

## Obsah

<b>Otevřená data Ministerstva financí</b> .....	1
1 Úvod.....	1
2 Interní předpis .....	1
2.1 Role.....	1
2.1.1 Poskytovatel dat.....	1
2.1.2 Kurátor vybrané datové sady .....	1
2.1.3 Správce dat .....	1
2.1.4 Analytický útvar .....	2
2.2 Technické parametry katalogu .....	2
3 Projekt.....	2
3.1 Obsah katalogu .....	2
3.1.1 Zákon č. 106/1999 Sb., o svobodném přístupu k informacím.....	3
3.1.2 Akademická obec .....	3
3.1.3 Neziskový sektor .....	3
3.1.4 Soukromý sektor .....	3
3.1.5 Analýza výstupů informačních systémů .....	4
3.1.6 Komunikace napříč organizací.....	4
3.2 IT řešení.....	4
3.3 Projektové řízení .....	4
3.3.1 IT řešení .....	4
3.3.2 Analýza datových sad a určení kurátorů .....	4
3.3.3 Komunikace s kurátory datové sady .....	5
3.3.4 Analýza možností zveřejňování dat.....	6
3.3.5 Komunikace mezi kurátory a správcem katalogu .....	6
3.3.6 Řízení rizik .....	6
4 Údržba a rozvoj.....	7
5 Příprava.....	7
5.1 Požadavky na katalog otevřených dat .....	7
5.2 Výběr technického řešení .....	7
5.3 Test proveditelnosti.....	8
6 Instalace .....	8
7 Úpravy.....	8
7.1 Lokalizace .....	8
7.2 Struktura datové sady .....	8
7.3 Vzhled.....	9
7.4 Analýzy návštěvnosti.....	9

7.5	Optimalizace výkonu .....	9
8	Publikace.....	9
8.1	Spuštění .....	9
8.2	Tisková konference .....	10

# 1 Úvod

Tento dokument popisuje projekt publikace otevřených dat Ministerstva financí ČR.

## 2 Interní předpis

Prvním krokem k publikaci otevřených dat je vytvoření interního předpisu, který definuje role, odpovědnosti, celý proces publikace a technické parametry katalogu. Příslušná směrnice, která vznikla ve spolupráci s poradci v oblasti otevřených dat, je přílohou č.1 tohoto návodu.

Během tvorby předpisu je vhodné využívat dobré praxe z jiných zemí a komunikovat s odbornou veřejností, která se touto problematikou dlouhodobě zabývá. Problematickou otevřených dat se v ČR zabývá zejména neziskový sektor a vybrané vysoké školy. Dále je pak vhodné využívat zkušeností ostatních organizací, které se rozhodly otevírat svá data.

### 2.1 Role

V procesu publikace a katalogizace otevřených dat figurují čtyři hlavní role:

- Poskytovatel dat
- Kurátor datové sady
- Správce dat
- Analytický útvar

#### 2.1.1 Poskytovatel dat

Organizace, která data otevírá. V našem případě je to tedy samotné Ministerstvo financí. Poskytovatelem dat může být magistrát, radnice, obec, příspěvková organizace, úřad atd.

#### 2.1.2 Kurátor vybrané datové sady

Kurátorem datové sady je konkrétní zaměstnanec, který má k datům nejbližší. V praxi se stává, že vzhledem k povaze dat není možné určit jednoznačně kurátora datové sady. Například, když je datová sada export z informačního systému, do kterého vkládá data více útvarů. V tomto případě by měl být kurátorem datové sady určen zaměstnanec, který spravuje informační systém a datovou sadu exportuje. Kurátor pak zajišťuje hlavně pravidelnost exportu, konzistenci datové sady a odesílání správci dat.

Pokud je kurátorem zaměstnanec, který nemůže ručit za věcnou správnost všech informací v datové sadě, jsou další povinnosti delegovány na jednotlivé útvary, které konkrétní data buď vkládají do systému, nebo je jakýmkoliv jiným způsobem odesílají. Pokud se jedná o systém, který organizace (poskytovatel dat) pouze provozuje, a data do něho vkládají jiné organizace, je nutné toto zdůraznit v popisu datové sady.

#### 2.1.3 Správce dat

Správce dat je útvar, který provozuje katalog z technického hlediska, většinou tedy ICT oddělení. Povinností správce dat je v první řadě zprovoznění katalogu po technické stránce a posléze jeho údržba, rozvoj a nahrávání souborů. V první fázi, zprovoznění katalogu, by měl správce dat úzce spolupracovat s analytickým útvarem, aby byla shoda na finální podobě a funkcích katalogu.

## 2.1.4 Analytický útvar

Analytický útvar zastřešuje a řídí celý proces. V praxi je nejdůležitější zodpovědností projektové řízení celé publikace a katalogizace otevřených dat. Je nutné, aby analytický útvar měl přehled a projekt řídil od chvíle vstoupení směrnice v platnost až do zprovoznění a zpřístupnění celého katalogu a dále po celou dobu fungování katalogu.

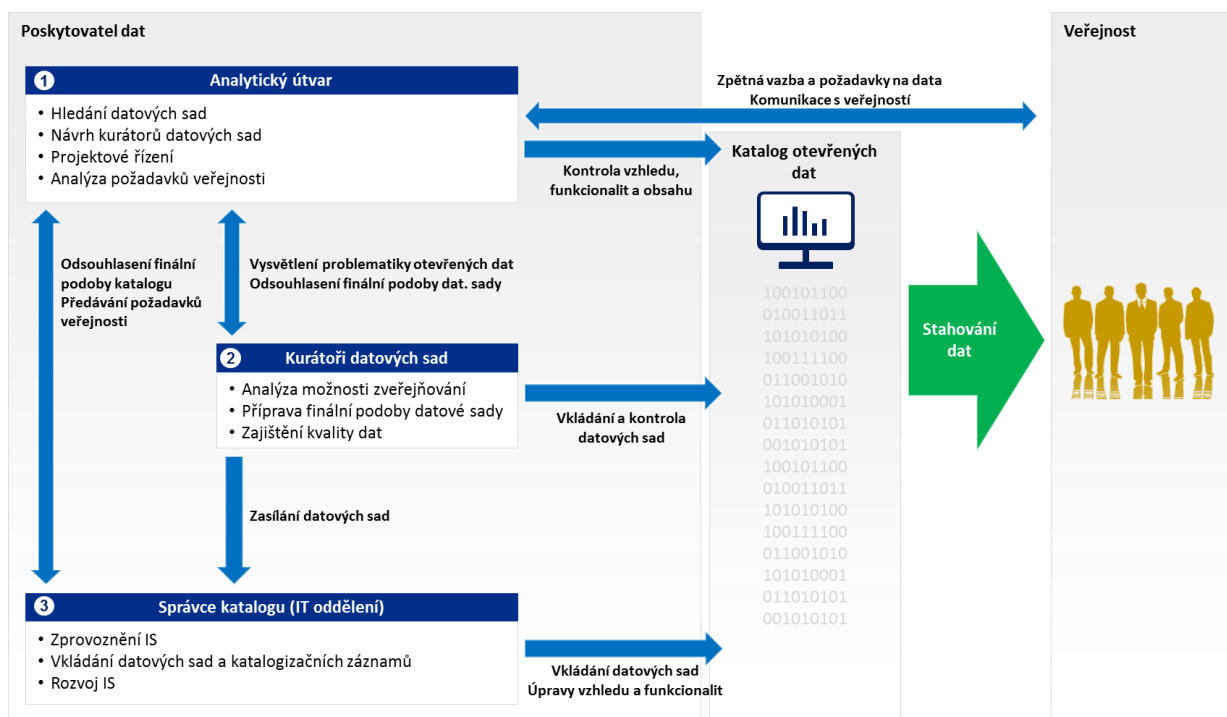
Analytický útvar navrhuje datové sady a jejich formát ke zveřejnění, komunikuje se správcem dat, s právním oddělením, s kurátory datových sad a zajišťuje vše potřebné pro co nejhladší průběh projektu.

## 2.2 Technické parametry katalogu

Technické parametry katalogu otevřených dat a povolené formáty datových sad vycházejí z definice otevřených dat, kterou lze, spolu s podrobným popisem jednotlivých formátů, nalézt například v metodice Ministerstva vnitra k otevřeným datům<sup>1</sup>. Podrobnější informace jsou uvedeny v kapitole 3.2 IT řešení.

## 3 Projekt

Následující schéma popisuje proces publikace otevřených dat:



### 3.1 Obsah katalogu

V této kapitole jsou popsány způsoby, jakými lze analyzovat poptávku po datech.

Na začátku projektu je potřeba najít poptávku po datech tak, aby výsledný obsah katalogu odpovídal tomu, co od organizace očekává široká veřejnost, akademická obec, neziskový sektor a soukromý sektor. Jednat se může jak o zveřejňování nových dat, ale často také může jít také o úpravy formátu či exporty z již existujících systémů.

<sup>1</sup> Metodika publikace otevřených dat veřejné správy ČR 2012, verze 1.0 na adrese <http://www.mvcr.cz/clanek/otevrena-data.aspx>

Dále je vhodné zjistit, jaká všechna data vlastně organizace vlastní a spravuje. V některých organizacích mohou být desítky různých informačních systémů, které mají možnost exportu zajímavých dat. Příklady dobré praxe z jiných zemí ukazují, že i data, která na začátku nikdo explicitně nepožadoval, mohou být nakonec velkým přínosem jak pro veřejnost, tak i pro organizaci samotnou. Například v Londýně byla zveřejněna data o městské hromadné dopravě, což mělo za následek vznik cca 500 projektů a aplikací postavených na těchto datech, pěti tisících pracovních míst a milionům ušetřených liber - obvykle díky rychlejší dopravě a efektivnějšímu využití.

Není nutné mít před zprovozněním katalogu kompletní seznam datových sad a zbytečně tak zprovoznění oddalovat. Je dobré mít k dispozici alespoň nějaké zajímavé datové sady (například přehledy různých výdajů organizace). To pomůže k lepší propagaci katalogu a většímu zájmu široké veřejnosti. Ostatní datové sady je možné nahrávat později a postupně katalog rozšiřovat na základě dalších analýz a poptávek, která bude reakcí na již funkční katalog. Je to jeden z ověřených postupů na vývoj webových stránek, který minimalizuje náklady na jejich vývoj a urychluje uvedení do provozu.

### 3.1.1 Zákon č. 106/1999 Sb., o svobodném přístupu k informacím

Jako první by měla organizace analyzovat dotazy veřejnosti vzniklé na základě zákona č. 106/1999 Sb., o svobodném přístupu k informacím. Umístění takových dat do katalogu organizaci usnadní práci a nebude muset vyřizovat tolik žádostí. Analytický útvar by měl projít a analyzovat všechny žádosti a z výsledku udělat přehled o nejčastějších žádostech a jejich vyhovění či zamítnutí. Pokud je v organizaci útvar přímo zodpovědný za komunikaci s veřejností, bude mít přehled o tom, co běžně poskytuje, co poskytovat nelze a co lze poskytovat za poplatek.

### 3.1.2 Akademická obec

Vysoké školy mají velký potenciál pro zpracování a analýzu různých dat, které veřejný sektor spravuje. Velmi často samotní studenti či doktorandi hledají data, se kterými by mohli pracovat. Výsledkem potom mohou být bakalářské, diplomové nebo disertační práce ale i různé aplikace nad daty. Tyto aplikace, které vzniknou bez jakýchkoliv nákladů, mohou být následně přínosem i pro samotnou organizaci.

Analytický útvar by proto měl komunikovat i s akademickou obcí. Je pravděpodobné, že většina vysokých škol, které se věnují výzkumu dat, bude ochotna spolupracovat a dodá své požadavky, které je posléze možné vyhodnotit. V České republice jsou to například Institut ekonomických studií, Vysoká škola ekonomická v Praze, České vysoké učení technické v Praze nebo Matematicko-fyzikální fakulta UK.

Je také vhodné nastavit komunikaci tak, aby fungovala i po uvedení katalogu do provozu, aby jednotlivé vysoké školy mohly připomínkovat zveřejněné datové sady a dodávat další podněty.

### 3.1.3 Neziskový sektor

Vybrané neziskové organizace se zabývají tématy spojenými s efektivitou fungování veřejné správy, korupcí, ale i například podporou podnikání a mohou proto mít také dobré nápady na otevírání dat. V České republice jsou hlavní organizace, které se zabývají otevřenými daty Fond Otakara Motejla a iniciativa opendata.cz. Tyto organizace poskytují konzultace ohledně otevřených dat a mohou dodat požadavky na konkrétní data, ale i doporučit technické řešení.

### 3.1.4 Soukromý sektor

Vzhledem k velkému počtu firem, které by mohly požadovat různá data, je velmi obtížné zjistit, co všechno by soukromý sektor mohl požadovat. Pro účely urychlení zprovoznění katalogu může být vhodné na začátku tuto část vynechat a zkoumat poptávku soukromého sektoru až po uvedení katalogu do provozu

na základě zpětné vazby. Za tímto účelem je vhodné zřídit speciální emailovou adresu, či možnost komentářů přímo v katalogu.

### 3.1.5 Analýza výstupů informačních systémů

Pokud má organizace k dispozici přehled IS, je velmi vhodné projít jednotlivé IS a podívat se na to, co obsahují za data, případně obejít jednotlivé správce IS a zeptat se na možnosti exportu dat. Vzhledem k tomu, že velmi často jsou data exportována právě z některého z IS, je toto velmi efektivní cesta, jak zjistit, co všechno organizace může zveřejňovat.

### 3.1.6 Komunikace napříč organizací

Dalším jednoduchým způsobem, jak lze také zjistit, co všechno organizace vůbec může veřejnosti poskytnout, je interní komunikace pomocí hromadných mailů, či vyvěšení na intranetu. Často mohou (což se během projektu otevírání MF potvrdilo) jednotliví tvůrci či vlastníci dat v organizaci dodat zajímavá data, jejichž zařazení do katalogu usnadní práci i jim.

## 3.2 IT řešení

Nástrojů, pomocí kterých lze provozovat katalog otevřených dat je více. Jedna z možností je vždy využití stránek, které organizace již provozuje, aby nevznikaly zbytečné náklady na vývoj dalšího systému. Je ovšem potřeba zvážit, jestli je tento postup optimální pro zveřejňování otevřených dat. Tím se rozumí například konvence URL entit, tj. aby měl každý datový soubor vlastní stabilní URL. Je také potřeba, aby katalog byl přehledný, aby bylo možné v něm vyhledávat, různě třídit atd.

Další možné řešení je využití již vytvořených opensourceových nástrojů. Využití takových nástrojů může být efektivní, účelné a hospodárné, protože tyto nástroje jsou většinou zdarma a jejich nasazení a nastavení není náročné. Důležité také je, že byly vytvořeny komunitou, která se otevřenými daty dlouhodobě zabývá a organizace si tak může být jistá, že použitím takového nástroje naplní požadavky otevřených dat. Jedná se například o katalogy DKAN a CKAN, které jsou volně k užití jako svobodný software (provozovatel neplatí licenční poplatky). Technický postup implementace katalogu otevřených dat na Ministerstvu financí je podrobněji popsán v příloze č.2 tohoto návodu.

## 3.3 Projektové řízení

Stejně jako u jiných projektů tohoto typu je potřeba mít jasně vymezené kompetence (v tomto případě to musí vymezovat zmíněný interní předpis) a určit analytický útvar, který bude schopen řídit celý projekt a převezme na sebe zodpovědnost za jeho dokončení. Analytický útvar musí potom řídit další rozvoj katalogu, jeho plnění a udržování.

### 3.3.1 IT řešení

Tvorba katalogu (volba nástroje a jeho finální podoby) by měla být v kompetenci IT oddělení, respektive správce katalogu. Je ovšem nutné, aby analytický útvar se správcem katalogu komunikoval již během vývoje a odsouhlasil jeho finální podobu a zajistil tak soulad s požadavky na otevřená data (viz kapitola 3.2 IT řešení).

### 3.3.2 Analýza datových sad a určení kurátorů

Způsob, jakým lze zjistit, co by bylo vhodné a co je možné v katalogu uveřejnit je popsán v kapitole 3.1 Obsah katalogu. Poté, co vznikne návrh seznamu datových sad, je potřeba zjistit, zda-li jsou data k dispozici a případně v jaké formě. To probíhá současně s určováním jednotlivých kurátorů datových sad. Většinou by z organizačního řádu mělo být zřejmé, který útvar za data odpovídá. Vedoucí útvaru by pak

měl potvrdit, že data má k dispozici a určit věcně příslušného zaměstnance (kurátora datové sady) nebo by měl potvrdit, že taková data nejsou k dispozici.

### 3.3.3 Komunikace s kurátory datové sady

V případě, že jsou navrhovaná data k dispozici, je nutné se s kurátorem datové sady sejit, vysvětlit mu pojem otevřená data a jejich smysl. Poté je potřeba si odsouhlasit finální podobu (strukturu) souboru, jeho periodicitu a také popis datové sady (katalogizační záznam).

U posuzování struktury datových sad je nutné přemýšlet nad tím, jaká struktura je nejvhodnější pro formát otevřených dat. V interním předpisu by mělo být uvedeno, jaké formáty jsou přípustné (minimálně to však musí být formáty, které lze otevřít v běžných tabulkových editorech), nicméně struktura (pokud to není XML formát nebo náročnější formát, jde pouze o to, jaké budou sloupce a řádky) je u každého souboru jiná a velmi často jsou data v podobě, která není pro uveřejňování nejvhodnější. Nežádoucí mohou být například sloučené buňky, zbytečné rozdělení do listů, více nesouvisejících dat v jednom listu, nevhodné formátování (formát data, čísel, textu atd.), nevhodně zvolené časové intervaly, názvy sloupců a řádků, z kterých není jasné, o co se jedná apod. Toto by měl posuzovat analytický útvar v součinnosti s kurátorem, nicméně v některých případech nemusí kurátor vnímat pojem otevřená data stejným způsobem a proto je potřeba, aby s ním toto analytický útvar konzultoval.

Dále je potřeba, aby analytický útvar kurátorovi vysvětlil, co přesně má obsahovat popis datové sady. Struktura popisu by měla být u všech datových sad stejná, což musí kontrolovat právě analytický útvar. Následující tabulka slouží jako vzor pro katalogizační záznam na Ministerstvu financí:

Položka	Popis
<b>Název datové sady</b>	Výstižné pojmenování datové sady.
<b>Popis</b>	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 nevyprávějí.
<b>Periodicita publikace datové sady</b>	Denně/Týdně/Měsíčně/Čtvrtletně/Pololetně/Ročně/Opakované s jinou periodicitou/Jednorázově.
<b>Aktualizace datové sady</b>	Datum, kdy naposledy došlo ke změně vlastních dat datové sady (např. k přidání datového souboru). Ve formátu: dd.mm.rrrr.
<b>Datum rozhodnutí o publikaci datové sady</b>	Datum, kdy ministr rozhodl o publikaci datové sady
<b>Poskytovatel dat</b>	Organizace poskytující otevřená data
<b>Kurátor</b>	Určený zaměstnanec příslušného útvaru, 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.
<b>Odpovědný útvar</b>	Číslo a název útvaru ministerstva, který datovou sadu vytvořil. Typický útvar ke kterému je příslušný kurátor.
<b>Související geografické území</b>	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.
<b>Klasifikace</b>	Každá datová sada je zařazena jedním nebo více identifikátory dle klasifikace EUROVOC ( <a href="http://eurovoc.europa.eu">http://eurovoc.europa.eu</a> ). 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)
<b>Podmínky užití</b>	Stanovení režimu užití datové sady odkazem na vhodnou otevřenou licenci nebo jako volné dílo.
<b>Schéma popisující strukturu datového souboru</b>	Odkaz na schéma tam, kde to bude mít smysl, tj. zejména u dat, která budou v XML. Pokud není schéma, bude kolonka prázdná.



### 3.3.4 Analýza možností zveřejňování dat

Po odsouhlasení finální podoby souboru s kurátorem datové sady je potřeba ověřit (a to by měl ověřovat kurátor datové sady v součinnosti s analytickým útvarem), že je dle zákona možné data takto zveřejnit a případně jestli je to vůbec vhodné. Buď jsou data již zveřejňována, nebo je zřejmé, že jejich zveřejnění neporušuje zákon, ale v některých případech (například zveřejňování faktur) je potřeba se společně s právním oddělením podívat na strukturu dat, jestli tam nejsou informace, které by neměly být veřejné (například jména zaměstnanců). Pokud právní oddělení dojde k rozhodnutí, že některé datová sada nemůže být zveřejněna, je vhodné na stránkách katalogu zmínit, že ji zveřejnit nelze a uvést důvod.

Základním kritériem pro rozhodování o možnostech zveřejňováním je zákon č. 106/1999 Sb., o svobodném přístupu k informacím (během roku 2015 by měla být přijata novela). Pokud organizace dané informace poskytuje nebo by v případě příslušného dotazu poskytovala na základě tohoto zákona, může je také zveřejnit v katalogu otevřených dat.

V některých případech může organizace mít k dispozici data, která nejsou její, ale mohla by je zveřejňovat. Typicky jde o export z IS, do kterého vkládají data jiné organizace, nebo export z IS, který je napojen na jiné IS a data pouze konsoliduje. Zde by také mělo rozhodnout právní oddělení, záleží totiž na konkrétních podmínkách, za kterých ostatní organizace data do IS vkládají nebo za kterých bylo domluveno propojení IS.

### 3.3.5 Komunikace mezi kurátory a správcem katalogu

Po odsouhlasení finální podoby datové sady, katalogizačního záznamu a potvrzení z právního oddělení, že je možné data zveřejnit, začíná kurátor datové sady komunikovat se správcem katalogu napřímo. Spolu si domluví způsob zasílání (mailem, na úložiště apod.) a termíny zasílání. Správce katalogu potom data spolu s popisem datové sady pouze nahraje do katalogu.

### 3.3.6 Řízení rizik

Stejně jako u každého projektu podobného typu, je nutné během implementace a po celou dobu provozování katalogu myslet na možná rizika a předcházet jim.

Jedná se zejména o následující rizika:

- 1) Zveřejnění důvěrných či tajných informací. Jak je zmíněno výše, je nutné vybrané datové sady posuzovat společně s právním oddělením. Toto riziko musí řídit především jednotliví kurátoři, nicméně analytický útvar by si měl u každé datové sady ověřit, že datová sada byla schválena k publikaci právním oddělením. Ministerstvo financí vycházelo při publikaci dat z předpokladu, že co zveřejnit lze, je veřejné a že, co by úřad musel poskytnout za informace na základě zákona 106/2000 Sb., může stejně tak zveřejnit rovnou v kompletní datové sadě.
- 2) Zveřejnění souborů, kde bude uvedeno jméno zaměstnance organizace. Každý soubor má svého autora, podle toho, kdo dokument založil. Je potřeba vysvětlit jednotlivým kurátorům, aby si na toto dávali pozor a autory smazali, případně přepsali na jméno organizace.
- 3) DDoS útok a následné vyřazení serveru z provozu. Jedná se o záměrné přetížení serveru na nějakou dobu a těžko se proti tomu dá bránit. Pokud se takový útok nastane, může to znamenat, že někdo nechce, aby ta data byla zveřejněna, což se dá považovat v podstatě za úspěch otevřených dat.
- 4) Nabourání katalogu a změna dat. V tomto případě je povinnost správce katalogu, aby otázku bezpečnosti měl vyřešenou.

- 5) Porucha katalogu. Pro tyto případy je nutné mít duplicitní kopii a zálohu.

## 4 Údržba a rozvoj

Po zprovoznění a zveřejnění katalogu je potřeba, aby analytický útvar službu nadále vylepšoval jak z pohledu obsahu, tak z pohledu funkcionalit katalogu. Jak je zmíněno v kapitole 3.1.4 Soukromý sektor, měla by být zřízena speciální emailová adresa na reakce veřejnosti. Na základě těchto reakcí je pak možné katalog upravovat, doplňovat či opravovat chyby.

## 5 Popis technické implementace

### 5.1 Příprava

#### 5.1.1 Požadavky na katalog otevřených dat

Finální podoba technické specifikace katalogu otevřených dat uvedená v příloze interní směrnice č.6/2014 ministra financí – Publikace a katalogizace otevřených dat Ministerstva financí (dále jen Směrnice) reflektuje požadavky veřejnosti, akademické obce a komunity otevřených dat. Zásadní jsou následující požadavky:

- a) data jsou dohledatelná běžnými informačními a komunikačními technologiemi
- b) data jsou poskytována ve formě ucelených souborů
- c) data jsou zveřejněny nejméně v jednom z následujících formátů: XLSX, CSV, JSON, XML,
- d) datové soubory jsou seskupeny do datových sad,
- e) ke každé datové sadě je zveřejněný katalogizační záznam, který obsahuje metadata dle Směrnice
- f) katalog dovoluje definovat uživatelská práva pro role dle Směrnice (kurátor, analytický útvar, správce katalogu otevřených dat),
- g) katalog bude veřejný na adrese <http://data.mfcr.cz/>
- h) jednotná konvence URL datových sad

Další požadavky zahrnovaly rychlou a levnou implementaci, snadnou možnost úprav, rozšiřování funkcionalit katalogu a rozhraní API pro možnost strojového napojení na katalog.

#### 5.1.2 Výběr technického řešení

Technické řešení jsme vybírali na základě výše zmíněných požadavků. Vzhledem k potřebě rychlé a levné implementace a dodržování principů 3E (economy, efficiency and effectivity) jsme ihned vyloučili vlastní vývoj softwaru či vývoj softwaru na zakázku od externího dodavatele. Z hotových řešení naše požadavky splňovaly systém CKAN<sup>2</sup>, který je vyvíjen společností Open Knowledge Foundation jako otevřený software a je ve světě standardem pro zveřejňování otevřených dat i pro soukromé společnosti i pro samosprávu. Dalším řešením, které splňovalo naše požadavky byl systém DKAN<sup>3</sup>, který je vyvíjen společností NüCivic jako klon systému CKAN pro použití na platformách LAMP (tj. serverech s Linux, Apache, MySQL a PHP).

---

<sup>2</sup> Více informací na <http://ckan.org/>

<sup>3</sup> Více informací na <http://getdkan.com/> a <http://docs.getdkan.com>

Vzhledem k jednoduchosti instalace a možnostem rozšiřitelnosti (je postavený na celosvětově rozšířeném systému na správu webového obsahu Drupal<sup>4</sup>) jsme zvolili DKAN.

### 5.1.3 Test proveditelnosti

Pro potvrzení výběru jsme provedli testovací instalaci katalogu DKAN na bezplatném serveru společnosti Acquia. Ověřili jsme funkce a strukturu katalogu pro potřeby Ministerstva financí a možnosti nastavení a přizpůsobení celého portálu. Po odzkoušení jsme se přesunuli k instalaci na lokální server.

## 5.2 Instalace

Doporučená konfigurace pro instalaci katalogu DKAN zahrnuje Linuxové prostředí s webovým serverem Apache a databázovým systémem MySQL. V našem případě nejsou servery s operačním systémem Linux ve standardní výbavě ministerstva a tak jsme se rozhodli pro instalaci na virtuálním Windows serveru. Při instalaci jsme postupovali dle instrukcí z dokumentace na internetových stránkách [docs.getdkan.org](https://docs.getdkan.org). Pouze v závěrečné fázi instalace jsme museli vyřešit problém s právy na zápis do některých adresářů, nejspíš způsobený pro instalátor neočekávaným Windows prostředím.

Instalace na interním virtuálním serveru nám přinesla spoustu výhod, zejména jednoduché uložení aktuální verze webu („klonování“) pro případ nutnosti vrácení změn. Na druhou stranu při dalších úpravách a instalaci doplňujících balíčků jsme postrádali připojení serveru k internetu a tak jsme balíčky museli nahrávat manuálně.

## 5.3 Úpravy

### 5.3.1 Lokalizace

Moduly DKAN, ani výchozí téma katalogu nejsou na webu dostupné v českém jazyce. Museli jsme tedy naprostou většinu webu vlastními silami přeložit. K tomu jsme využili následující moduly Drupalu:

- Locale<sup>5</sup>
- Localization Update<sup>6</sup>
- Localization Client<sup>7</sup>

Modul Localization client dovoluje překlad přímo na stránce (kliknutím na tlačítko Translate text v pravém dolním rohu) a tedy je po uživatelské stránce velmi přívětivý. Největším problémem lokalizace byla specifika českého jazyka. Jako příklad může posloužit anglická fráze „Filter by“, která se překládá jednotně pro všechny filtry na stránce s vyhledáváním datových sad a tedy při přeložení vzniká problém s tvary (např. překlad „Třídít dle“ dělá problém při použití „Třídít dle štítky“) a proto bylo nutné vymýšlet alternativy. V současné době pracujeme na zaslání námi vytvořených překladů do centrálního úložiště překladů na stránkách Drupalu, aby byly k použití pro všechny budoucí uživatele DKAN.

### 5.3.2 Struktura datové sady

Struktura metadat datové sady ve výchozím nastavení DKAN je odlišná od struktury katalogizačního záznamu dle Směrnice. Změna polí a jejich formátů probíhá v administraci systému Drupal<sup>8</sup> a je

---

<sup>4</sup> Více informací na <https://www.drupal.org/>

<sup>5</sup> Ve výchozí instalaci Drupal 7

<sup>6</sup> Ke stažení zde: [https://www.drupal.org/project/l10n\\_client](https://www.drupal.org/project/l10n_client)

<sup>7</sup> Ke stažení zde [https://www.drupal.org/project/l10n\\_update](https://www.drupal.org/project/l10n_update)

doporučené změny provést před naplněním katalogu jednotlivými datovými sadami. Zároveň jsou některá pole na pevně zakódovaná ve výchozích šablonách, a proto jejich odebrání způsobuje problémy na stránce. Např. se ve výchozím nastavení šablony zobrazuje na stránce datové sady blok s licenci, který používá pole *field\_license* a při jeho smazání hlásí chybu.

Pole pro datové sady (příp. i datové soubory a skupiny dat) je potřeba nastavit jak v záložce Správa polí, tak v záložce Správa zobrazení. První ovlivňuje jejich vykreslení ve formuláři pro vytvoření/editaci datové sady, druhé pak zobrazení pro veřejnost.

Změnu polí u datové sady je potřeba reflektovat i v nastavení indexování obsahu<sup>9</sup>. Je tedy potřeba vytvořit indexy na nově vytvořená pole, případně zapnout možnost třídění těchto polí v záložce Facets a následně přidat nové bloky pro třídění na stránku s vyhledáváním datových sad<sup>10</sup>.

### 5.3.3 Vzhled

Vzhledem k časové tísní jsme grafický vzhled webových stránek oproti výchozímu tématu DKANu NuBoot měnili minimálně. Nastavení tématu dovoluje komfortní změnu obrázku na úvodní stránce, loga a ikony stránky. Zbytek je nutné upravit v kódu tématu v adresáři DKANu.

Z hlediska struktury jsme využívali vestavěné správy bloků a kontextů, která je popsána v dokumentaci Drupalu<sup>11</sup>. Obsah bloků lze upravovat v jazyce HTML.

### 5.3.4 Analýzy návštěvnosti

Kromě statistik z webového serveru jsme do webových stránek implementovali bezplatné analýzy chování a technického vybavení návštěvníků Google Analytics. Implementace byla opět provedena instalací modulu do systému Drupal. Statistiky zobrazují kromě aktuálního a historického počtu návštěvníků i vstupní a výstupní stránky, nejčastěji navštěvované stránky, země původu návštěvníků, zdroje návštěvnosti, technické parametry zařízení a další. Podmínkou je vytvoření účtu Google.

### 5.3.5 Optimalizace výkonu

V rámci testování výkonu jsme použili Apache HTTP server benchmarking tool<sup>12</sup>, nástroj který simuluje zátěž na serveru zasíláním nastaveného počtu žádostí. Po zapnutí výchozího kešování v administraci Drupal byly výsledky velmi dobré.

## 5.4 Publikace

### 5.4.1 Spuštění

Pro ostrou verzi katalogu jsme vytvořili další server, tentokrát s přístupem na internet. Nastavili jsme přesměrování domény, ale zablokovali přístup na server z jiných adres, než Ministerstva financí. Přenesení katalogu DKAN proběhlo odklonováním verze z vývojového serveru. Po otestování všech funkcí a provedení posledních drobných úprav jsme v předvečer tiskové konference zrušili omezení na přístup a

---

<sup>8</sup> Administrační menu -> Struktura -> Typy obsahu -> Datová sada (Dataset)

<sup>9</sup> Administrační menu -> Nastavení -> Vyhledávání -> Search API -> Rejstřík Datasets

<sup>10</sup> Bloky se upravují v: Administrační menu -> Struktura -> Kontext -> Upravit záznam dataset\_search -> Reactions -> Bloky

<sup>11</sup> Zde: <https://www.drupal.org/documentation/blocks> a zde: <https://www.drupal.org/project/context>

<sup>12</sup> Více informací a stažení: <http://httpd.apache.org/docs/2.2/programs/ab.html>

ještě jednou otestovali funkčnost z internetu. Do tiskové konference jsme nechali server veřejný, nicméně dle statistik na něj nebylo přistupováno jinými uživateli než námi.

#### **5.4.2 Tisková konference**

Na serveru bylo alokováno pro jistotu více výpočetního výkonu, nicméně ani to nestačilo a tak již během tiskové konference server neustál nápor uživatelů a nebyl schopen vypořádat žádosti uživatelů v povoleném čase. Většina uživatelů se tedy k webovým stránkám po několik hodin nedostala. Poté již webový server fungoval standardně bez jakýchkoliv náznaků nestability.