

**OTEVŘENÁ DATA**

# Standardy publikace a katalogizace otevřených dat VS ČR

Ministerstvo vnitra České republiky

2015

Projekt: „Implementace strategií v oblasti otevřených dat veřejné správy ČR“  
CZ.1.04/4.1.00/D9.00001

# 1 Obsah

2	Úvod .....	5
3	Co jsou otevřená data? .....	6
3.1	Co nelze chápat jako otevřená data? .....	6
3.2	Jaké jsou přínosy otevřených dat? .....	7
3.3	Mýty a omyly při otevírání dat .....	7
4	Stupně otevřenosti datových sad .....	9
4.1.1	Specifikace stupňů otevřenosti datových sad .....	10
4.1.2	Stupně otevřenosti v kontextu veřejné správy České republiky .....	12
5	Důležité pojmy v oblasti otevřených dat .....	12
6	Role a jejich činnosti .....	14
6.1	Vedení poskytovatele dat .....	14
6.2	Koordinátor otevírání dat .....	14
6.3	Kurátor dat .....	14
6.4	Správce katalogu otevřených dat .....	14
6.5	IT specialista .....	15
7	Postup otevírání dat .....	15
8	Vytvoření publikačního plánu .....	15
8.1	Rozhodnutí o otevírání dat .....	16
8.1.1	Přínosy pro veřejnost .....	17
8.1.2	Přínosy pro ekonomiku .....	17
8.1.3	Přínosy pro poskytovatele dat .....	17
8.2	Obsazení role Koordinátora otevírání dat .....	18
8.3	Příprava interní legislativy .....	18
8.4	Schválení interní legislativy .....	18
8.5	Obsazení dalších rolí v kontextu otevřených dat .....	19
8.5.1	Kurátor dat .....	19
8.5.2	Správce katalogu otevřených dat .....	19
8.5.3	IT specialista .....	20
8.6	Příprava publikačního plánu .....	20
8.6.1	Vzorové podklady .....	21
8.7	Vytvoření seznamu datových sad .....	21
8.8	Určení přínosů otevření datových sad .....	21
8.9	Určení rizik otevření datových sad .....	22
8.10	Revize seznamu datových sad určených k otevření .....	23
8.11	Stanovení podmínek užití datových sad .....	23
8.11.1	Sekce 1 – Autorská práva .....	24
8.11.2	Sekce 2 – Autorskoprávní ochrana databáze jako díla .....	25
8.11.3	Sekce 3 - zvláštní databázová práva pořizovatele databáze .....	25

8.11.4	Závěr .....	26
8.11.5	Přehledová tabulka .....	26
8.12	Sestavení harmonogramu publikace datových sad .....	26
8.13	Návrh datových sad ke zveřejnění .....	27
8.14	Spolupráce na návrhu datových sad k publikaci.....	28
8.15	Schválení publikačního plánu .....	28
8.16	Rozhodnutí o způsobu katalogizace otevřených dat .....	28
8.16.1	Modelová situace 1: Malá obec s jednoduchou statickou webovou stránkou bez možnosti jakýchkoliv úprav.....	29
8.16.2	Modelová situace 2: Ministerstvo .....	30
8.17	Konzultace s IT specialistou pro určení způsobu katalogizace .....	30
8.18	Spolupráce IT specialisty na určení způsobu katalogizace datových sad.....	30
8.19	Zveřejnění publikačního plánu .....	30
8.20	Periodické vyhodnocování plnění publikačního plánu a informování vedení .....	31
8.21	Vedení statistiky přístupů k otevřeným datovým sadám .....	31
9	Publikace datové sady v otevřené podobě .....	31
9.1	Analýza datové sady a návrh technického způsobu jejího zveřejnění .....	32
9.2	Určení stupně otevřenosti datové sady .....	33
9.3	Vytvoření datového schématu datové sady.....	33
9.3.1	Datová schémata doporučených datových sad .....	33
9.3.2	Vytvoření nového datového schématu .....	34
9.4	Příprava katalogizačního záznamu datové sady .....	34
9.4.1	Katalogizační záznamy doporučených datových sad .....	34
9.4.2	Vyplňování hodnot atributů katalogizačního záznamu.....	34
9.5	Vyhledání EUROVOC konceptu.....	37
9.6	Výběr MIME typu .....	38
9.7	Spolupráce na analýze datové sady .....	38
9.8	Příprava datové sady k publikaci.....	39
9.9	Převod dat ze stupně otevřenosti 1 do vyššího stupně otevřenosti .....	40
9.10	Převod dat ze stupně otevřenosti 2 do vyššího stupně otevřenosti .....	40
9.10.1	Převod tabulkové datové sady .....	40
9.10.2	Převod datové sady s hierarchickou strukturou.....	41
9.10.3	Převod datové sady v podobě kolekce textových dokumentů .....	41
9.10.4	Převod ostatních typů datových sad .....	41
9.11	Rozhodnutí o podobě aktualizací datové sady .....	41
9.12	Konzultace přípravy datové sady s IT specialistou .....	42
9.13	Validace datového souboru.....	43
9.14	Spolupráce na přípravě datové sady k publikaci .....	43
9.15	Kontrola datové sady připravené k publikaci .....	43
9.16	Zajištění zveřejnění datové sady.....	44

9.17	Zveřejnění datové sady .....	44
9.18	Vytvoření a správa katalogizačního záznamu v lokálním katalogu .....	44
9.19	Vytvoření a správa katalogizačního záznamu v Národním katalogu otevřených dat (NKOD) .....	44
9.19.1	Katalogizace nové datové sady .....	45
9.19.2	Úprava katalogizačního záznamu .....	45
9.19.3	Zrušení katalogizačního záznamu.....	45
10	Zprovoznění lokálního katalogu otevřených dat a jeho registrace v NKOD.....	46
10.1	Instalace lokálního katalogu.....	46
10.2	Příprava lokálního katalogu.....	47
10.3	Provoz lokálního katalogu .....	47
10.4	Registrace vlastního katalogu v NKOD .....	47
10.4.1	Registrace lokálního katalogu .....	47
10.4.2	Zrušení registrace lokálního katalogu .....	48
11	Nástroje pro publikaci otevřených dat.....	49
11.1	Národní katalog otevřených dat (NKOD).....	49
11.2	CKAN a jeho rozšíření .....	50
11.2.1	Instalace .....	50
11.3	DKAN a jeho rozšíření .....	50
11.4	Použití widgetu NKOD .....	51
11.5	Použití NKOD API.....	51
11.6	Služba Datahub.io.....	51
12	Nástroje pro publikaci otevřených prostorových dat.....	52
13	Technické standardy .....	52
13.1	Sdílené datové typy .....	52
13.1.1	Datový typ: AnoNe .....	52
13.1.2	Datový typ: Čas .....	52
13.1.3	Datový typ: Celé číslo .....	52
13.1.4	Datový typ: Datum .....	52
13.1.5	Datový typ: Desetinné číslo.....	52
13.1.6	Datový typ: Řetězec.....	53
13.1.7	Datový typ: URL.....	53
13.2	Technické standardy pro datové sady na stupni otevřenosti 3 .....	53
13.2.1	Standardy pro tabulkové datové sady .....	53
13.2.2	Standardy pro datové sady s hierarchickou strukturou.....	54
13.2.3	Standardy pro datové sady v podobě kolekce textových dokumentů .....	54
13.2.4	Standardy pro datové sady v podobě kolekce obrázků .....	54
13.2.5	Standardy pro prostorové datové sady.....	55
13.2.6	Standardy pro ostatní typy datových sad .....	57

13.3	Technické standardy vytváření datových schémat pro datové sady na stupni otevřenosti 3.....	57
13.4	Technické standardy pro datové sady na stupni otevřenosti 4 a 5 .....	58
13.4.1	Datový model RDF.....	58
13.4.2	Standardy pro zápis dat v datovém modelu RDF .....	58
13.4.3	Standardy tvorby IRI .....	59
13.4.4	Standardy propojených dat .....	59
13.5	Technické standardy vytváření datových schémat pro datové sady na stupni otevřenosti 5.....	60
13.5.1	Standardy pro vyjádření a zápis ontologií .....	60
13.5.2	Standardy pro tvorbu ontologií .....	60
13.5.3	Standardy propojování prvků ontologií.....	61
13.5.4	Standardy tvorby IRI tříd a vlastností definovaných v ontologiích.....	61
13.6	Technické standardy pro aplikační rozhraní lokálních katalogů.....	62
13.6.1	package_list.....	62
13.6.2	package_show.....	62
15	Přílohy.....	66
15.1	Vzor směrnice "Publikace a katalogizace otevřených dat" .....	66

## 2 Úvod

Tento dokument je určen orgánům veřejné správy, které plánují otevřít svá data. Dokument detailně popisuje doporučený postup otevírání dat včetně obsazení rolí a jejich povinností v kontextu otevřených dat.

Dokument dále popisuje jednotlivé nástroje a jejich možnosti pro katalogizaci otevřených dat včetně technických standardů pro klíčové stupně otevřenosti otevřených dat, standardy vytváření datových schémat a technické standardy pro rozhraní lokálních katalogů.

Pro potřeby integrace procesu otevírání dat a zajištění interní legislativy, byla vytvořena vzorová interní směrnice, která je součástí tohoto dokumentu.

Standardy publikace a katalogizace otevřených dat VS ČR vznikly jako výstup projektu „Implementace strategií v oblasti otevřených dat veřejné správy ČR“ CZ.1.04/4.1.00/D9.00001. Projekt je financován v rámci Operačního programu lidské zdroje a zaměstnanost z Evropského sociálního fondu a státního rozpočtu ČR. Realizátorem projektu je Ministerstvo vnitra ČR.



## 3 Co jsou otevřená data?

Nezisková organizace [Open Knowledge definuje](#) otevřená data jako data zveřejněná na internetu způsobem, který neomezuje žádné uživatele ve způsobu jejich použití pro nekomerční i komerční účely (technicky ani legislativně) a opravňuje všechny uživatele k jejich dalšímu šíření, pokud tímto šířením nedojde k omezení práv ostatních uživatelů.

**Orgány veřejné správy by měly publikovat jako otevřená data takové datové sady, které jsou významné pro dosažení hospodářského prospěchu, zlepšení kvality života nebo zvýšení efektivity a transparentnosti veřejné správy.**

[Metodika publikace otevřených dat veřejné správy ČR](#), která byla vytvořena v roce 2012 v rámci plnění Akčního plánu Partnerství pro otevřené vládnutí, schváleném v [usnesení Vlády České republiky ze dne 4. dubna 2012 č. 243](#), definuje otevřená data jako data poskytovaná způsobem, který umožňuje dálkový přístup a která jsou:

- *úplná*, tj. jsou zveřejněna v maximálním možném rozsahu, tj. nejsou záměrně odstraňovány vybrané záznamy a data jsou udržována co možná nejaktuálnější,
- *snadno dostupná*, tj. jsou dostupná na internetu a dohledatelná běžnými ICT nástroji a prostředky,
- *strojově čitelná*, tj. jsou ve formátu, který je strukturovaný takovým způsobem, že pomocí programové aplikace z nich lze získat žádané (vybrané) údaje,
- *používající standardy s volně dostupnou specifikací (tzv. otevřené standardy)*, tj. musí být ve formátu, který je volně (bezplatně) dostupný pro libovolné použití nebo do takového formátu převoditelný volně (bezplatně) dostupnou aplikací,
- *zpřístupněna za jasně definovaných podmínek užití dat (licence) s minimem omezení*, tj. podmínky jejich užití musí být jasně a zřetelně definovány a zveřejněny a musí umožňovat využití dat pro komerční i nekomerční účely,
- *dostupná uživatelům při vynaložení minima možných nákladů na jejich získání*, tj. jejich poskytovatelé jsou oprávněni žádat úhradu maximálně ve výši, která nesmí přesáhnout náklady spojené s jejich zpřístupněním uživateli; poskytovatel dat může jednorázově vyžádat i úhradu za mimořádně náročné pořízení dat, pokud si uživatel zpřístupnění těchto dat vyžádá. Zpoplatnění je však předpokládáno ve zcela výjimečných případech.

### 3.1 Co nelze chápat jako otevřená data?

Otevřená data jsou určena pro specialisty (programátory, datové analytiku, novináře, vědce atd.), kteří s jejich pomocí tvoří zajímavé softwarové aplikace či odborné analýzy. Ty pak slouží široké veřejnosti. Proto je důležité, aby bylo možné co nejjednodušším způsobem s daty pracovat - vytvářet nad nimi aplikace, analyticky je zpracovávat pomocí běžně používaných softwarových nástrojů apod.

Proto nelze jako otevřená data chápat např.:

- tabulky zveřejněné v podobě PDF souborů,
- tabulky zveřejněné v podobě HTML stránky,
- vyhledávací HTML formulář určený pro koncové uživatele.

Všechny výše uvedené způsoby lze sice chápat jako způsoby zveřejňování informací, jsou ale zcela nevhodné pro následné využití dat v aplikacích, pro analýzy apod. Nejedná se o otevřená data. Některé orgány veřejné správy však právě takto zveřejněné informace zcela chybně prezentují jako otevřená data.

Aby bylo možné potenciál otevřených dat využít co nejvíce, je nutné zveřejňovat data z různých oblastí veřejné správy. Často ale pod pojmem otevřená data různé orgány veřejné správy chápou pouze zveřejňování smluv, objednávek a faktur. Ty jsou jen malým výsekem zajímavých dat, která lze zveřejňovat v otevřené podobě.

## 3.2 Jaké jsou přínosy otevřených dat?

Otevřená data jsou především šetrným způsobem jak poskytovat informace. Vytváření PDF souborů, webových stránek či aplikací s sebou totiž nese dodatečné náklady, které v případě publikace dat v otevřené podobě odpadají. Principy otevřených dat požadují pouze zveřejnění dat v surové podobě, nikoliv dodatečnou tvorbu PDF souborů, webových stránek či aplikací.

- Otevřená data mohou být opakovaně využívána v různých mobilních a jiných aplikacích, které slouží veřejnosti. Veřejná správa s tvorbou takových aplikací nemá náklady, ty přebírá odborná veřejnost.
- Veřejná správa může veřejnosti ukázat, na čem všem pracuje a jakou hodnotnou službu poskytuje.
- Občané mohou snadno využívat otevřená data veřejné správy v běžném životě.
- Zlevní a urychlí se procesy sdílení dat v rámci veřejné správy.
- Veřejnost může lépe dohlížet na fungování veřejné správy.
- Občané se mohou kvalifikovaně podílet na fungování státu.
- Novináři mají relevantní zdroje informací a mohou tak psát kvalitní a na faktech založené články.

## 3.3 Mýty a omyly při otevírání dat

### 1. Naše IT systémy neumí vytvářet otevřená data.

Pokud umíte z vašeho informačního systému exportovat tiskovou sestavu v podobě formátu XLS (Excel) či jiného podobného formátu, umíte exportovat i data v otevřené podobě. Pokud máte přístup k databázi, ve které informační systém svá data ukládá, lze snadno napsat SQL dotaz, který data v otevřené podobě vyexportuje.

### 2. O data nikdo nemá zájem, veřejnosti jsou k ničemu.

Záleží na úhlu pohledu, kdo se na data dívá a kdo s nimi pracuje. Vám mohou připadat nezajímavá pro zveřejnění, ale veřejnost, studenti, firmy ale i akademická sféra může data využít dále a ve spojení s dalšími daty mohou vznikat nové služby, inovace a projekty. Příklady ze zahraničí ale i z Česka nás o tom neustále přesvědčují. Podívejte se na [příklady](#) ze zahraničí nebo [aplikace](#) ze soutěže Společně otevíráme data.

### 3. Je to příliš složité pro naši obec.

Není, zvládají to velké orgány veřejné správy i malé obce. Jde jen o to, jaký rozsah a způsob publikace zvolíte. Nemusíte otevřít všechno. Vyberte si datové sady, o kterých si myslíte, že o ně bude největší zájem a postupně přidávejte další. Všechny kroky komunikujte ven a získejte tak zpětnou vazbu na vaše data.

### 4. Data jsou majetkem orgánu veřejné správy, nelze je zveřejnit v otevřené podobě pro komerční užití.

Data jsme si jako daňoví poplatníci již jednou zaplatili na daních. Není tedy důvod, aby s nimi mohl nakládat pouze jeden orgán veřejné správy, který navíc nemá kapacity na jejich další využití a rozpracování. Když data zveřejníte, umožníte, aby ve veřejném sektoru vznikaly nové a zajímavé služby založené na vašich datech. Například NASA v nedávné době prodávala snímky z Hubbleova teleskopu, než soud řekl, že NASA je placena z veřejných peněz, a tak i její výstupy mají povahu veřejného majetku.



## **5. Máme webové stránky umožňující prohlížet smlouvy, objednávky a faktury - máme tedy otevřená data.**

Nemáte, ale jste na dobré cestě. Naskenované PDF smlouvy bez strojově čitelných údajů o smluvních stranách, částce apod. a bez správných podmínek užití nejsou otevřená data.

## **6. Když na web dáme data, tak je někdo může nevhodně interpretovat a zneužít, špatně je vyloží a může tím způsobit škodu a ta bude vymáhána na nás?**

Pokud k datům vytvoříte metadata a vhodnou dokumentaci, v níž vysvětlíte, co znamenají a co je korektní s nimi dělat, nemusíte se bát – ani žaloby za způsobené škody, ani ztráty reputace. Podle dosavadních zkušeností k takovým dezinterpretacím prakticky nedochází.

## **7. Naše data jsou ve velmi špatné kvalitě, takže je nemůžeme a ani nechceme publikovat.**

Pokud jsou data v tak špatné kvalitě, že nelze publikovat ani jejich část, tak k čemu je vlastně máte, a jak jste schopni podle nich plánovat a řídit váš orgán veřejné správy? Dejte alespoň vědět, že data máte. Pokud uživatel neví, co hledá, nežádá to. Otevření dat vám pomůže analyzovat vaše data a zjistit jejich aktuální stav. Následně zjistíte, že některá data můžete zveřejnit rovnou, jiná budou potřebovat ještě vyčistit a upravit.

## **8. Některé datové sady nelze publikovat, neboť se jedná o složité a veliké soubory, které nejsou v technických možnostech orgánu veřejné správy.**

To může být skutečný problém, který je ale snadno řešitelný. Data nemusíte mít na svém serveru, ale můžete využít cloudových řešení a na svá data budete jen odkazovat ze svého lokálního katalogu nebo využijete rovnou Národního katalogu otevřených dat. Více informací najdete [zde](#).

## **9. Nemáme na to kapacity a otevírání dat nás zahltlí.**

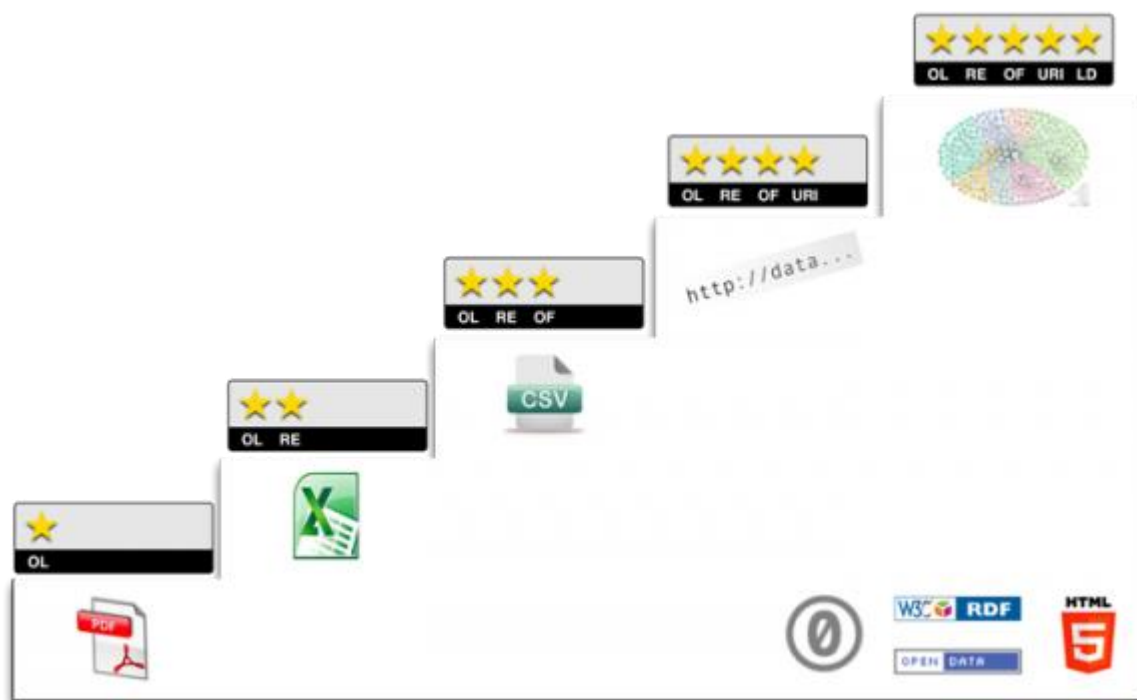
Záleží na tom, jak si nastavíte procesy a priority. Při správném plánování je zátěž nízká. Nemusíte otevírat všechna data. Vyberte si ze [seznamu datových sad](#) ty, o kterých si myslíte, že o ně bude největší zájem a nebude s nimi tolik práce při jejich publikaci, a ty publikujte. Zjistěte si zpětnou vazbu a postupně přidávejte další datové sady.

## **10. Naším právníkům se to nebude líbit. Budou chtít vlastní licenci a kontrolu všech datových sad.**

Ano, určení podmínek užití dat je velmi důležité, ale je řešitelné za pomoci našeho [postupu určení podmínek užití](#).

## 4 Stupně otevřenosti datových sad

Míru otevřenosti dat je možné vyjádřit pomocí 5 stupňů otevřenosti, které jsou znázorněny na následujícím obrázku (stupeň je vyjádřen počtem hvězdiček).



Převzato z <http://5stardata.info>

- stupeň 1 - datová je dostupná v síti WWW s vhodnými podmínkami užití otevřených dat (viz kapitola: „Stanovení podmínek užití datových sad“),
- stupeň 2 - datová sada je poskytována ve strojově čitelném formátu, který umožňuje automatizované strojové zpracování,
- stupeň 3 - datová sada je poskytována v otevřeném formátu, tj. ve formátu s volně dostupnou specifikací,
- stupeň 4 - na identifikaci entit v datové sadě se používají IRI,
- stupeň 5 - datová sada splňuje standard propojených dat.

Datová sada je tvořena sadou spolu souvisejících údajů. Při jejím zveřejňování v síti WWW je typicky distribuována zájemcům v podobě souborů ke stažení, které nazýváme **distribuce datové sady**. Distribucí datové sady může být více. Pokud je obsah datové sady příliš velký, je možné jej rozdělit do několika distribucí. Nebo může mít smysl poskytnout obsah datové sady ve více formátech. Je tedy např. možné poskytnout distribuci datové sady ve formátu XLS a jinou distribuci ve formátu CSV. První formát je vhodný pro analytiku, kteří chtějí s údaji pracovat ve svém tabulkovém editoru. Druhý formát je vhodný pro programátory, kteří chtějí údaje automatizovaně převádět do svojí databáze, nad kterou budují svojí softwarovou aplikaci.

Vždy platí, že se snažíme poskytnout jednu distribuci s celým obsahem datové sady a v pravidelných periodách ji nahrazovat novou distribucí vždy s aktuálním obsahem. Je to nejvhodnější způsob publikace pro uživatele, kteří potřebují pracovat s obsahem datové sady jako s jedním celkem. Není vhodné např. pro každý záznam v datové sadě poskytovat jednu distribuci, kterou lze stáhnout jedině tak, že uživatel zná identifikátor záznamu a ten použije jako hodnotu parametru v odkazu na distribuci. Takový přístup je příliš komplikovaný a získání celého obsahu datové sady je velmi složité a neefektivní jak pro uživatele tak poskytovatele dat. Pouze u velkých datových sad je možné rozdělit jejich obsah do více distribucí z důvodů jejich neúměrné velikosti. Případně u datových sad, kde je důležité

přesně informovat o provedených změnách je doporučováno zveřejnit jednu distribuci s iniciálním obsahem datové sady a poté zveřejňovat distribuce se seznamem provedených změn (tj. jaké záznamy byly smazány a jaké byly vytvořeny či aktualizovány a jak).

### 4.1.1 Specifikace stupňů otevřenosti datových sad

#### *Stupeň otevřenosti 1*

Stupeň otevřenosti 1 vyžaduje, aby byly distribuce datové sady dostupné on-line a s jasným vymezením podmínek užití. Neklade však žádné požadavky na datové formáty, ve kterých jsou distribuce datových sad zveřejňovány. Proto tento stupeň není považován za dostatečný stupeň otevřenosti.

V případě publikace prostorových dat sem spadají mimo jiné i často používané webové služby [OGC WMS](#) (Web Mapping Service) a [OGC WMTS](#) (Web Mapping Tile Service). Služby nezpřístupňují vlastní data, ale pouze obrázky generované z těchto dat - ty tedy v žádném případě nelze považovat za strojově zpracovatelné (viz stupeň 2).

#### *Stupeň otevřenosti 2*

Stupeň otevřenosti 2 je charakteristický tím, že distribuce datové sady jsou poskytovány ve strojově čitelném formátu. Pojem strojově čitelnosti je v oblasti IS dobře známý. Distribuce datové sady je vždy vytvářena za účelem zaznamenání určité množiny údajů. Strojová čitelnost distribuce datové sady pak znamená, že je distribuce vyjádřena v takovém formátu, který umožňuje co nejsnazší přístup k jednotlivým zaznamenaným údajům pomocí běžných a volně dostupných programovacích prostředků (programovacích jazyků a knihoven).

Příkladem je datová sada, která má charakter tabulky. Distribuce musí být vyjádřena v takovém formátu, který umožňuje pomocí běžných programovacích prostředků získávat z tabulky jednotlivé řádky a obsah buněk v řádcích. Jiným příkladem je datová sada tvořená kolekcí textových dokumentů. Musí být zaznamenána ve formátu, který umožňuje pomocí běžných programovacích prostředků získávat jednotlivé dokumenty, jejich jednotlivé znaky a značky označující sémantiku jejich částí jako jsou nadpisy (různých úrovní), tabulky, seznamy, obrázky apod.

Co nejsnazší přístup k údajům znamená, že údaje reprezentované v distribuci datové sady je možné přímo číst pomocí programovacích instrukcí bez nutnosti jakéhokoliv předzpracování distribuce, které nesouvisí s čtením jednotlivých údajů. Předzpracováním distribuce, která nesouvisí se čtením jednotlivých údajů, může být např. nutnost:

- zpracovat distribuci pomocí OCR nástrojů:
  - např. kolekci dokumentů v podobě naskenovaných stránek není možné považovat za distribuci datové sady tvořenou kolekcí dokumentů na stupni otevřenosti 2,
  - podobně např. tabulku reprezentovanou v podobě obrázku není možné považovat za distribuci datové sady s tabulkou na stupni otevřenosti 2 (ať se jedná o samostatný obrázek nebo je obrázek vložen do jiného souboru, např. do XLS či DOC souboru);
- vyhledat tabulku zanořenou v textu či mezi jinými objekty:
  - např. dokument ve formátu HTML obsahující tabulku vyjádřenou v podobě HTML elementu <table> vnořenou v textu a mezi jinými HTML elementy nelze považovat za distribuci datové sady s tabulkou na stupni otevřenosti 2;
- vyhledat tabulku mezi jinými tabulkami:
  - např. tabulku v souboru ve formátu XLS či XLSX, který obsahuje více tabulek nelze považovat za distribuci datové sady s tabulkou na stupni otevřenosti 2,
- rekonstruovat obsah tabulky ze struktur, které nejsou primárně určeny pro reprezentaci tabulkových dat, ale pro vyjádření vizuálních objektů ve tvaru tabulek,

- o např. dokument ve formátu DOC či DOCX obsahující tabulku nelze považovat za distribuci datové sady s tabulkou na stupni otevřenosti 2.

Je-li to vhodné a účelné, lze distribuce na stupni otevřenosti 2 poskytovat v komprimovaném tvaru (např. jako ZIP archiv apod.). To připadá v úvahu především v případě velkých distribucí nebo v případě kolekcí textových dokumentů.

Protože stupeň otevřenosti 2 neklade žádné další požadavky na datové formáty, ve kterých jsou distribuce datových sad zveřejňovány, a umožňuje i využití různých proprietárních či komerčních formátů, není považován za dostatečný stupeň otevřenosti.

### Stupeň otevřenosti 3

Stupeň otevřenosti 3 oproti stupni 2 navíc vyžaduje, aby specifikace formátu, ve kterém je distribuce datové sady vyjádřena, byla otevřená. Tzn.: aby byla vyhledatelná a zdarma dostupná v síti WWW a aby existovaly volně dostupné programovací nástroje pro jejich zpracování (tj. knihovny apod.). Příkladem mohou být např. formáty definované v podobě RFC (např. formát CSV, který je definovaný v RFC 4180) nebo formáty definované konsorciem WWW (např. formát XML, který je definovaný v W3C Recommendation Extensible Markup Language (XML) 1.0 (Fifth Edition)).

V distribuci datové sady mohou být vyjádřeny pouze údaje tvořící datovou sadu a konstrukce vyjadřující jejich typ či sémantiku (sémantiku vyjadřuje např. hlavička tabulky v CSV souboru nebo XML tagy a atributy v XML elementu ohraničující údaj s určitým významem). Do distribuce datové sady nepatří konstrukce vyjadřující formátování (barvy nebo styl písma, v tabulkových datových sadách pak navíc ohraničení či slučování buněk, atd.), konstrukce s grafickými prvky, které nejsou součástí údajů v datové sadě, apod.

Do stupně otevřenosti 3 nespadá formát PDF. Taktéž sem nespádají formáty kancelářských aplikací bez otevřené specifikace, jako jsou např. formáty Microsoft Office ve verzích do roku 2003 (tj. DOC, XLS apod.).

Do stupně otevřenosti 3 spadají formáty kancelářských aplikací s otevřenou specifikací (např. formáty Office Open XML či OpenDocument). Distribuce datových sad vyjádřené v těchto formátech však mohou obsahovat pouze údaje tvořící datovou sadu, nikoliv instrukce, které definují jejich formátování při zobrazování na obrazovce či tisku.

Pro prostorová data je vhodné zvolit některý z otevřených formátů OGC (Open Geospatial Consortium), jako je [GML](#), [KML](#) nebo [GeoPackage](#).

Za otevřený a pro některé typy dat i vhodný formát můžeme považovat proprietární ESRI Shapefile (firma ESRI neklade žádná licenční omezení na jeho použití) nebo komunitní GeoJSON (ten zatím nebyl žádným oficiálním orgánem standardizován). Oba tyto formáty mají širokou podporu v dalších softwarech, nejsou ale vhodné na všechny typy datových sad.

### Stupeň otevřenosti 4

Stupeň otevřenosti 4 zavádí povinnost v distribuci datové sady identifikovat entity, kterých se týkají údaje obsažené v datové sadě. Identifikátory musí mít tvar Internationalized Resource Identifier (IRI). Např. v distribuci datové sady o knihách musí mít všechny entity jako knihy, autoři či vydavatelé přiřazen identifikátor ve tvaru IRI.

### Stupeň otevřenosti 5

Stupeň otevřenosti 5 je nejvyšším stupněm otevřenosti. Dále rozšiřuje stupeň 4. Vyžaduje, aby distribuce splňovala standardy propojených dat (angl. Linked Data). Standardy propojených dat jsou postaveny nad dobře známými standardy sítě WWW a umožňují vyjadřovat souvislosti mezi různými datovými sadami v podobě strojově zpracovatelných odkazů, které se podobají hypertextovým odkazům v síti WWW.



## 4.1.2 Stupně otevřenosti v kontextu veřejné správy České republiky

V kontextu veřejné správy České republiky je nejnižším stupněm otevřenosti stupeň 3. Datové sady publikované na stupních otevřenosti 0, 1 a 2 nejsou považovány za otevřené.

Jednotlivé orgány veřejné správy mohou publikovat své datové sady na nižších stupních otevřenosti, ale tyto datové sady pak nejsou otevřené. Neznamena to však, že otevřená datová sada nesmí být publikována na stupni otevřenosti 2. Je vždy nutno zajistit publikaci jejího kompletního obsahu na stupni otevřenosti alespoň 3 a vedle toho lze její obsah publikovat i na stupni 2 (příp. 1, pokud to je smysluplné a účelné).

V praxi to znamená, že uživatelům dat je dostupná jak varianta na úrovni 2, tak varianta na úrovni 3. V případě datové sady s tabulkou je např. možné zveřejnit distribuci datové sady v podobě XLSX souboru, který obsahuje řadu grafických prvků a formátování pro přehledné zobrazení tabulky na obrazovce uživatele - čtenáře (stupeň 2) a vedle toho další distribuci této datové sady v podobě jednoduchého CSV souboru, který žádné grafické prvky a formátování neobsahuje a umožňuje uživatelům - programátorům aplikací a datovým analytikům co nejsnazší přístup k údajům v tabulce (stupeň 3).

Dále je možné, aby poskytovatel provozoval nad svými otevřenými daty nějakou webovou, mobilní či jinou softwarovou aplikaci, která data zpřístupňuje široké veřejnosti (kterou často nezajímají stroje čitelná data, ale právě aplikace nad nimi vytvořené). Existence nebo neexistence takových aplikací však nijak neovlivňuje otevřenost datové sady. Tj. pokud poskytovatel dat provozuje pouze takovou aplikaci (např. HTML formulář pro vyhledávání záznamů a jejich následné zobrazení) a neposkytuje kompletní obsah datové sady na stupni otevřenosti 3, neposkytuje otevřená data.

Z hlediska prostorových dat je zásadní zákon 123/1998 Sb. o právu na informace o životním prostředí, který zajišťuje transpozici evropské Směrnice INSPIRE do české legislativy. Směrnice INSPIRE se zabývá standardizací prostorových dat v Evropě, určuje jejich strukturu, definuje způsob jejich publikace a usnadňuje jejich vyhledávání prostřednictvím vyhledávací služby nad katalogem metadatových záznamů. Směrnice INSPIRE neukládá povinnost všechny datové sady otevírat, určuje však způsob, jak se mají prostorová data publikovat. Publikace prostorových dat je dle Směrnice INSPIRE zakotvena mezi úrovněmi 3 (otevřené formáty a webové služby pro poskytování dat) a 4 (jednotlivé datové sady a služby jsou vyhledatelné na úrovni metadatových záznamů). Směrnice INSPIRE nepřináší nové standardizační postupy. Využívá stávajících standardů a norem a na jejich základě vytváří komplexní návod pro standardizaci a publikaci.

## 5 Důležité pojmy v oblasti otevřených dat

**Datová sada** - Množina souvisejících dat, vztahující se k jednomu věcnému tématu.

**Distribuce datové sady** - Datový zdroj v podobě elektronického souboru, který zpřístupňuje data datové sady nebo jejich část.

**Katalogizační nástroj** - Softwarový nástroj umožňující provozovat lokální katalog otevřených dat.

**Lokální katalog otevřených dat** - Softwarová aplikace umožňující zobrazení, vytváření, úpravu, správu a rušení katalogizačních záznamů o otevřených datech poskytovaných povinným subjektem. Katalog umožňuje katalogizační záznamy poskytovat v plném rozsahu pomocí dálkového přístupu, včetně funkcionality vyhledávání a získání přehledového seznamu všech katalogizačních záznamů.

**Metadata datové sady** - Data popisující datovou sadu, zejména její věcný obsah, časové, územní a další souvislosti.

**Metadata distribuce datové sady** - Data popisující distribuci datové sady, zejména její strukturu, časové a další souvislosti.

**Míra otevřenosti dat** - Míra otevřenosti dat je vyjádřena pomocí 5 stupňů otevřenosti. Stupně 1 - 5 pak znamenají různou úroveň otevřenosti dat (s rostoucím stupněm roste i otevřenost). Někdy bývá přidáván stupeň 0, který znamená, že datová sada není poskytována v podobě otevřených dat.

**Národní geoportál INSPIRE** - Národní metadatový katalog dat o životním prostředí publikovaných orgány veřejné správy. Obsahuje metadata o odkazy na datové sady, webové služby a aplikace. Ne všechny zde nalezené datové sady jsou publikovány v režimu otevřených dat, na druhou stranu se jedná o cenný zdroj informací o tom, jaká data veřejná správa má k dispozici. <http://geoportal.gov.cz/>

**Národní katalog otevřených dat** - Centrální katalog otevřených dat provozovaný a spravovaný Ministerstvem vnitra ČR, který je dostupný způsobem umožňujícím vzdálený přístup a obsahuje katalogizační záznamy o datových sadách a jejich distribucích. Národní katalog je přístupný prostřednictvím [Portálu veřejné správy](#) a může obsahovat rovněž distribuce datových sad. Ministerstvo vnitra ČR zajišťuje zálohování v katalogu uložených metadat a datových sad.

**Otevřená data** - Informace poskytované tak, že jsou poskytovány v maximálním možném rozsahu, způsobem umožňujícím dálkový přístup, ve strojově čitelném a otevřeném formátu, za jasně definovaných podmínek užití neomezujících způsob ani účel následného využití poskytovaných informací včetně umožnění jejich využití pro přímý a nepřímý hospodářský prospěch a dostupné uživatelům při vynaložení minima možných nákladů na jejich získání.

**Ontologie** - Ontologie (někdy též nazývaná slovník) je obdobou datového schématu ve světě tabulkových dat nebo XML dat. Zatímco však ve světě tabulkových či XML dat popisuje datové schéma striktní syntaxi např. CSV či XML distribucí datových sad, ontologie popisuje datovou sadu na konceptuální úrovni. Konkrétně definuje třídy a vlastnosti, které lze v RDF distribucích datových sad použít.

**Propojená data** - Datové sady poskytované v takové podobě, kdy odkazují na sebe navzájem, a konečný uživatel obdrží po zadání požadavku v aplikačním rozhraní informaci složenou z dat pocházejících z mnoha různých datových sad poskytovaných mnoha různými subjekty.

**Publikační plán** - Publikační plán je seznamem datových sad plánovaných k publikaci v podobě otevřených dat, sestavovaný Koordinátorem otevírání dat. Stanovuje, jaké datové sady Poskytovatele dat budou publikovány, jaký bude harmonogram otevírání, a pro každou datovou sadu určuje podmínky užití datové sady. Publikační plán je aktualizován v roční periodě.

**Směrnice INSPIRE** - Směrnice Evropské Komise a Rady zabývající se prostorovými daty, jejich strukturou a formou jejich publikace. Do české legislativy je transponována novelou [zákona 123/1998 Sb.](#), o právu na informace o životním prostředí. Zákon nenařizuje otevírání dat, ale ukládá pořizovatelům dat povinnost tato data poskytovat (§ 11b, odstavec 3). Zároveň umožňuje tato data zpoplatnit částkou, která však nesmí přesáhnout minimální výši nezbytnou k zajištění potřebné kvality a zpřístupňování prostorových dat a služeb založených na prostorových datech (§ 11c, odstavec 3). Směrnice kategorizuje prostorová data do 34 témat, definuje strukturu, ve které mají být data publikována, a způsob, jakým mají být zpřístupněna a nařizuje jejich opatření metadaty, vše dle platných norem. Konkrétní způsoby implementace INSPIRE jsou popsány v „implementačních pravidlech“ a „technických průvodcích“.



## 6 Role a jejich činnosti

### 6.1 Vedení poskytovatele dat

Vedení poskytovatele dat spouští proces otevírání dat svým rozhodnutím, obsazuje pracovníky organizace do jednotlivých rolí a schvaluje publikační plán.

Při vytváření publikačního plánu Vedení poskytovatele dat zajišťuje následující činnosti:

- rozhoduje o otevírání dat organizace,
- obsazuje roli koordinátora,
- schvaluje interní legislativu,
- obsazuje další role v kontextu otevřených dat,
- schvaluje publikační plán.

### 6.2 Koordinátor otevírání dat

Koordinátor otevírání dat řídí celý proces otevírání dat organizace a je za něj zodpovědný. Některé kroky realizuje s pomocí ostatních rolí.

V rámci vytvoření publikačního plánu Koordinátor otevírání dat zajišťuje následující činnosti:

- připravuje interní legislativu,
- připravuje publikační plán,
- rozhoduje o způsobu katalogizace otevřených dat,
- zveřejňuje publikační plán,
- periodicky vyhodnocuje plnění publikačního plánu a informuje vedení.

V rámci zveřejnění konkrétní datové sady v otevřené podobě Koordinátor otevírání dat zajišťuje následující činnosti:

- kontroluje připravenou datovou sadu k publikaci,
- zajišťuje zveřejnění datové sady.

### 6.3 Kurátor dat

Kurátor dat je garantem jemu příslušejících datových sad. Spolupracuje s Koordinátorem otevírání dat na umístění těchto datových sad do publikačního plánu a připravuje je ke zveřejnění. Vykonává následující činnosti:

V rámci vytvoření publikačního plánu Kurátor dat zajišťuje následující činnost:

- navrhuje datové sady ke zveřejnění

V rámci zveřejnění konkrétní datové sady v otevřené podobě Kurátor dat zajišťuje následující činnosti:

- analyzuje datovou sadu,
- připravuje datovou sadu k publikaci.

### 6.4 Správce katalogu otevřených dat

Správce katalogu otevřených dat je zodpovědný za náplň katalogu otevřených dat organizace.

V rámci zprovoznění lokálního katalogu otevřených dat správce katalogu otevřených dat zajišťuje následující činnosti:

- připravuje lokální katalog otevřených dat.

V rámci zveřejnění konkrétní datové sady v otevřené podobě správce katalogu otevřených dat zajišťuje následující činnosti:

- spravuje katalogizační záznam.

## 6.5 IT specialista

IT specialista pomáhá Kurátorům dat při přípravě datových sad ke zveřejnění a Správcí katalogu otevřených dat s provozem datového katalogu.

V rámci vytvoření publikačního plánu IT specialista zajišťuje následující činnosti:

- spolupracuje na návrhu datových sad k publikaci,
- spolupracuje na určení způsobu katalogizace.

V rámci zprovoznění lokálního katalogu otevřených dat IT specialista zajišťuje následující činnosti:

- instaluje lokální katalog,
- provozuje lokální katalog.

V rámci zveřejnění konkrétní datové sady v otevřené podobě IT specialista zajišťuje následující činnosti:

- spolupracuje na analýze datové sady,
- spolupracuje na přípravě datové sady k publikaci,
- zveřejňuje datovou sadu na internetu.

## 7 Postup otevírání dat

Otevírání dat je proces, který nezahrnuje pouze zveřejnění několika vybraných datových sad. Měl by nejprve zahrnovat fázi, ve které jsou zanalyzovány datové sady a je rozhodnuto, které je možno otevřít a kdy. Výsledkem této fáze by měl být publikační plán. Další fáze potom znamená samotné zveřejňování datových sad. Ta zahrnuje návrh technické podoby datových sad (formáty a datová schémata), jejich publikaci v síti WWW a katalogizaci. Je možné se rozhodnout katalogizovat v Národním katalogu otevřených dat nebo ve vlastním lokálním katalogu. Ve druhém případě je nutno lokální katalog otevřených dat zprovoznit, což zahrnuje i volbu vhodného softwarového nástroje. Nedílnou součástí zprovoznění lokálního katalogu je vždy jeho registrace v Národním katalogu otevřených dat (NKOD).

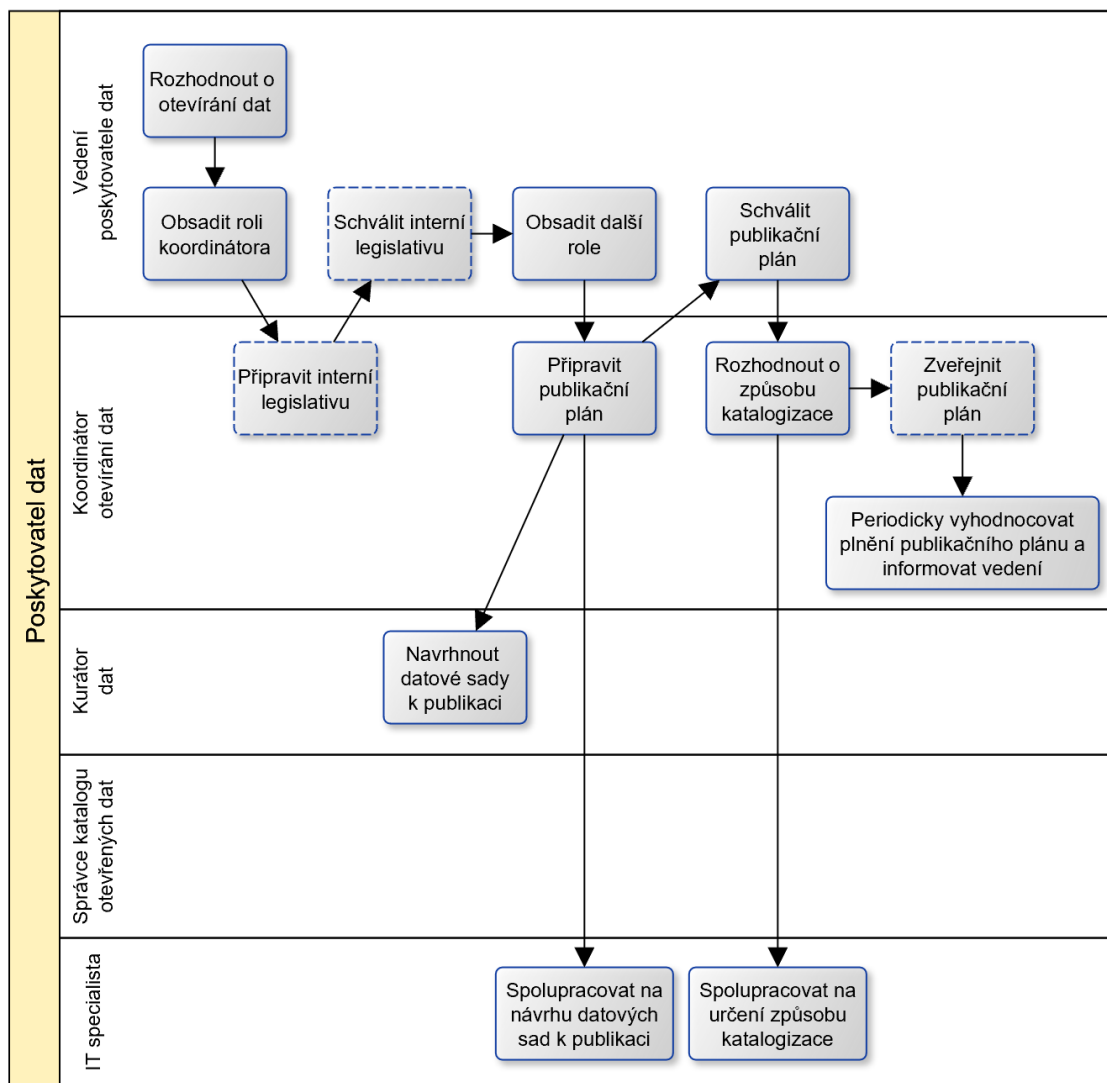
Jednotlivé fáze jsou detailně popsány v následujících kapitolách:

- Vytvoření publikačního plánu
- Publikace datové sady v otevřené podobě
- Zprovoznění lokálního katalogu otevřených dat

## 8 Vytvoření publikačního plánu

Před samotnou publikací prvních datových sad v otevřené podobě je vhodné nejprve definovat postup otevírání dat a role, které se jej účastní, včetně jejich obsazení. Poté je nutno provést analýzu toho, jaké datové sady budou zveřejněny v otevřené podobě. Výsledkem analýzy by měl být publikační plán. Publikační plán je seznamem datových sad plánovaných k publikaci v podobě otevřených dat.

Jednotlivé činnosti jsou zachyceny v následujícím diagramu. Po kliknutí na činnost se dostanete na detailní popis její doporučené náplně. Jednotlivé plavecké dráhy odpovídají rolím, které by se měly procesu otevírání dat účastnit. Po kliknutí na roli se dostanete na její popis.



Legenda:

- Obdélník s plným okrajem značí povinnou činnost.
- Obdélník s přerušovaným okrajem značí nepovinnou činnost.
- Šipky značí návaznost mezi činnostmi.

## 8.1 Rozhodnutí o otevírání dat

Proces otevírání dat začíná vedení příslušného orgánu veřejné správy rozhodnutím, které zavazuje daný orgán veřejné správy k publikaci otevřených dat.

Může se stát, že pro schválení rozhodnutí není dostatečná většina. Je tedy potřeba příslušnou část vedení přesvědčit, aby s otevíráním dat souhlasila. Jako argument lze použít skutečnost, že řada orgánů veřejné správy na všech úrovních již otevřená data poskytuje. Jedná se například o tyto orgány veřejné správy:

- [Ministerstvo financí](#)
- [Ministerstvo vnitra](#)
- [Česká obchodní inspekce](#)
- [Český telekomunikační úřad](#)
- [Kraj Vysočina](#)
- [Město Děčín](#)

Jako argumenty lze použít řadu přínosů, které otevřená data přináší.

### 8.1.1 Přínosy pro veřejnost

- **Posílení transparentnosti** - lepší dostupnost a přístup k informacím o fungování orgánů veřejné správy, například přístup ke smlouvám a veřejným zakázkám;
- **Zefektivnění veřejné správy a zlepšení kvality života** - veřejný sektor skrze otevřená data informuje veřejnost o plánovaných změnách a akcích a tím vytváří větší důvěru mezi občanem a orgánem veřejné správy;
- **Kontrola veřejné správy** - zveřejněná data umožňují kontrolu, jak se hospodaří s daněmi nebo jaké jsou náklady organizací, které podporujeme;
- **Zapojení občanů do rozhodování** - občané se mohou díky datům a analýzám kvalifikovaněji podílet na fungování státu;
- **Datová žurnalistika** - otevřená data jsou nezastupitelným zdrojem informací pro novináře a datové analytiky.
- **Služby** - nad otevřenými daty vytvářejí specialisté nové aplikace a služby s přidanou hodnotou, které novým způsobem zpřístupňují veřejnosti data vytvářená veřejnou správou.

### 8.1.2 Přínosy pro ekonomiku

Otevřená data **přináší nové ekonomické příležitosti a jsou zdrojem, pracovních nabídek a nových služeb** – lze je využít třeba v dopravě, logistice, zdravotnictví či bankovníctví. Firmy pracují s daty jako se surovinou, vytváří nad nimi aplikace, které generují přidanou hodnotu, zisk a nová pracovní místa.

**Otevřená data v číslech:**

- 100 miliard eur ročně na straně institucí a 450 miliard eur na straně klientů mohou západním ekonomikám ušetřit otevřená data ([Studie McKinsey & comp](#)),
- 16 miliard liber přineslo využití dat veřejného sektoru ve Velké Británii v roce 2011, kdy vláda otevřela přes 8000 datasetů ([Deloitte 2012](#)),
- 2 miliardy liber ročně přináší otevřená data britské ekonomice ([Rufus Pollock 2011](#)),
- až 40 miliard eur ročně mohou přinést ekonomice v Evropské unii otevřená data (Studie EU z roku 2011),
- 20 milionů dolarů dosáhly úspory Kalifornie poté, co návštěvníci portálu zaměřeného na zvyšování transparentnosti poukázali na nepotřebné náklady (Studie Capgemini Consulting).

### 8.1.3 Přínosy pro poskytovatele dat

- **snížení počtu dotazů dle zák. č. 106/1999 Sb.**, - zveřejněním otevřených dat se může výrazně snížit počet žádostí dle zák. č. 106/1999 Sb. a tím se výrazně ulehčí poskytovatelům dat,
- **zlepšení vnímání veřejné správy veřejností a zapojení veřejnosti do zlepšování práce úřadů a veřejnosti** - veřejný sektor skrze otevřená data informuje veřejnost o

plánovaných změnách a akcích a tím vytváří větší důvěru mezi občanem a organizací,

- **zlepšení procesů a dat veřejné správy, zlepšení komunikace a spolupráce VS** - publikace dat ve formátu otevřených dat může napomoci i komunikaci mezi jednotlivými orgány veřejné správy. Publikace otevřených dat může nahradit někdy zbytečně složité předávání dat mezi jednotlivými organizacemi,
- **zlepšení služeb veřejné správy/zlepšení kvality života** - veřejný sektor skrze otevřená data informuje veřejnost o plánovaných změnách a akcích a tím vytváří větší důvěru mezi občanem a orgánem veřejné správy,
- **pořádek ve vlastních datech** - analýza dat, katalogizace a publikace umožní orgánu veřejné správy uspořádat svá vlastní data a optimalizovat tak vlastní interní procesy.

Zajímavé zdroje inspirací najdete na webu [www.otevrenadata.cz](http://www.otevrenadata.cz).

Podívejte se také na [příklady zajímavých aplikací](#) postavených na otevřených datech v soutěži [Společně otevíráme data](#), kterou každoročně pořádá [Fond Otakara Motejla](#).

## 8.2 Obsazení role Koordinátora otevírání dat

Po vydání rozhodnutí pověří Vedení poskytovatele dat vybraný útvar (osobu) rolí Koordinátora otevírání dat. Koordinátor otevírání dat zajišťuje součinnost a kontrolu výstupů všech ostatních rolí, které se na otevírání dat podílejí.

Koordinátor otevírání dat zodpovídá za:

- přípravu publikačního plánu;
- návrh Kurátorů dat pro jednotlivé datové sady;
- poskytování metodické podpory Kurátorům dat při plnění povinností v oblasti otevřených dat;
- zveřejnění publikačního plánu;
- rozhodnutí o způsobu katalogizace (Národní katalog otevřených dat nebo i lokální katalog otevřených dat);
- zajištění finální kontroly dat před publikací;
- zajištění publikace datových sad a katalogizačních záznamů (metadat);
- komunikaci připomínek uživatelů otevřených dat a reportování o publikaci otevřených dat Vedení poskytovatele dat a příp. tiskovému oddělení Poskytovatele dat.

Vedení poskytovatele dat může Koordinátora otevírání dat pověřit přípravou interní legislativy, která bude upravovat proces otevírání dat, vymezí jednotlivé role a stanoví jejich povinnosti. Doporučujeme vytvořit legislativu v podobě směrnice. Vzor směrnice je k dispozici k použití (viz Vzor směrnice "Publikace a katalogizace otevřených dat").

## 8.3 Příprava interní legislativy

Prvním úkolem Koordinátora otevírání dat je navrhnout interní legislativu pro oblast otevřených dat. Doporučujeme připravit ji v podobě směrnice. Při vytváření je možné použít připravený návrh (viz Vzor směrnice "Publikace a katalogizace otevřených dat").

Návrh Koordinátor otevírání dat předá Vedení poskytovatele dat ke schválení.

## 8.4 Schválení interní legislativy

Koordinátor otevírání dat jako svůj první výstup připraví návrh interní legislativy. Vedení poskytovatele dat nyní musí návrh zkontrolovat a schválit.



Při kontrole doporučujeme se zaměřit především na následující:

- Pokud byla využita Vzorová interní směrnice publikace a katalogizace otevřených dat, potom je třeba zkontrolovat, že byly správně obsazeny zúčastněné role.
- Pokud byla vytvářena nová směrnice nezávisle na vzorovém podkladu, doporučujeme ustanovení směrnice verifikovat vůči základním krokům postupu publikace a katalogizace otevřených dat:

IT specialista pomáhá Kurátorům dat při přípravě datových sad ke zveřejnění a Správci katalogu otevřených dat s provozem datového katalogu.

V rámci vytvoření publikačního plánu IT specialista zajišťuje následující činnosti:

- spolupracuje na návrhu datových sad k publikaci,
- spolupracuje na určení způsobu katalogizace.

V rámci zprovoznění lokálního katalogu otevřených dat IT specialista zajišťuje následující činnosti:

- instaluje lokální katalog,
- provozuje lokální katalog.

V rámci zveřejnění konkrétní datové sady v otevřené podobě IT specialista zajišťuje následující činnosti:

- spolupracuje na analýze datové sady,
- spolupracuje na přípravě datové sady k publikaci,
- zveřejňuje datovou sadu na internetu.

## 9 Postup otevírání dat

Otevírání dat je proces, který nezahrnuje pouze zveřejnění několika vybraných datových sad. Měl by nejprve zahrnovat fázi, ve které jsou zanalyzovány datové sady a je rozhodnuto, které je možno otevřít a kdy. Výsledkem této fáze by měl být publikační plán. Další fáze potom znamená samotné zveřejňování datových sad. Ta zahrnuje návrh technické podoby datových sad (formáty a datová schémata), jejich publikaci v síti WWW a katalogizaci. Je možné se rozhodnout katalogizovat v Národním katalogu otevřených dat nebo ve vlastním lokálním katalogu. Ve druhém případě je nutno lokální katalog otevřených dat zprovoznit, což zahrnuje i volbu vhodného softwarového nástroje. Nedílnou součástí zprovoznění lokálního katalogu je vždy jeho registrace v Národním katalogu otevřených dat (NKOD).

Jednotlivé fáze jsou detailně popsány v následujících kapitolách:

- Vytvoření publikačního plánu
  - Publikace datové sady v otevřené podobě
  - Zprovoznění lokálního katalogu otevřených dat
1. Vytvoření publikačního plánu
  2. Publikace datové sady v otevřené podobě
  3. **Chyba! Nenalezen zdroj odkazů.**

### 9.1 Obsazení dalších rolí v kontextu otevřených dat

Schválená interní legislativa, např. směrnice, určuje role v procesu otevírání dat, které nyní musí Vedení poskytovatele dat obsadit.



### 9.1.1 Kurátor dat

zodpovídá za:

- analýzy datových sad předcházející jejich zveřejnění;
- zajištění, že otevřené datové sady budou publikovány v souladu s platnými právními předpisy ČR a Standardy publikace a katalogizace otevřených dat VS ČR (zveřejněných na tomto portálu);
- určení případné nutnosti transformace datových sad, např. anonymizaci, agregaci nebo jinou změnu struktury datových sad;
- správu a kontrolu kvality datových sad (tj. mj. kontrola správnosti, úplnosti, konzistentnosti) před jejich publikací a katalogizací;
- spolupráci při procesu zveřejnění datové sady s Koordinátorem otevírání dat a IT specialistou;
- předání datových souborů s datovými sadami a návrhu obsahu katalogizačních záznamů (metadat) Správci katalogu otevřených dat.

Kurátor dat je povinen poskytnout IT specialistovi a Správci katalogu otevřených dat potřebnou součinnost při:

- transformaci datové sady a její kontrole;
- určení pracnosti a přínosů plynoucích z publikace otevřených datových sad;
- vytváření a aktualizaci katalogizačních záznamů.

### 9.1.2 Správce katalogu otevřených dat

zodpovídá za:

- formální kontrolu a správu katalogizačních záznamů (metadat);
- vkládání katalogizačních záznamů a příp. i datových souborů do Národního katalogu otevřených dat;
- správu katalogizačních záznamů v Národním katalogu otevřených dat.

V případě, že bude v organizaci implementován i lokální katalog otevřených dat, Správce katalogu otevřených dat dále zodpovídá za

- zajištění výběru softwarového nástroje pro lokální katalog otevřených dat;
- zajištění implementace lokálního katalogu otevřených dat;
- registraci lokálního katalogu v Národním katalogu otevřených dat;
- archivaci katalogizačních záznamů a dat uložených v lokálním katalogu otevřených dat.

### 9.1.3 IT specialista

zodpovídá za:

- poskytování technické podpory a součinnosti Kurátorovi dat při plnění povinností v oblasti otevřených dat, zejména při analýze předcházející zveřejnění dat a přípravě datových sad k otevření;
- případné zajištění instalace a provozu lokálního katalogu otevřených dat.

## 9.2 Příprava publikačního plánu

Publikační plán organizace sestavuje Koordinátor otevírání dat. Plán stanovuje, jaké datové sady Poskytovatele dat budou publikovány v podobě otevřených dat a jaký bude harmonogram otevírání. Pro každou datovou sadu navíc určuje podmínky užití datové sady.

Koordinátor otevírání dat ve spolupráci s Kurátory dat vytváří seznam datových sad, které navrhuje k otevření. Jedná se o prvotní návrh, který obsahuje různé náměty na datové sady, které budou v dalších krocích potvrzeny nebo zamítnuty.

*Jak vytvořit seznam datových sad k otevření je uvedeno v kapitole „Vzorové podklady*

*Vzorový publikační plán pro ministerstva*

- Vzorový publikační plán pro centrální orgány
- Vzorový publikační plán pro kraje
- Vzorový publikační plán pro obce s rozšířenou působností
- Vzorový publikační plán pro ostatní obce

Vytvoření seznamu datových sad “.

Jednotlivé datové sady by měl Koordinátor otevírání dat ve spolupráci s Kurátory dat analyzovat z pohledu přínosů a rizik jejich publikace.

*Jak určit přínosy publikace datových sad se dozvíte v kapitole „Koordinátor otevírání dat nejprve ve spolupráci s Kurátory dat vytvoří seznam datových sad organizace, které navrhuje k otevření. Není nutné sepisovat všechny datové sady, které v daném orgánu veřejné správy existují. V první fázi stačí vybrat několik datových sad a později je rozšiřovat. Také není nutné sepisovat pouze ty, u kterých je dopředu jasné, že je možné je otevřít. V dalších krocích jsou pak hodnoceny přínosy a rizika otevření a datová sada může být ze seznamu vyjmuta.*

Při vytváření seznamu doporučujeme postupovat následovně:

1. Koordinátor otevírání dat předá jednotlivým Kurátorům dat seznam doporučených datových sad uvedený ve vzorovém publikačním plánu, aby vybrali datové sady, které spadají do jejich kompetencí a je možné je zveřejnit v podobě otevřených dat, a případně seznam rozšířili o další datové sady v jejich kompetenci.
2. Koordinátor otevírání dat konsoliduje datové sady vybrané Kurátory dat do jednoho seznamu.
3. Koordinátor otevírání dat kontroluje, že pro každou datovou sadu v seznamu jsou doplněny následující atributy (většinu lze převzít ze vzorového publikačního plánu). Pokud nejsou, komunikuje s příslušnými Kurátory dat o jejich doplnění:
  - název datové sady,
  - popis datové sady,
  - Kurátor dat.

Určení přínosů otevření datových sad“.

*Jak určit rizika publikace datových sad je vysvětleno v kapitole „Určení rizik otevření datových sad“.*

Na základě provedené analýzy by měl Koordinátor otevírání dat provést revizi navrženého seznamu. To znamená především odstranit ze seznamu datové sady, u nichž jsou rizika publikace příliš vysoká nebo přínosy příliš nízké.

Jak revidovat seznam datových sad k otevření je uvedeno v kapitole „Revize seznamu datových sad určených k otevření“.

Pro datové sady v revidovaném seznamu musí Koordinátor otevírání dat stanovit podmínky jejich užití.

*Jak stanovit podmínky užití datových sad je uvedeno v kapitole „Výsledek analýzy přínosů a rizik otevření jednotlivých datových sad může ukázat, že přínosy jsou nízké a/nebo rizika jsou příliš vysoká.*

Doporučujeme tedy vytvořený seznam datových sad revidovat dle následujících pravidel:

1. Pokud se nepodařilo identifikovat výraznější přínos otevření datové sady, zvažte, zda se vyplatí ji vůbec otevřít;
2. Pokud bylo identifikováno významné riziko a nejste schopni nalézt způsob, jak riziko eliminovat, zvažte, zda je rozumné datovou sadu v její aktuální podobě otevřít.

Současně je nutno jak přínosy, tak i rizika otevřených dat revidovat vždy při aktualizaci publikačního plánu, tj. nejméně jednou ročně.

Stanovení podmínek užití datových sad“.

Nakonec by měl Koordinátor otevírání dat výsledný seznam datových sad navržených k otevření utřídit dle priorit a stanovit tak harmonogram publikace datových sad.

*Jak sestavit harmonogram publikace datových sad je vysvětleno v kapitole „Sestavení harmonogramu publikace datových sad“.*

Výsledný harmonogram publikace vybraných datových sad vytváří publikační plán, který bude postoupen Vedení poskytovatele dat ke schválení. Publikační plán ještě není v tuto chvíli kompletní, neboť neobsahuje pro datové sady technické parametry (stupeň otevřenosti a datový formát), které však nejsou pro Vedení poskytovatele dat důležité.

### 9.2.1 Vzorové podklady

- Vzorový publikační plán pro ministerstva
- Vzorový publikační plán pro centrální orgány
- Vzorový publikační plán pro kraje
- Vzorový publikační plán pro obce s rozšířenou působností
- Vzorový publikační plán pro ostatní obce

## 9.3 Vytvoření seznamu datových sad

Koordinátor otevírání dat nejprve ve spolupráci s Kurátory dat vytvoří seznam datových sad organizace, které navrhuje k otevření. Není nutné sepisovat všechny datové sady, které v daném orgánu veřejné správy existují. V první fázi stačí vybrat několik datových sad a později je rozšiřovat. Také není nutné sepisovat pouze ty, u kterých je dopředu jasné, že je možné je otevřít. V dalších krocích jsou pak hodnoceny přínosy a rizika otevření a datová sada může být ze seznamu vyjmuta.

Při vytváření seznamu doporučujeme postupovat následovně:

4. Koordinátor otevírání dat předá jednotlivým Kurátorům dat seznam doporučených datových sad uvedený ve vzorovém publikačním plánu, aby vybrali datové sady, které spadají do jejich kompetencí a je možné je zveřejnit v podobě otevřených dat, a případně seznam rozšířili o další datové sady v jejich kompetenci.
5. Koordinátor otevírání dat konsoliduje datové sady vybrané Kurátory dat do jednoho seznamu.
6. Koordinátor otevírání dat kontroluje, že pro každou datovou sadu v seznamu jsou doplněny následující atributy (většinu lze převzít ze vzorového publikačního plánu). Pokud nejsou, komunikuje s příslušnými Kurátory dat o jejich doplnění:
  - název datové sady,
  - popis datové sady,
  - Kurátor dat.

## 9.4 Určení přínosů otevření datových sad

Pro otevřená data je důležité, aby jejich existence přinášela nějaký užitek. U každé navrhované datové sady je proto důležité analyzovat, jaké přínosy její otevření přinese. Analýzu provádí Koordinátor otevírání dat ve spolupráci s Kurátory dat.

Při analýze přínosů doporučujeme postupovat následovně:

1. Pro datové sady uvedené ve vzorovém publikačním plánu jsou přínosy identifikovány a lze je převzít.
2. Pokud jste do seznamu přidali vlastní datové sady, ohodnoťte pásmově (0 - žádný přínos, 1 - nízký přínos, 2 - střední přínos, 3 - významný přínos) následující přínosy:
  - **Posílení transparentnosti** - lepší dostupnost a přístup k informacím o fungování orgánů veřejné správy, například přístup ke smlouvám a veřejným zakázkám;
  - **Podpora ekonomického růstu** - zveřejněním informací veřejného sektoru jako zdroje pro nové inovace, služby, produkty a podnikatelských příležitostí. Výsledkem může být hospodářský a ekonomický růst;
  - **Zlepšení služeb veřejné správy a zlepšení kvality života** - veřejný sektor skrze otevřená data informuje veřejnost o plánovaných změnách a akcích a tím vytváří větší důvěru mezi občanem a orgány veřejné správy;
  - **Podpora opětovného použití dat** - strojová čitelnost, dostupnost metadat a právní otevřenost dat zajišťují větší znovuvyužitelnost dat veřejného sektoru. Díky tomu je jednodušší vyvíjet aplikace a služby postavené na otevřených datech,
  - **Zlepšení vnímání veřejné správy veřejností** - díky otevřeným datům jsou orgány veřejné správy otevřenější a transparentnější, což zlepšuje reputaci poskytovatele dat;
  - **Zlepšení procesů a dat veřejné správy, zlepšení komunikace a spolupráce veřejné správy** - publikace dat ve formátu otevřených dat může napomoci i komunikaci mezi jednotlivými orgány veřejné správy. Publikace otevřených dat může nahradit někdy zbytečně složité předávání dat mezi jednotlivými orgány veřejné správy;
  - **Zamezení chyb vzniklých při práci s daty** - dostupnost dat ve strojově čitelném formátu minimalizuje potřebu manuálně přepisovat data z dokumentů do jiných systémů a celkově se sníží chybovost;
  - **Snížení počtu dotazů dle zák. č. 106/1999 Sb.** - zveřejněním otevřených dat se může výrazně snížit počet žádostí dle zák. č. 106/1999 Sb. a tím se výrazně ulehčí poskytovatelům dat;
  - **Pořádek ve vlastních datech** - analýza dat, katalogizace a publikace umožní organizaci uspořádat svá vlastní data a optimalizovat tak vlastní interní procesy;
  - **Zvýšení hodnoty dat** strojová čitelnost a otevřenost dat spolu s metadaty přidávají datům novou hodnotu, která může být ještě více rozvinuta v kontextu Otevřených propojitelných dat.

## 9.5 Určení rizik otevření datových sad

Otevřená data přinášejí i rizika. Závažná rizika je proto nutno dopředu identifikovat, aby bylo možné je řídit, až budou data otevřena. Analýzu rizik opět provádějí Koordinátor otevírání dat s Kurátory dat.

Při analýze rizik doporučujeme postupovat následovně:

1. Pro datové sady uvedené ve vzorovém publikačním plánu jsou rizika identifikována a můžete je převzít.
2. Pokud jste do seznamu přidali vlastní datové sady, ohodnoťte pásmově (0 - žádné riziko, 1 - nízké riziko, 2 - střední riziko, 3 - významné riziko) následující rizika:
  - **Zveřejnění dat v rozporu se zákonem** - vždy nutno ověřit, zda publikaci datové sady nebrání související legislativa;



- **Porušení ochrany obchodního tajemství** - dále je nutno zvážit, zda publikace datové sady neporušuje obchodní tajemství, specifikované např. ve smlouvě s dodavatelem;
- **Porušení ochrany osobních údajů** - je nutno ověřit, zda součástí otevřených dat nejsou chráněné osobní údaje;
- **Zveřejnění nevhodných dat či informací** - je třeba zvážit, zda publikovaná data nebudou mít negativní dopady, např. z hlediska kvality publikovaných dat;
- **Dezinterpretace dat** - zvážit, zda otevřená data nebude možno dezinterpretovat, např. nevysvětlením používaných pojmů;
- **Absence konzumentů dat** - ověřit, zda je o publikovaná data zájem ze strany budoucích uživatelů dat;
- **Překrývání dat** - ověřit, zda se datové sady nemohou překrývat s jinými již publikovanými datovými sadami.
- **Ohrožení bezpečnosti státu / majetku / osob** - ověřit, zda datové sady nemohou ohrozit bezpečnost státu, majetku anebo osob.

Celkové zhodnocení rizikovitosti datové sady musí být následováno návrhem opatření pro eliminaci či omezení pravděpodobnosti výskytu nebo negativního dopadu rizik. Možná opatření pro zmírnění rizik:

- Anonymizace/agregace dat (rizika č. 1, 2, 3, 8) - pokud nelze poskytnout data primární kvůli požadavkům na jejich ochranu či bezpečnostní požadavky, je vhodné zvážit, zda lze publikovat alespoň anonymizovaná data či statistiky;
- Návrh vhodné komunikační strategie (rizika č. 4, 6) - určení jak bude o datové sadě informováno, identifikovat potenciální negativní reakce a připravit odpovědi. Oslovení cílových skupin uživatelů dat;
- Vytvoření vhodných metadat a doplňujících informací (riziko č. 5) - uvést do metadat popis předpokladů a omezení využití dat. Publikovat doplňující informace – např. metodiku pořízení a zpracování dat, použité výpočty, definice ukazatelů apod.;
- Vhodné vymezení datové sady a propojení na související datové sady (riziko č. 7) - vymezení datové sady s ohledem na minimalizaci překryvů s jinými datovými sadami. Využití již publikovaných datových sad. Propojování na související datové zdroje.

## 9.6 Revize seznamu datových sad určených k otevření

Výsledek analýzy přínosů a rizik otevření jednotlivých datových sad může ukázat, že přínosy jsou nízké a/nebo rizika jsou příliš vysoká.

Doporučujeme tedy vytvořený seznam datových sad revidovat dle následujících pravidel:

3. Pokud se nepodařilo identifikovat výraznější přínos otevření datové sady, zvažte, zda se vyplatí ji vůbec otevřít;
4. Pokud bylo identifikováno významné riziko a nejste schopni nalézt způsob, jak riziko eliminovat, zvažte, zda je rozumné datovou sadu v její aktuální podobě otevřít.

Současně je nutno jak přínosy, tak i rizika otevřených dat revidovat vždy při aktualizaci publikačního plánu, tj. nejméně jednou ročně.

## 9.7 Stanovení podmínek užití datových sad

S publikací datových sad je spojeno možné porušení autorských a databázových práv, která se k dané datové sadě mohou vázat. Před publikací je nezbytné vypořádat se s těmito překážkami poskytnutím licencí tam, kde je to nezbytné. V případě, že datová sada, ani žádná její součást, nejsou jakkoliv chráněny, je přesto vhodné uvést tuto informaci, za účelem zvýšení právní jistoty budoucích uživatelů dat.

Tento návod provede Poskytovatele dat problematikou licencování práv duševního vlastnictví k datovým sadám určeným k publikaci v režimu otevřených dat. V této podkapitole jsou představeny doporučené možnosti nastavení podmínek užití. Samotný návod je rozdělený do tří sekcí, podle práv, se kterými se při zveřejňování otevřených dat setkáváme. Každá sekce se pak skládá z několika kroků a komentáře k relevantním zákonným ustanovením. První sekce se týká možné autorskoprávní ochrany prvků databáze, druhá sekce se týká autorskoprávní ochrany databází a jejich struktury a třetí sekce se týká zvláštních databázových práv pořizovatele databáze.

Před publikací otevřených dat je nezbytné připojit k datové sadě informaci o tom, jak je chráněna z hlediska práv duševního vlastnictví. K tomu doporučujeme následující možnosti nastavení podmínek užití:

I. Veřejná licence CC-BY 4.0, kterou držitel autorského práva umožňuje volné užití svých děl za předpokladu, že uživatel díla jej uvede jako autora (<https://creativecommons.org/licenses/by/4.0/>). Dále je pro její korektní použití třeba uvést rovněž název díla, odkaz na jeho původní umístění a odkaz na znění licence. Zkušenost zatím říká, že použití CC-BY 4.0 bude spíše výjimečné. Tuto licenci použijte v následujících případech:

1. Datová sada je chráněna jako tvůrčí databáze dle § 2 odst. 2 Autorského zákona (tedy, je chráněna jeho struktura) a zároveň nemůže být úředním dílem dle § 3 písm. a) Aut. Z.
2. Datová sada obsahuje autorskoprávně chráněná díla a zároveň můžete tato díla dále licencovat. Pokud je licencovat nemůžete, nelze je ani zveřejnit.

II. Veřejná licence CC0, která slouží jako prostředek ke vzdání se databázových práv pořizovatele databáze (<https://creativecommons.org/publicdomain/zero/1.0/>). Díky tomu bude moci každý vytěžovat a zužítkovat obsah databáze – což je výsledek, kterého chceme otevřenými daty dosáhnout. Použijte se častěji než licence CC-BY 4.0, zejména proto, že se na databázová práva pořizovatele nevztahuje výjimka úředního díla. Použijte se tehdy, když:

- Při tvorbě datové sady byl učiněn pořizovatelem databáze (orgánem veřejné správy) pořízením obsahu (zanesením dat do databáze), jeho ověřením nebo předvedením (zveřejněním) podstatný vklad. Ten může být jak vyjádřitelný penězi, tak nehmotný.

III. V případě že datová sada není jakkoli chráněna, je doporučeno k ní připojit místo licence informaci o tom, že datová sada není dle autorského zákona chráněna (příklad zde: <https://portal.gov.cz/portal/ostatni/volny-pristup-k-ds.html>). Tato varianta je pouze informací, kterou veřejná správa poskytuje a která pomůže následným uživatelům datových sad. Použijte se tehdy, když není nezbytné použít některou z výše uvedených variant. Pro použití této varianty musí být splněny následující podmínky:

1. Pokud datová sada není tvůrčí databází; NEBO pokud je datová sada tvůrčí databází a zároveň existuje veřejný zájem na vyloučení autorskoprávní ochrany této datové sady. Díky tomu se stane úředním dílem dle § 3 písm. a) Aut. Z. Tato varianta bude většinová oproti výše uvedené variantě I. 1).
2. Datová sada neobsahuje autorskoprávně chráněná díla.
3. Datová sada není chráněna databázovými právy pořizovatele databáze.

Obecně platí, že licence je jen potřebná tam, kde je poskytován obsah chráněný některým právem duševního vlastnictví. V opačných případech by se licence neměla nikde u datové sady objevit (“jen tak pro jistotu”), protože to jen povede ke zmatení příjemců dat. Právě pro takové případy, kdy datová sada není chráněna, slouží “Informace o volném přístupu k datové sadě”.



## 9.7.1 Sekce 1 – Autorská práva

Tato sekce řeší otázky případné autorsko-právní ochrany samotných prvků datové sady. Většina případů poskytovaných dat nebude spadat pod definici autorského díla dle § 2 autorského zákona (121/2000 Sb.), nebude se na ně proto vztahovat autorskoprávní ochrana a nebude proto pro ně potřeba jakékoli licence. Ve výjimečných případech se však může stát, že datová sada bude obsahovat autorsko-právně chráněná díla. Příkladem datové sady obsahující autorsko-právně chráněná díla by mohl být seznam příspěvků literární soutěže pořádané radnicí, který by byl uveřejněný včetně plnotextového znění přihlášených soutěžních děl.

**Krok 1** – Zjistěte, zda je součástí datové sady autorské dílo dle § 2 odst. 1 Aut. Z. Pokud není, nebo pokud se jedná o úřední dílo dle § 3 písm. a) Aut. Z., lze v tomto ohledu bez obav obsah databáze zveřejnit.

**Krok 2** – Pokud datová sada autorské dílo obsahuje, musíte zjistit, zda lze toto dílo licencovat. Pokud jste autorem vy (tj. např. orgán veřejné správy), můžete licenci CC-BY 4.0 udělit bez dalšího. Pokud je autorem třetí osoba, musíte zjistit, zda licenční smlouva, která vás opravňuje k užívání díla, obsahuje výslovně uvedenou možnost udělovat podlicence a jak je tato podlicence specifikovaná. Pokud to podlicenční právo dovolí, můžete dílo licencovat licencí CC-BY 4.0. Pokud ne, užití vlastní licenci odpovídající možnostem udělení podlicence.

**Krok 3** – Pokud dílo licencovat nelze, musíte ho z datové sady odstranit.

## 9.7.2 Sekce 2 – Autorskoprávní ochrana databáze jako díla

Tato ochrana je založena § 2 odst. 2 a 5 Autorského zákona a chráněny jsou takové databáze, které jsou „způsobem výběru nebo uspořádáním obsahu autorovým vlastním duševním výtvořem“. Chráněna je tady originální struktura, systém řazení údajů a podobně, tedy kostra, do které se samotné údaje vkládají. Takové databázi se říká „kreativní databáze“. Příkladem kreativní databáze může být obchodní rejstřík, který je však zároveň úředním dílem.

**Krok 1** – Zjistěte, jestli datová sada, kterou publikujete, naplňuje znaky kreativní databáze. Pokud tyto znaky nenaplňuje, můžete ji zveřejnit.

**Krok 2** – Pokud naplňuje, ověřte, zda existuje veřejný zájem na vyloučení autorskoprávní ochrany dle § 3 písm. a) autorského zákona, tak aby databáze byla úředním dílem. Ve většině případů tento zájem existovat bude.

**Krok 3** – Pokud je databáze úředním dílem, nebrání vám autorskoprávní ochrana kreativní databáze v jejím zveřejnění v režimu otevřených dat.

**Krok 4** – Pokud kreativní databáze není úředním dílem, použijte licenční podmínky CC-BY 4.0. Licenční podmínky použijte tak, že k datové sadě připojíte: „<Název datové sady> je jako kreativní databáze autorským dílem <Název organizace nebo jméno autora> a je zveřejněn pod licencí CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). Datovou sadu je možné použít za splnění licenčních podmínek.“

## 9.7.3 Sekce 3 - zvláštní databázová práva pořizovatele databáze

### databáze

Tato ochrana je založena § 88 a následujícími Autorského zákona. Zvláštní databázová práva chrání investice vynaložené při zřízení databáze, ověření uvedených dat, nebo jejich publikaci. Práva vznikají, pokud pořízení, ověření nebo předvedení obsahu databáze představuje kvalitativně nebo kvantitativně podstatný vklad. Držitelem práv je pořizovatel databáze, což je osoba, která na svou odpovědnost pořídí databázi, nebo pro kterou tak

z jejího podnětu učiní jiná osoba. Zvláštních databázových práv se dá plně vzdát, na rozdíl od autorských práv. Příkladem databáze chráněné zvláštními právy pořizovatele databáze je databáze právních předpisů.

**Krok 1** – Ověřte, zda k datové sadě vznikla zvláštní databázová práva pořizovatele databáze. Pokud byl učiněn při vkládání dat, jejich validaci, nebo publikaci kvantitativně nebo kvalitativně podstatný vklad, tak práva vznikla. Tento vklad, obvykle vyjádřen formou finanční investice, musí být podstatný, zdaleka ne všechny databáze na tuto ochranu dosáhnou.

**Krok 2** – Pokud je takový vklad přítomný a zvláštní databázová práva pořizovatele databáze vznikla, použijte licenci CC0, čímž se těchto práv vzdáte. Licenci udělíte tak, že k datové sadě připojíte: “<Název organizace> zveřejňuje datovou sadu <název datové sady> pod licencí CC0 (<https://creativecommons.org/publicdomain/zero/1.0/>) a tím se vzdává zvláštních práv pořizovatele databáze k tomuto datasetu.”

**Krok 3** - Pokud taková zvláštní databázová práva nevznikla, můžete datovou sadu volně publikovat.

### 9.7.4 Závěr

Pokud jste nepoužili žádný z nástrojů CC-BY 4.0, nebo CC0, připojte k datové sadě *Informaci o volném přístupu k datové sadě* (<https://portal.gov.cz/portal/ostatni/volny-pristup-k-ds.html>).

Je třeba nezapomenout, že může nastat situace, kdy dojde ke kombinaci výše uvedených ochran - např. datová sada bude chráněná jak jako kreativní databáze, tak zvláštními databázovými právy. V takovém případě je nezbytné ošetřit všechna práva. Vhodným prostředkem je kombinace výše uvedených licenčních oprávnění (např.: “<Název datové sady> je jako kreativní databáze autorským dílem <Název organizace nebo jméno autora> a je zveřejněna pod licencí CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). Datovou sadu je možné použít za splnění licenčních podmínek. Zároveň <Název organizace> zveřejňuje datovou sadu <název datové sady> pod licencí CC0 (<https://creativecommons.org/publicdomain/zero/1.0/>) a tím se vzdává zvláštních práv pořizovatele databáze k této datové sadě.”)

### 9.7.5 Přehledová tabulka

Datová sada:	Licenční podmínky:	Text při zveřejnění:
Obsahuje autorské dílo jako součást datasetu	Pokud mohu, udělte licenci dle podlicenčního oprávnění	“Dílo <Jméno díla> autora <jméno autora>, které je součástí datové sady <název datové sady> je zveřejněno za podmínek licence <uvést licenci - dle oprávnění>.”
Je kreativní databází	Pokud není úředním dílem (většinou je), CC-BY 4.0	“<Název datové sady> je jako kreativní databáze autorským dílem <Název organizace nebo jméno autora> a je zveřejněn pod licencí CC BY 4.0 ( <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a> ). Datovou sadu je možné použít za splnění licenčních podmínek.”
Je chráněna zvláštními databázovými právy	CC0	“<Název organizace> zveřejňuje datovou sadu <název datové sady> pod licencí CC0 ( <a href="https://creativecommons.org/publicdomain/zero/1.0/">https://creativecommons.org/publicdomain/zero/1.0/</a> ) a tím se vzdává zvláštních práv pořizovatele databáze

		k této datové sadě.”
Neobsahuje chráněná díla, ani není chráněna jako kreativní databáze, ani zvláštními databázovými právy nebo jde o úřední díla	Připojte <i>Informaci o volném přístupu k datové sadě</i>	<a href="https://portal.gov.cz/portal/ostatni/volny-pristup-k-ds.html">Informace o volném přístupu k datové sadě</a> (https://portal.gov.cz/portal/ostatni/volny-pristup-k-ds.html)

## 9.8 Sestavení harmonogramu publikace datových sad

Seznam datových sad určených ke zveřejnění obsahuje několik položek. Jejich zveřejnění může být poměrně pracné a není tak doporučováno zveřejňovat vše najednou. Harmonogram určuje priority a informuje veřejnost o tom, kdy může očekávat zveřejnění jakých datových sad. Harmonogram sestavuje Koordinátor otevírání dat otevírání ve spolupráci s Kurátory dat.

Při sestavení harmonogramu doporučujeme postupovat následovně:

- 1) Kurátoři dat odhadnou pracnost otevření datových sad, které vybrali k otevření.
  - a) Doporučujeme pásmový odhad pracnosti, např. nízká pracnost (do 1 pracovního dne), střední pracnost (do 1 pracovního týdne), vysoká pracnost (více než 1 pracovní týden)
- 2) Koordinátor otevírání dat setřídí seznam datových sad dle jejich priority:
  - a) Prioritu určí s ohledem na pracnost a na výsledky analýzy přínosů a rizik otevření datové sady;
  - b) Nejvyšší prioritu mají datové sady s nízkou pracností, nejvyššími přínosy a nejnižšími riziky;
  - c) Koordinátor otevírání dat může s vedením komunikovat, jaké jsou preferované přínosy a nejobávanější rizika.
- 3) Koordinátor otevírání dat každé datové sadě přiřadí s ohledem na priority datum, ve kterém bude otevřena.
  - a) Postačí, když si zvolí 2 data (např. za 1 měsíc a za 6 měsíců) a každé datové sadě dá buď jedno z těchto dat, nebo uvede obecnější informaci, že ke zveřejnění dojde později (u těch datových sad, u kterých nelze v současnosti určit přesné datum).

## 9.9 Návrh datových sad ke zveřejnění

Koordinátor otevírání dat nezná věcný obsah jednotlivých datových sad, a proto není často schopen kvalifikovaně rozhodnout, jaké datové sady by měly být zahrnuty do publikačního plánu. Věcnou znalost a odpovědnost za datové sady nesou Kurátoři dat. Ti by tedy měli i rozhodovat o umístění datových sad do publikačního plánu. Kurátor dat navrhuje takové datové sady, za které v rámci svojí běžné pracovní činnosti nese věcnou zodpovědnost (tzn., vytváří je, koordinuje jejich sběr apod.).

Doporučovaný postup:

1. Od Koordinátora otevřených dat obdrží Kurátor dat seznam doporučených datových sad k otevření.
2. Kurátor dat vybere v seznamu datové sady, které spadají do jeho kompetence.
3. Kurátor dat z nich vybere ty datové sady, které je možné zařadit do publikačního plánu a zveřejnit v podobě otevřených dat (i když v tom vidí rizika).

4. Kurátor dat analyzuje, zda nespravuje další datové sady, které by bylo možné publikovat v podobě otevřených dat, a přitom nejsou v seznamu doporučených datových sad uvedeny. Nově identifikované datové sady umístí na seznam.
5. Kurátor dat pro každou vybranou datovou sadu doplní následující atributy (většinu lze převzít ze vzorového publikačního plánu):
  - a. název datové sady,
  - b. popis datové sady,
  - c. Kurátor dat.
6. Kurátor dat předá seznam datových sad Koordinátorovi otevírání dat a pomáhá mu s analýzou přínosů a rizik otevření jednotlivých datových sad, pokud o to Koordinátor otevírání dat požádá.

Při návrhu datových sad může Kurátor dat požádat o spolupráci IT specialistu. IT specialista by měl být schopen popsat, jaké informační systémy a databáze jsou pro agendu spravovanou Kurátorem dat provozovány a jaká data jsou v nich uchovávána. Tuto informaci by měl Kurátor dat zvážit a analyzovat, zda se nejedná o další datové sady, které by bylo možné otevřít.

## 9.10 Spolupráce na návrhu datových sad k publikaci

IT specialista může být požádán Kurátory dat o spolupráci při návrhu datových sad k publikaci. Měl by jim popsat data, která jsou vedena v informačních systémech a databázích provozovaných pro potřeby agend vykonávaných Kurátory dat. Kurátoři dat použijí poskytnutou informaci při analýze datových sad, které by bylo možné otevřít.

## 9.11 Schválení publikačního plánu

Po obsazení všech rolí Koordinátor otevírání dat vypracovává s jednotlivými Kurátory dat publikační plán organizace. Po jeho vypracování musí Vedení poskytovatele dat plán schválit. Celá organizace pak bude podle schváleného publikačního plánu postupovat při otevírání dat. Plán popisuje, jaké datové sady budou organizací otevřeny a v jakém harmonogramu. Vedení poskytovatele dat by mělo schválit datové sady navržené v plánu i harmonogram jejich publikace. Publikační plán je aktualizován v roční periodě.

Pro každou datovou sadu jsou zejména uvedeny:

- název datové sady;
- popis datové sady;
- Kurátor dat zodpovědný za analýzu a přípravu datové sady;
- přínosy a rizika publikace datové sady ve formátu otevřených dat;
- periodicita aktualizace datové sady;
- podmínky užití datové sady;
- plánovaný termín první publikace datové sady.

Pro různé typy poskytovatelů dat byly připraveny vzorové publikační plány, které významně urychlí a usnadní proces otevírání dat. Vzorový publikační plán předepisuje doporučené datové sady k otevření a pro ně nabízí předpřipravené katalogizační záznamy a datová schémata.

- Vzorový publikační plán pro ministerstva
- Vzorový publikační plán pro centrální orgány
- Vzorový publikační plán pro kraje
- Vzorový publikační plán pro obce s rozšířenou působností
- Vzorový publikační plán pro ostatní obce

## 9.12 Rozhodnutí o způsobu katalogizace otevřených dat



Zásadní vlastností otevřených dat je, že jsou veřejností snadno dohledatelná. Katalog otevřených dat je prostředek, který dohledatelnost významně usnadňuje. Je důležité, aby všechny datové sady poskytovatele byly katalogizovány v alespoň jednom datovém katalogu.

Povinná pro každého Poskytovatele dat je katalogizace v [Národním katalogu otevřených dat \(NKOD\)](#), který provozuje Ministerstvo vnitra České republiky na Portálu veřejné správy (PVS).

Poskytovatel může dále zvolit některou z následujících možností:

- Lokální katalog otevřených dat: Poskytovatel může na svých WWW stránkách provozovat svůj vlastní lokální katalog, ve kterém bude katalogizovat svoje datové sady. Pokud navíc lokální katalog splňuje technické parametry popsané v těchto standardech, nemusí poskytovatel předávat záznamy do NKOD ručně. Je možné zajistit automatickou synchronizaci obsahu lokálního katalogu s NKOD.
- Katalog otevřených dat jako služba: Existují katalogy otevřených dat, které nabízejí poskytovatelům možnost katalogizovat jejich datové sady. Služby bývají poskytovány zdarma. Takovou službou je např. [Datahub.io](#).
- Jak spolupracovat s IT specialistou na určení způsobu katalogizace naleznete v kapitole „Ministerstvo má možnost zřídit subdoménu (např. data.ministerstvo.cz) a na ní zprovoznit technologie PHP a MySQL.“
- Ke katalogizaci datových sad volí možnost provozovat vlastní lokální datový katalog postavený na technologii DKAN.

Konzultace s IT specialistou pro určení způsobu katalogizace“.

Při rozhodování o způsobu katalogizace doporučujeme spolupracovat s IT specialistou a postupovat podle následující praxe:

- Pokud nechcete provozovat žádnou vlastní technologii pro katalogizaci datových otevřených datových sad, zvolte možnost katalogizovat datové sady přímo v Národním katalogu otevřených dat (NKOD).
  - Postup katalogizace datové sady v Národním katalogu otevřených dat (NKOD) viz Kapitola „Národní katalog otevřených dat (NKOD)“.
- Pokud Vám nepostačuje způsob katalogizace pouze v NKOD, můžete provozovat vlastní lokální katalog otevřených dat. Při volbě způsobu provozu lokálního katalogu postupujte následovně:
  - Na svém serveru již provozujete nebo můžete provozovat technologie Linux, Python, Java, PostgreSQL a nebo chcete záznamy o datových sadách automaticky aktualizovat pomocí jiného software ⇒ Provozujte vlastní lokální katalog na zřízené doméně postavený na technologii CKAN (viz kapitola „CKAN a jeho rozšíření“).
  - Na svém serveru již provozujete nebo můžete provozovat technologie PHP a MySQL a stačí vám spravovat záznamy o datových sadách ručně ⇒ Provozujte vlastní lokální katalog na zřízené doméně postavený na technologii DKAN (viz kapitola „DKAN a jeho rozšíření“).
  - Nemůžete nebo nechcete zasahovat do svých webových serverů, ale můžete si pořídit levný externí webhosting s podporou PHP a MySQL a stačí vám spravovat záznamy o datových sadách ručně ⇒ Provozujte vlastní lokální katalog na zřízené doméně na externím webhostingu na technologii DKAN (viz kapitola „DKAN a jeho rozšíření“).
  - Alternativou k vlastnímu lokálnímu katalogu je využití některé z volně dostupných služeb pro katalogizaci datových sad, např. Datahub.io (viz kapitola „Služba Datahub.io“).



- Je však nutné upozornit, že služby jako Datahub.io podporují pouze omezenou strukturu katalogizačního záznamu. Důležitá metadata o svých datových sadách tak nebudete mít možnost vyplnit.

Pokud budete provozovat vlastní lokální katalog otevřených dat (na bázi technologií DKAN/CKAN) nebo využíváte ke katalogizaci službu DataHub.io či podobnou, nemusíte navíc ručně katalogizovat jednotlivé datové sady v NKOD. Postačí v NKOD svůj katalog zaregistrovat (viz kapitola „Registrace vlastního katalogu v NKOD“).

Pro rozhodování můžete použít následující modelové situace:

### 9.12.1 Modelová situace 1: Malá obec s jednoduchou statickou webovou stránkou bez možnosti jakýchkoliv úprav

- Obec provozuje jednoduchou WWW stránku se základními informacemi o obci.
- Nemá možnost ani prostředky nijak WWW stránku rozšiřovat.
- Nemá možnost na WWW stránce publikovat datové soubory ke stažení.
- Nechce pracovat v cizojazyčném prostředí a chce garanci, že zvolený způsob katalogizace bude fungovat a nebude měněn.
- Ke katalogizaci datových sad volí možnost katalogizovat datové sady pomocí NKOD.

### 9.12.2 Modelová situace 2: Ministerstvo

- Ministerstvo má možnost zřídit subdoménu (např. data.ministerstvo.cz) a na ní zprovoznit technologie PHP a MySQL.
- Ke katalogizaci datových sad volí možnost provozovat vlastní lokální datový katalog postavený na technologii DKAN.

## 9.13 Konzultace s IT specialistou pro určení způsobu katalogizace

IT specialista je zodpovědný za technické prostředky Poskytovatele. Musí tedy poskytnout informace o stávajících technických možnostech Poskytovatele a možnostech jejich rozšíření vzhledem k záměru provozovat lokální katalog.

IT specialista musí být schopen podat informace o možnostech:

- zřízení subdomén pod doménou poskytovatele (poskytovatel.cz)
- automatizovaného exportu dat z existujících systému indikujících potřebu zapisovatelného katalogového API pro tvorbu katalogizačních záznamů (CKAN)
- instalace a správy katalogu CKAN u Poskytovatele
- instalace a správy katalogu DKAN u Poskytovatele
- instalace a správy katalogu DKAN u externího webhostingu
- vložení NKOD widgetu do webových stránek Poskytovatele

## 9.14 Spolupráce IT specialisty na určení způsobu katalogizace datových sad

IT specialista musí dodat Koordinátorovi otevírání dat informace o technických prostředcích poskytovatele a jejich využitelnosti a rozšiřitelnosti pro katalogizaci datových sad. Tyto informace pomohou Koordinátorovi otevírání dat (viz kapitola „Rozhodnutí o způsobu katalogizace otevřených dat“).

Informace zahrnují možnost a nákladnost:

- Zřízení subdomény data.poskytovatel.cz, na které bude případný lokální katalog;
- Provozu a správy vlastního (virtuálního) serveru s OS Linux a technologiemi Python, Java, PostgreSQL pro provoz lokálního katalogu CKAN;
- Provozu a správy vlastního webového serveru s technologiemi PHP a MySQL pro provoz lokálního katalogu DKAN;
- Využití případného existujícího webového serveru Poskytovatele dat pro provoz lokálního katalogu DKAN;
- Zřízení externího webhostingu s technologiemi PHP a MySQL pro provoz lokálního katalogu DKAN;

Dále IT specialista musí informovat Koordinátora otevírání dat o případných existujících systémech automaticky produkujících data k otevření. Pokud se totiž budou tato data připravovat automatizovaně, vzniká potřeba automatizované správy katalogizačních záznamů přes API katalogu. Takovou správu umožňuje jen CKAN, DKAN má API pouze pro čtení.

## 9.15 Zveřejnění publikačního plánu

Poté, co Vedení poskytovatele dat připravený publikační plán schválí, může jej Koordinátor otevírání dat zveřejnit. Pokud Poskytovatel dat nepovažuje za vhodné zveřejňovat identifikovaná rizika a přínosy, není nutné zveřejňovat kompletní publikační plán. Postačí zveřejnit tabulku se seznamem datových sad určených k otevření, kdy pro každou datovou sadu budou uvedeny základní informace (název, popis, Kurátor dat) a plánované datum otevření. Zveřejnění publikačního plánu je nepovinným krokem. Je třeba říci, že jeho zveřejnění znamená určitý závazek orgánu veřejné správy vůči veřejnosti.

## 9.16 Periodické vyhodnocování plnění publikačního plánu a informování vedení

Koordinátor otevírání dat periodicky vyhodnocuje plnění publikačního plánu (doporučujeme jednou ročně). Při vyhodnocení by se měl věnovat především následujícímu:

- Podařilo se otevřít všechny datové sady dle předepsaného harmonogramu?
- Pokud se některé datové sady otevřít nepodařilo, proč?
- Podařilo se otevřít některé datové sady nad rámec předepsaného harmonogramu?
- Existují uživatelé již otevřených datových sad?
- Vznikly nad otevřenými datovými sadami nějaké aplikace?
- Jaké byly náklady na otevření datových sad?

Vyhodnocení lze zdokumentovat ve Výroční zprávě.

S vyhodnocením mohou Koordinátorovi otevírání dat pomoci statistiky přístupů k otevřeným datovým sadám.

*Jak vést statistiku přístupů uživatelů k otevřeným datovým sadám (viz kapitola „Vedení statistiky přístupů k otevřeným datovým sadám“)?*

Vyhodnocení nákladů na otevření datových sad je důležitou součástí celkového vyhodnocení. Do procesu otevírání dat je zapojena řada pracovníků Poskytovatele dat, kteří otevíráním dat strávili nemalou část svého času.

Koordinátor otevírání dat předává výsledky vyhodnocení nákladů Vedení poskytovatele dat pro informaci.

Vyhodnocením proces otevírání dat nekončí. Koordinátor otevírání dat musí dále zajišťovat, že již zveřejněné datové sady jsou nadále aktualizovány v určených periodách. Může také dále rozšiřovat publikační plán o nové datové sady a zajistit jejich otevření.

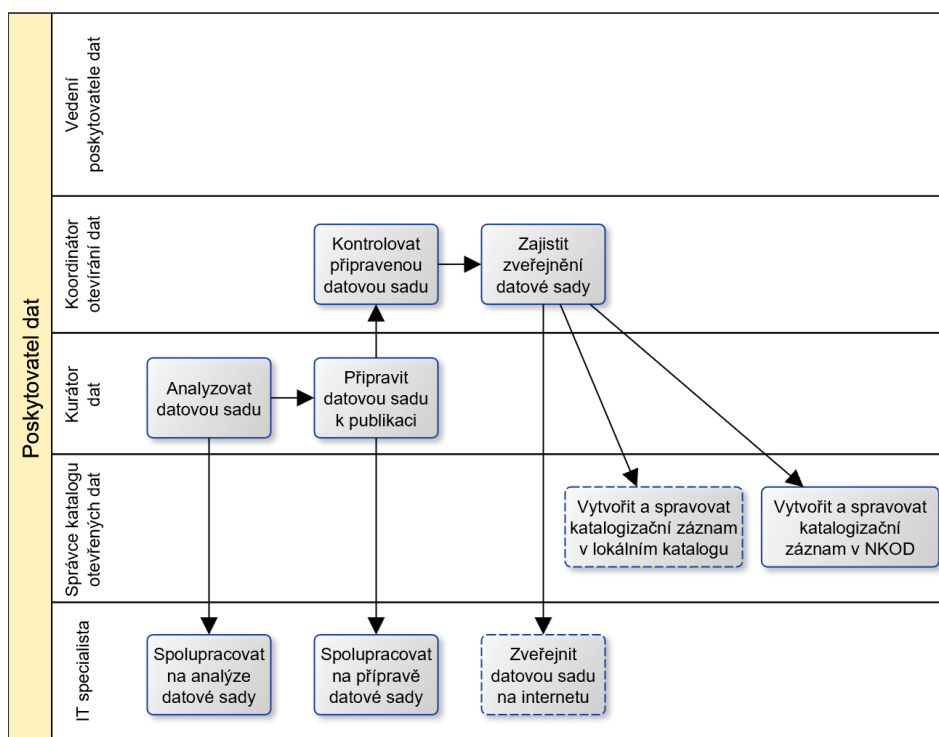
## 9.17 Vedení statistiky přístupů k otevřeným datovým sadám

Pro sběr statistik o přístupech k otevřeným datovým sadám lze snadno využít volně dostupné služby pro sledování přístupu k webovým stránkám, jako například [Google Analytics](#), [Piwik](#), [KISSMetrics](#), [Clicky](#) a další. Obvykle je potřeba se u vybrané služby zaregistrovat a vložit její sledovací JavaScript kód do každé webové stránky.

## 10 Publikace datové sady v otevřené podobě

Po vytvoření publikačního plánu je nutné postupně jednotlivé datové sady publikovat dle stanoveného harmonogramu. Postup publikace datové sady v otevřené podobě v sobě zahrnuje její analýzu a návrh technické podoby její publikace, přípravu datové sady do navržené podoby, její publikaci a katalogizaci.

Jednotlivé činnosti jsou zachyceny v následujícím diagramu. Po kliknutí na činnost se dostanete na detailní popis její doporučené náplně. Jednotlivé plavecké dráhy odpovídají rolím, které by se měly procesu otevírání dat účastnit. Po kliknutí na roli se dostanete na její popis.



Legenda:

- Obdélník s plným okrajem značí povinnou činnost.
- Obdélník s přerušovaným okrajem značí nepovinnou činnost.
- Šipky značí návaznost mezi činnostmi.

### 10.1 Analýza datové sady a návrh technického způsobu jejího zveřejnění

Před samotnou publikací datové sady je nutné ji analyzovat. Analýzu zajišťuje její Kurátor dat. Analyzuje obsah a strukturu datové sady. Na základě analýzy pak (ve spolupráci s IT specialistou) navrhuje:

- stupeň otevřenosti, na kterém bude datová sada zveřejněna
- datové schéma určující strukturu, ve které bude datová sada zveřejněna
- katalogizační záznam datové sady

Pokud se jedná o datovou sadu uvedenou ve vzorovém publikačním plánu, je pro ni také k dispozici již předpřipraven navržený stupeň otevřenosti, datové schéma i katalogizační záznam.

Při provádění vlastní analýzy by měl Kurátor dat nejprve určit stupeň otevřenosti datové sady. Minimální stupeň otevřenosti je stupeň 3.

*Jak určit stupeň otevřenosti datové sady (viz kapitola „Určení stupně otevřenosti datové sady“)?*

Dále je třeba pro datovou sadu navrhnout její strojově čitelné datové schéma. Datové schéma popisuje strukturu jednotlivých položek datové sady. Jeho tvorba je technickou záležitostí. Pokud nemá Kurátor dat dostatečné technické znalosti, musí tuto činnost převzít určený IT specialista. Přístup k tvorbě schématu datové sady se liší dle zvoleného stupně otevřenosti.

*Jak vytvořit datové schéma (viz kapitola „Vytvoření datového schématu datové sady“)?*

V posledním kroku analýzy připraví Kurátor katalogizační záznam datové sady, který bude při jejím zveřejnění publikován v datovém katalogu.

*Jak připravit katalogizační záznam datové sady (viz kapitola „Příprava katalogizačního záznamu datové sady“)?*

## 10.2 Určení stupně otevřenosti datové sady

Pro každou datovou sadu je třeba rozhodnout, na jakém stupni otevřenosti bude otevřena. Pro tento účel je definováno 5 stupňů otevřenosti.

V kontextu veřejné správy České republiky je nutno pro každou datovou sadu, kterou chcete publikovat v podobě otevřených dat, zvolit stupeň otevřenosti alespoň 3. Stupně otevřenosti 4 a 5 s sebou nesou větší pracnost při přípravě datové sady k publikaci. Je možno volit i více stupňů najednou (včetně 1 či 2), platí však, že alespoň jeden ze stupňů musí být stupeň alespoň 3.

V souvislosti s určením stupně otevřenosti je nutno zvolit i technickou podobu zveřejnění datové sady. To zahrnuje určit, v jakých distribucích bude obsah datové sady zveřejněn a v jakém datovém formátu budou distribuce vyjádřeny. Postup určení technické podoby se liší v závislosti na zvoleném stupni otevřenosti datové sady (technické standardy viz kapitola „Technické standardy“).

## 10.3 Vytvoření datového schématu datové sady

Datové schéma určuje, jak jsou strukturovány jednotlivé záznamy v datové sadě. Struktura datové sady by měla být vyjádřena v podobě strojově čitelného datového schématu. Je to běžnou praxí ve světě databází a výměny dat mezi softwarovými systémy. Protože se řada typů datových sad často opakuje napříč různými orgány veřejné správy (OVS), má navíc smysl datová schémata pro tyto typy datových sad předpřipravit a nabídnout všem OVS k využití. To výrazně zvyšuje potenciál otevřených dat. Pokud někdo vytvoří softwarovou aplikaci, která rozumí danému datovému schématu, bude z toho profitovat každá organizace, která dané datové schéma použije pro svoji datovou sadu, neboť s ní bude aplikace fungovat.



### 10.3.1 Datová schémata doporučených datových sad

Pro doporučené datové sady ze vzorových publikačních plánů jsou již jejich datová schémata předpřipravena a můžete je využít v případě, že se rozhodnete pro publikaci na stupni otevřenosti 3. Předpřipravené datové schéma je doporučeno, nikoliv striktně povinné. Snažte se však strukturou svojí datové sady co nejvíce přiblížit struktuře definované předpřipraveným datovým schématem.

- Nepřejmenovávávejte ani jinak nepředdefinovávejte datové prvky definované předpřipraveným datovým schématem.
- Pokud některý datový prvek ve své datové sadě evidujete, ale s jiným názvem, doporučujeme jej ve vaší datové sadě přejmenovat na název definovaný předpřipraveným datovým schématem.
- Pokud některý datový prvek ve své datové sadě neevidujete, nemusíte jej doplňovat.
- Pokud některý datový prvek ve své datové sadě evidujete, ale strukturuje jej jinak, doporučujeme jej ve vaší datové sadě přestrukturovat. Pokud však na to nemáte kapacity, upravte datové schéma tak, aby odpovídalo vaší struktuře.

Pokud využijete předpřipravené datové schéma beze změny, nemusíte jej kopírovat. V katalogizačním záznamu datové sady, kde musí být datové schéma uvedeno, se budete na předpřipravené datové schéma pouze odkazovat. Pokud předpřipravené datové schéma potřebujete jakkoliv změnit, musíte si vytvořit vlastní kopii a tu upravit požadovaným způsobem. Vytvořené datové schéma budete muset při publikaci datové sady také publikovat a odkazovat se na něj.

### 10.3.2 Vytvoření nového datového schématu

Postup přípravy datového schématu se liší v závislosti na zvoleném stupni otevřenosti. Technické standardy viz kapitola „Technické standardy“.

## 10.4 Příprava katalogizačního záznamu datové sady

Katalogizace datové sady je důležitá pro její potenciální uživatele. Umožní jim datovou sadu nalézt. Katalogizační záznam může být zveřejněn v různých katalogích. Tím zajistíte, že se informace o existenci datové sady dostane k co nejširšímu okruhu uživatelů.

Struktura katalogizačního záznamu odpovídá evropskému standardu DCAT-AP. Ten rozlišuje datovou sadu a distribuci datové sady. Datová sada je tvořena jednou či více distribucemi. Distribuce obsahuje samotná data. Může to být datový soubor určený ke stažení nebo to může být datové API, které umožňuje přímé dotazování do obsahu datové sady.

Katalogizační záznam se dělí na 2 části:

1. První část obsahuje metadata o samotné datové sadě. Jedná se např. o název a popis datové sady.
2. Druhá část obsahuje metadata o distribucích datové sady (datových zdrojích). Jedná se např. o specifikaci datového formátu datového zdroje či o jeho datové schéma.

### 10.4.1 Katalogizační záznamy doporučených datových sad

Pro doporučené datové sady ze vzorového publikačního plánu jsou již katalogizační záznamy předpřipraveny. Můžete je tedy použít. Konkrétně jsou předvyplněné hodnoty těch atributů, které jsou nezávislé na konkrétní organizaci:

- název datové sady (můžete doplnit název organizace či jinak upravit),
- popis datové sady (můžete doplnit název organizace či jinak upravit),



- klíčová slova (můžete doplnit vlastními klíčovými slovy),
- [EUROVOC](#) koncepty (můžete doplnit dalšími EUROVOC koncepty),
- datový formát,
- odkaz na dokumentaci (můžete nahradit vlastní dokumentací, k jejímu vytvoření můžete využít tuto předpřipravenou),
- odkaz na strojem čitelné datové schéma (můžete jej nahradit vlastním datovým schématem).

Hodnoty atributů, které nejsou předvyplněné je nutno doplnit.

## 10.4.2 Vyplňování hodnot atributů katalogizačního záznamu

Při vyplňování hodnot atributů dodržujte následující pravidla.

### *Atributy datové sady*

- *název datové sady* (povinný atribut)
  - Vyplňte krátký název datové sady, ze kterého je zřejmé, co je obsahem datové sady.
  - V názvu je vhodné použít i název vaší organizace.
  - Název by měl být optimálně max. v rozsahu jednoho řádku textu.
- *popis datové sady* (povinný atribut)
  - Vyplňte delší text, který detailněji popisuje, co je obsahem datové sady.
  - Pokud je to pro správné pochopení nutné, uveďte i o čem data nevyprávějí.
  - Popis by měl být optimálně max. v rozsahu jednoho odstavce do 10 řádků textu.
- *periodicita aktualizace* (povinný atribut)
  - Vyberte, jak často budete aktualizovat datové zdroje (přidávat nové či aktualizovat obsah dřívějších distribucí).
  - Např. pokud vyberete týdenní periodicitu, pak to znamená, že každý týden přidáváte nový datový zdroj nebo aktualizujete obsah dřívějšího datového zdroje.
  - Periodicita aktualizace se vyplňuje jako [kód dle standardu ISO 8601](#).  
Příklady:
    - Denně = R/P1D
    - Týdně = R/P1W
    - Měsíčně = R/P1M
    - Čtvrtletně = R/P3M
    - Pololetně = R/P6M
    - Ročně = R/P1Y
- *související geografické území* (povinný atribut)
  - Vyberte geografické území, kterého se datová sada týká. Může to být např.:
    - území, které by datová sada zabírala, pokud by byla zobrazena na mapě,
    - správní území orgánu veřejné správy, který datovou sadu publikuje, pokud její zobrazení na mapě nemá smysl.
  - K identifikaci geografického území používáme identifikátory RÚIAN, např.:
    - Datová sada s veřejnými parkovišti na území obce má jako související geografickou oblast přiřazenu obec (pokud bychom zobrazili parkoviště na mapě, rozprostírala by se na území obce);
    - Datová sada s rozpočtem krajského úřadu má jako související geografickou oblast přiřazeno území vyššího samosprávného celku, tj. kraje (rozpočet nemá smysl zobrazovat na mapě, volíme tedy správní území kraje).
  - K identifikaci se používá se dvojice hodnot:
    - typ prvku RÚIAN ([číselník typů prvku RÚIAN](#))

- kód prvku RÚIAN - pro vyhledání prvku lze využít [Veřejný dálkový přístup k datům RÚIAN](#)
- Příklady:
  - Česká republika
    - Typ prvku: ST
    - Kód prvku: 1
  - Praha:
    - Typ prvku: OB
    - Kód prvku: 554782
- *klíčová slova* (povinný atribut)
  - vyplňte jedno nebo více libovolných klíčových slov, které charakterizují obsah datové sady,
- *klasifikace dle EUROVOC* (povinný atribut)
  - Vyberte koncepty z taxonomie EUROVOC, což je standardní slovník pojmů EU přeložený do všech jazyků EU;
  - Vyberte alespoň jeden koncept, který charakterizuje obsah datové sady.

Jak vyhledat správné EUROVOC koncepty (viz kapitola „Vyhledání EUROVOC konceptu“)?

- *jméno Kurátora dat* (doporučený atribut)
  - vyplňte jméno konkrétní osoby nebo název útvaru organizace, který je Kurátorem dat pro danou datovou sadu,
- *email Kurátora dat* (doporučený atribut)
  - vyplňte kontaktní email Kurátora dat,
  - může být jednotný pro celou organizaci,
- *odkaz na dokumentaci datové sady* (doporučený atribut)
  - zadejte odkaz na dokument určený pro člověka, který popisuje obsah datové sady a vysvětluje jeho správné použití/interpretaci,
- *dotčené časové období* (doporučený atribut)
  - vyplňte časové období (datum od, případně do), které pokrývá obsah datové sady,
  - jako formát pro reprezentaci data použijte datový typ používaný v jazyce XML Schema [xsd:date](#).
  - vyplňujte pouze, pokud má smysl, např.:
    - Datová sada s demografickými údaji za posledních 5 let v jednotlivých letech může mít nastaven „datum od“ (1.1. prvního roku) a „datum do“ (31.12. posledního roku), konkrétně například 2010-01-01 a 2014-12-31
    - Datová sada s odtaženými auty může mít nastaven jako „datum od“ datum, od kdy je evidence vedena, „datum do“ nastaven nemá, protože se jedná o aktuální datum a evidence je průběžně aktualizována.
    - Datová sada s aktuální polohou dopravních spojů nemá nastaven „datum od“ ani „datum do“, neboť je zveřejňován vždy aktuální stav.

#### *Atributy datového zdroje*

- odkaz na stažení datového zdroje (povinný atribut)
  - vyplňte odkaz na datový soubor ke stažení nebo odkaz na datové API
- formát datového zdroje (povinný atribut)
  - vyplňte MIME typ popisující formát datového zdroje
- odkaz na podmínky užití (povinný atribut)
  - vyplňte odkaz na webovou stránku uvádějící podmínky užití či licenci upravující užití datové sady,
- název distribuce (doporučený atribut)
  - vyplňte název datového souboru ke stažení nebo název API

- odkaz na stránku o distribuci (doporučený atribut)
  - vyplňte odkaz na webovou stránku, která informuje o existenci distribuce a na které je případně umístěn i odkaz ke stažení


*Jak vybrat MIME typ (viz kapitola „Výběr MIME typu“)?*


- odkaz na strojově čitelné datové schéma datového zdroje (doporučený atribut)
  - vyplňte odkaz na strojově čitelné datové schéma popisující strukturu datového zdroje,
- formát datového schématu (doporučený atribut)
  - vyplňte MIME typ popisující formát datového schématu,
- dotčené časové období (doporučený atribut)
  - vyplňte časové období (datum od, případně do), které pokrývá obsah dané distribuce,
  - platí stejná pravidla jako u dotčeného časového období datové sady,
  - použijte se, pokud jednotlivé datové zdroje pokrývají různá časová období,
  - příklad: datová sada Rozpočet má datové zdroje za jednotlivé roky. Pak bude u každého datového zdroje uvedeno datum začátku a konce roku, kterého se týká.

## 10.5 Vyhledání EUROVOC konceptu

[Tezaurus EUROVOC](#) obsahuje koncepty, jejichž názvy (pojmy, termíny) jsou udržovány ve všech jazycích Evropské unie, a je proto vhodný pro klasifikaci datových sad. Pro identifikaci EUROVOC konceptu se používá jeho URI. Pojmy (termíny) mají také svá URI označující konkrétní pojem v konkrétním jazyce, datové sady se ale klasifikují pomocí URI konceptu, nikoliv URI pojmu v konkrétním jazyce. Například koncept pro veřejnou správu má URI <http://eurovoc.europa.eu/77> a <http://eurovoc.europa.eu/131636> je URI termínu pro veřejnou správu v češtině, nikoliv URI konceptu. Vhodný EUROVOC koncept [lze vyhledat na stránkách tezauru](#). Měly by být zvoleny takové koncepty, které nejlépe vystihují, o čem datová sada vypovídá. Hledat lze pomocí vyhledávání v levé části stránky:

← → ↻ 🏠

Tato stránka je součástí  **Europa**

 **EuroVoc** Mnohojazyčný tezaurus Evropské unie

Europa > Úvodní stránka EuroVoc > Vyhledávání > veřejná správa

**Jazyk:**  
(cs) Čeština

**Vyhledávání**

- Rozšířené vyhledávání

**Procházet**

- Procházet verzí řazenou podle témat

**Stáhnout**

- Podle oblasti
- Abecední permutovaná verze
- Vícejazyčný seznam
- Abecední rejstřík
- EuroVoc SKOS/RDF
- EuroVoc XML

**Vaše návrhy**

**Výsledek hledání (1 pojmy)**

---

MT 0436 výkonná moc a státní správa

**veřejná správa**  
**veřejná správa**

Po vybrání pojmu se lze navigovat stromovou strukturou konceptů a nalézt tak ten nejlépe odpovídající. URI konceptu je pak uvedeno ve spodní části stránky.

← → ↻ 🏠 [eurovoc.europa.eu/drupal/?q=cs/request&concepturi=http%3A%2F%2Feurovoc.europa.eu%2F77&termu](http://eurovoc.europa.eu/drupal/?q=cs/request&concepturi=http%3A%2F%2Feurovoc.europa.eu%2F77&termu)

Tato stránka je součástí 

 **EuroVoc** Mnohojazyčný tezaurus Evropské unie

Europa > Úvodní stránka EuroVoc > Oblasti B, MT > veřejná správa

**Jazyk:**  
(cs) Čeština

**Vyhledávání**  
  
Rozšířené vyhledávání

**Procházet**  
Procházet verzí řazenou podle témat

**Stáhnout**  
Podle oblastí  
Abecední permutovaná verze  
Vícejazyčný seznam  
Abecední rejstřík  
EuroVoc SKOS/RDF  
EuroVoc XML

**Vaše návrhy**  
Zaslat příspěvek  
Nové schválené pojmy

**veřejná správa**

UF *orgány státní správy státní správa*

**04 POLITICKÝ ŽIVOT**

MT 0436 výkonná moc a státní správa  
NT1 elektronická správa  
NT1 místní orgány státní správy  
NT1 místní poslanec  
NT1 občansko-správní vztahy  
NT2 ombudsman  
NT2 průhlednost administrativy  
NT2 správní formality  
NT2 správní odvolání  
NT1 partnerství mezi veřejným a soukromým sektorem  
NT1 regionální orgány státní správy  
NT1 sídlo správy [4.4]  
NT1 správní dohled  
NT1 ústřední orgány státní správy  
NT2 ministerstvo  
NT1 veřejná instituce  
NT1 veřejná politika  
NT1 veřejná služba

RT *orgány veřejné správy [0406 ]  
právo volné úvahy při výkonu pravomocí [0406 ]  
správní věda [3611]*

**URI** <http://eurovoc.europa.eu/77>

**Has Exact Match**  
*veřejná správa (AGROVOC)  
Veřejná správa (ECLAS)  
public administration (UNBIS)*

**EKVIVALENTY V OSTATNÍCH JAZYCÍCH**

BG	държава администрация
ES	administración pública
<b>CS</b>	<b>veřejná správa</b>
DA	offentlig forvaltning
DE	öffentliche Verwaltung
ET	avalik haldus
EL	δημόσια διοίκηση
EN	public administration
FR	administration publique
HR	javna uprava
IT	pubblica amministrazione
LV	valsts pārvalde
LT	viešasis valdymas
HU	közigazgatás
MT	amministrazzjoni pubblika
NL	overheidsadministratie
PL	administracja publiczna
PT	administração pública
RO	administrație publică
SK	štátna správa
SL	javna uprava
FI	julkisen hallinto
SV	offentlig förvaltning
SR	javna uprava
MK	javna administracija
SQ	administratë publike

## 10.6 Výběr MIME typu

MIME (media) typ souboru popisuje formát souboru standardizovaným způsobem. [Oficiální seznam všech MIME typů](#) spravuje organizace IANA. Při výběru MIME typu souboru je možné se řídit příponou daného souboru (pokud nebyla uměle změněna).

Příklady:

- CSV = text/csv
- XML = application/xml
- JSON = application/json
- RDF (Turtle) = text/turtle
- XLS = application/vnd.ms-excel
- XLSX = application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
- DOCX = vnd.openxmlformats-officedocument.wordprocessingml.document
- ODS = application/vnd.oasis.opendocument.spreadsheet

## 10.7 Spolupráce na analýze datové sady

Analýzu datové sady provádí její Kurátor dat. Nicméně, Kurátor dat není IT specialista a nemusí tak mít potřebné know-how pro všechny kroky, které má v analýze provést. Může proto požádat IT specialistu o pomoc.



IT specialista by měl poskytnout součinnost při určení stupně otevřenosti datové sady dle 5\* schématu. Měl by pomoci Kurátorovi dat zjistit, v jaké podobě je v datová sada uchovávána a jaký stupeň otevřenosti je tedy reálně dosáhnout.

Dále by měl IT specialista spolupracovat s Kurátorem dat na návrhu datového schématu datové sady. Kurátor dat je schopen provést analýzu obsahu datové sady a popsat slovně její strukturu. Vyjádření ve strojově čitelné podobě pomocí k tomu vhodného jazyka (např. XML Schema) ale musí většinou provést IT specialista, který k tomu má potřebné znalosti.

Kurátor dat v závěru analýzy sestavuje katalogizační záznam. Jeho součástí jsou i technické detaily datové sady, jako např. datový formát, MIME typ, odkaz na vytvořené datové schéma apod. S tím by měl také pomoci IT specialista.

## 10.8 Příprava datové sady k publikaci

Po provedení analýzy Kurátor dat připraví datovou sadu k publikaci. To znamená připravit distribuci datové sady (příp. více distribucí, pokud se Kurátor při analýze rozhodl rozdělit obsah datové sady do více distribucí či publikovat obsah datové sady ve více různých formátech). Příprava distribuce znamená buď vytvořit novou distribuci, nebo aktualizovat již jednou publikovanou. Připravenou distribuci Kurátor dat předá Koordinátorovi otevírání dat spolu s připraveným katalogizačním záznamem k formální kontrole a k zajištění zveřejnění.

Při přípravě distribuce je nutno dodržet pravidla zvoleného stupně otevřenosti a datového formátu a také strukturu danou zvoleným či vytvořeným datovým schématem.

Často se stává, data tvořící datovou sadu již existují v datových souborech. Ty však nelze přímo zveřejnit, protože jejich formát odpovídá nižšímu stupni otevřenosti, než byl zvolen pro publikaci.

*Jak převést data odpovídající stupni otevřenosti 1 do vyššího stupně otevřenosti (viz kapitola „Převod dat ze stupně otevřenosti 1 do vyššího stupně otevřenosti“)?*

*Jak převést data odpovídající stupni otevřenosti 2 do vyššího stupně otevřenosti (viz kapitola „Převod dat ze stupně otevřenosti 2 do vyššího stupně otevřenosti“)?*

Distribuci či distribuce datové sady je nutno pravidelně aktualizovat dle nastavené periodicity aktualizace v katalogizačním záznamu. Aktualizace znamená buď aktualizace již jednou publikované distribuce, nebo vytvoření nové distribuce obsahující aktualizaci (tj. nově přidané záznamy, změny v již publikovaných záznamech včetně informace o odstranění existujících záznamů).

*Jak rozhodnout o podobě aktualizací datové sady (dále viz kapitola „Rozhodnutí o podobě aktualizací datové sady“)?*

Přípravu datového souboru lze zajistit ručně nebo ji lze automatizovat. Možnosti automatizace by měl Kurátor dat konzultovat s IT specialistou.

*Jak konzultovat přípravu datové sady s IT specialistou (viz kapitola „Konzultace přípravy datové sady s IT specialistou“)?*

Poté, co je datový soubor připraven, by měl Kurátor dat zkontrolovat, že formát datového souboru je v pořádku a že jeho struktura odpovídá navrženému datovému schématu. Kontrolu nazýváme validace datového souboru. Jedná se o technickou záležitost. Pokud nemá Kurátor dat dostatečné technické znalosti, je třeba, aby validaci provedl IT specialista.

*Jak validovat datový soubor (viz kapitola „Validace datového souboru“)?*

Dále může být potřeba aktualizovat atributy katalogizačního záznamu. Datový soubor je v katalogizačním záznamu reprezentován jako distribuce datové sady. Pokud existuje pro datovou sadu pouze jeden aktualizovaný datový soubor, je třeba aktualizovat atributy příslušné distribuce. Pokud je v každé periodě vytvářen nový datový soubor, je třeba rozšířit katalogizační záznam datové sady o novou distribuci.

## 10.9 Převod dat ze stupně otevřenosti 1 do vyššího stupně otevřenosti

V případě, že data tvořící datovou sadu nejsou k dispozici ve strojově čitelné podobě, je nutné je ručně rekonstruovat. Ujistěte se nejprve, že je tomu skutečně tak. Pokud ano, je nutno zvážit, zda datovou sadu publikovat v otevřené podobě. Ruční rekonstrukce je velmi pracná a je tedy nutno zvážit, zda se náklady rekonstrukce vyrovnají přínosům otevření datové sady. Můžete přistoupit pouze ke zveřejnění datové sady na stupni otevřenosti 1. V tom případě však nebudete zveřejňovat datovou sadu v podobě otevřených dat.

Následující příklady nastiňují, jak se Kurátor dat může ujistit, zda obsah datové sady neexistuje ve strojově čitelné podobě:

- Pokud chcete otevřít datovou sadu tvořenou tabulkou, která je dostupná pouze v PDF souboru z výroční zprávy, pokuste se dohledat zdrojový tabulkový soubor (typicky ve formátu XLS), který byl použit pro tvorbu výroční zprávy. Pokud jej najdete, získali jste reprezentaci datové sady odpovídající stupni otevřenosti 2. Pokud jej nenajdete, je nutno přistoupit k ruční rekonstrukci obsahu tabulky.
- Pokud chcete otevřít datovou sadu tvořenou tabulkou, která je dostupná pouze v podobě HTML stránky, pokuste se zjistit, zda HTML stránka není automaticky generována skriptem z databáze. Pokud ano, IT specialista musí umět zařídit export tabulky do vyššího stupně otevřenosti. Pokud ne, je nutno přistoupit k ruční rekonstrukci obsahu tabulky.
- Pokud chcete zveřejnit kolekci dokumentů, ale máte k dispozici pouze jejich naskenované PDF verze, pokuste se zjistit, zda někde neexistují např. ve formátu DOC(X), ve kterém byly smlouvy původně editovány. Pokud je najdete, získali jste reprezentaci datové sady odpovídající stupni otevřenosti 2. Pokud je nenajdete, je nutno přistoupit k ruční rekonstrukci textu dokumentů.

Pokud jste nuceni přistoupit k ruční rekonstrukci obsahu datové sady, doporučujeme následující postup:

- V případě tabulkové datové sady použijte tabulkový editor (Microsoft Excel apod.), ve kterém vytvoříte novou tabulku, jejíž struktura odpovídá zdrojové tabulce a do ní přepokopírujte (pokud to formát zdrojové tabulky umožňuje) nebo přepište hodnoty ze zdrojové tabulky.
- V případě datové sady tvořené kolekcí dokumentů je nutno jednotlivé dokumenty přepokopírovat (pokud to formát zdrojové kolekce dokumentů umožňuje) či přepsat.

Ruční práci lze urychlit využitím vhodných OCR nástrojů.

## 10.10 Převod dat ze stupně otevřenosti 2 do vyššího otevřenosti

Pokud jsou data tvořící datovou sadu k dispozici ve strojově čitelné podobě, ale pouze v proprietárním formátu, je nutné je převést do vhodného otevřeného formátu.

### 10.10.1 Převod tabulkové datové sady

- Pokud je datová sada reprezentována v podobě souboru tabulkového editoru s jednoduchou strukturou, lze ji přímo v tabulkovém editoru exportovat do CSV formátu anebo lze poměrně snadno vytvořit ve spolupráci s IT specialistou skript, který uložení do CSV formátu provede automaticky. Jednoduchá struktura znamená, že:
  - Tabulka je rozdělena na dvě po sobě následující části: hlavička a datová část;

- Hlavička je tvořena N prvními řádky tabulky a definuje strukturu datové části, ve které každý sloupec odpovídá jednomu typu údajů;
- Datová část je tvořena řádky, jejichž struktura odpovídá struktuře definované hlavičkou. Každá buňka na řádku odpovídá jednomu údaji. Žádné buňky nejsou sloučeny ani přes řádky ani přes sloupce.
- Pokud je struktura tabulkového souboru složitější, je nutné konzultovat s IT specialistou možnosti vytvoření automatického skriptu pro konverzi do CSV souboru. Ne vždy je to možné.
- Pokud je datová sada reprezentována v podobě souboru v proprietárním datovém formátu, jehož specifikace není otevřená, je nutné konzultovat možnosti převodu s IT specialistou. Pro vytvoření tabulky byl použit SW nástroj, který by měl umožnit uložení tabulky i v některém z otevřených formátů.
- Pokud není možné vytvořit skript pro konverzi, je nutno převod provést ručně.
- Může se stát, že stávající datový soubor ve skutečnosti obsahuje více tabulek s různou strukturou. V tom případě je nutno rozložit soubor na více datových sad, protože datové sady by měly zůstat co nejjednodušší.
- Převod samozřejmě musí zajistit, že struktura výsledného CSV souboru odpovídá navrženému datovému schématu.

### 10.10.2 Převod datové sady s hierarchickou strukturou

Pro reprezentaci hierarchické struktury používáte nějaký proprietární strojově čitelný formát, jehož specifikace není otevřená. V tom případě je nutné konzultovat možnosti převodu s IT specialistou.

### 10.10.3 Převod datové sady v podobě kolekce textových dokumentů

Pro reprezentaci textových dokumentů používáte nějaký proprietární strojově čitelný formát, jehož specifikace není otevřená. Textové dokumenty byly jistě vytvořeny v nějakém textovém editoru, který tento formát podporuje. Textový editor by měl umožňovat i uložení dokumentů v nějakém formátu s otevřenou specifikací. Konzultujte proto s IT specialistou možnost takového uložení.

### 10.10.4 Převod ostatních typů datových sad

V ostatních případech postupujte obdobně jako v případě textových dokumentů a obrázků.

## 10.11 Rozhodnutí o podobě aktualizací datové sady

Všechny datové sady se v čase mění a je potřeba zajistit jejich periodickou aktualizaci. Některé datové sady se mění tak, že přibývají nové záznamy a ty, které již existovaly, se nemění. Jiné datové sady se naopak mění tak, že se mění existující záznamy. Pro každý takový typ datové sady může být vhodné reprezentovat aktualizace jiným způsobem.

- 1) Pokud se datová sada mění pouze tak, že přibývají nové položky a ty již publikované se nemění, vytvořte při každé periodě aktualizace nový datový soubor s novými položkami.
  - a) Nově vytvořený datový soubor je nutné v katalogizačním záznamu popsat jako novou distribuci datové sady.
  - b) Do názvu nebo popisu distribuce dejte informaci, že se jedná o přírůstková data za dané období.
- 2) Pokud se datová sada mění nejen tím, že přibývají nové položky, ale mohou se měnit i ty stávající (aktualizují se, zneplatní či jsou přímo mazány), existují dvě možnosti:

- a) Udržujte pouze jeden datový soubor. Při každé periodě vytvořte nový, kterým nahradíte ten předchozí.
  - i) V katalogizačním záznamu bude jen jedna distribuce popisující tento jediný datový soubor. Při nahrazení pouze příslušným způsobem metadata o distribuci upravte.
  - ii) Do názvu či popisu distribuci uveďte informaci, že je v datovém souboru dostupný celý obsah datové sady najednou.
- b) Pro každou periodu vytvořte tři datové soubory:
  - i) Soubor se záznamy, které se v periodě změnily.
  - ii) Soubor se záznamy, které v periodě nově přibyly.
  - iii) Soubor se záznamy, které byly v periodě odstraněny či zneplatněny.
  - iv) V katalogizačním záznamu je nutné tyto tři nové soubory popsat jako tři nové distribuce datové sady.
  - v) Do názvu nebo popisu každé distribuce dejte informaci, že se jedná o přírůstek/upravené záznamy/odstraněné či zneplatněné záznamy.
  - vi) V této variantě publikace aktualizací je nutné jednotlivé záznamy jednoznačně identifikovat pomocí identifikátorů, aby bylo možné jednoznačně propojit záznamy v jednotlivých verzích.

## 10.12 Konzultace přípravy datové sady s IT specialistou

Automatizace přípravy datové sady může výrazně urychlit a zlevnit proces opakovatelné publikace datové sady a pomůže zamezit řadě chyb, které by vznikly při ruční přípravě.

Automatizace znamená vytvoření opakovatelně spustitelného skriptu, který vytváří datový soubor v požadované podobě a ukládá jej na domluvené místo, ke kterému má Kurátor dat přístup (lze realizovat i zasláním emailu s datovým souborem v příloze, pokud to jeho velikost dovolí). Skript lze vytvořit např. v následujících případech:

1. Pokud jsou data uložena v databázovém systému (relačním či jiném typu), ke kterému má IT specialista přístup, měl by být schopen vytvořit skript (v případě relační databáze to může být sada SQL příkazů), který data z databáze extrahuje a uloží do souboru v požadované podobě.
2. Pokud jsou data uložena v databázovém systému, ke kterému IT specialista nemá přístup, protože jej dodavatel neudělil, je nutno zajistit, aby dodavatel přístup k systému dodatečně udělil nebo je nutno se s žádostí o vytvoření skriptu na dodavatele obrátit.
3. Pokud jsou data IT specialistovi přístupná pouze prostřednictvím datového API a není možné mu zajistit přístup přímo k databázi, prodiskutujte s ním možnost vytvoření datového souboru pomocí sady dotazů do datového API. Pokud API umožňuje získat seznam všech datových položek datové sady (např. řádků tabulky) a poté detailní data ke každé datové sadě, měl by být IT specialista schopen skript vytvořit. Takové funkce běžně datová API nabízejí.
4. Pokud jsou již data v nějakém datovém souboru uložena a je nutné je převést do jiné podoby, lze přímočaré transformace automatizovat. Zejména by měl být IT specialista schopen pomocí skriptu automatizovat následující operace:
  - konverze jednoduché tabulky uložené v datovém souboru tabulkového editoru (např. XLS(X), ODS, atd.) do formátu CSV.
  - konverze formátu (CSV, XML či JSON) do jiného formátu (CSV, XML či JSON),
  - přejmenování datových prvků (přejmenování sloupců v CSV souboru, přejmenování elementů v XML či JSON souboru),
  - konverze datových typů (např. konverze jednoho formátu datumů do jiného formátu).



## 10.13 Validace datového souboru

Validaci datového souboru provádí Kurátor dat, nebo pokud nemá dostatečné znalosti, IT specialista. Validace spočívá v následujících kontrolách:

- zda formát datového souboru (CSV, XML, RDF, JSON, ...) odpovídá publikačnímu plánu a katalogizačnímu záznamu;
- zda má soubor správné kódování - pro CSV, XML, RDF a JSON to je UTF-8;
- kontrola validity datového souboru vůči datovému schématu - například validace XML souboru odpovídajícím XML schématem (DTD, XSD, ...)

Pro validaci CSV lze použít například nástroj [CSV Lint](#)

Pro validaci RDF lze použít například nástroj [RDF Translator](#)

Pro validaci JSON lze použít například nástroj [JSONLint](#)

Pro validaci XML lze použít například nástroj [XSD validation](#)

## 10.14 Spolupráce na přípravě datové sady k publikaci

Přípravu datové sady k publikaci provádí její Kurátor dat. Může být ale možné přípravu automatizovat, např. exportem datových souborů z databázového systému, ve kterém jsou data uložena, nebo z informačního systému, který se používá k jejich správě. IT specialista musí v tomto ohledu Kurátorovi dat poskytnout podporu.

Kurátor dat především může potřebovat:

- Vytvořit skripty zajišťující pravidelný automatický export datových souborů tvořících datovou sadu z databázového či informačního systému (např. SQL skript);
- Vytvořit skripty zajišťující automatickou konverzi datových formátů a úpravu jejich struktury (např. XSLT skript);
- Provést validaci vytvořených datových souborů tvořících datovou sadu vůči datovému schématu datové sady.

## 10.15 Kontrola datové sady připravené k publikaci

Poté co Kurátor dat připravil datovou sadu k publikaci, předá ji Koordinátorovi otevírání dat. Ten zkontroluje, že obdržel kompletní materiály:

- informaci o zvoleném stupni otevřenosti datové sady (v případě, že je datová sada publikována poprvé),
- strojově čitelný soubor s definicí datového schématu datové sady (pokud standardy vyžadují jeho vytvoření),
- strojově čitelný datový soubor s obsahem datové sady či datový soubor s aktualizací obsahu datové sady,
- informaci o tom, že proběhla validace datového souboru (tj. kontrola, že datový soubor dodržuje všechna pravidla daná zvoleným datovým formátem a že jeho struktura odpovídá definovanému datovému schématu),
- katalogizační záznam s vyplněnými všemi aktuálně známými atributy.



## 10.16 Zajištění zveřejnění datové sady

Poté, co Koordinátor otevírání dat zkontroloval datovou sadu připravenou Kurátorem dat k publikaci, zajistí její samotné zveřejnění.

V případě, že Poskytovatel dat zveřejňuje datové soubory tvořící datové sady na svých webových stránkách, pověří Koordinátor otevírání dat IT specialistu zveřejněním následujících souborů na webu:

- datový soubor s obsahem datové sady či datový soubor s aktualizací obsahu datové sady,
- soubor s definicí datového schématu datové sady.

Dále Koordinátor otevírání dat pověří Správce katalogu otevřených dat založením (v případě prvního zveřejnění datové sady) či úpravou (v případě opakovaného zveřejnění) katalogizačního záznamu v určeném katalogu.

## 10.17 Zveřejnění datové sady

IT specialista zveřejňuje datovou sadu nahráním jejího souboru na k tomu určené místo na webových stránkách Poskytovatele dat tak, aby byl soubor přístupný z Internetu přes URL. Výjimku tvoří případ, kdy bude soubor datové sady nahrán přímo do zvoleného katalogového řešení.

## 10.18 Vytvoření a správa katalogizačního záznamu v lokálním katalogu

Pokud Poskytovatel provozuje lokální katalog otevřených dat splňující požadavky na rozšířené CKAN API a má katalog zaregistrován v NKOD, Správce katalogu otevřených dat postupuje s vytvořením nebo aktualizací datové sady v jeho lokálním katalogu dle zvoleného řešení (např. CKAN) a dle katalogizačního záznamu od Kurátora dat.

Tento způsob katalogizace šetří agendu spojenou s posíláním datových zpráv do NKOD pro každou datovou sadu a zároveň umožňuje prezentovat data Poskytovatele v rámci jeho webové prezentace. Lokální katalog Poskytovatele se do NKOD načítá celý (všechny datové sady). Při zrušení záznamu o registraci lokálního katalogu dojde i ke zrušení záznamů datových sad z tohoto katalogu.

## 10.19 Vytvoření a správa katalogizačního záznamu v Národním katalogu otevřených dat (NKOD)

Správce katalogu otevřených dat po zveřejnění datové sady IT specialistou musí zajistit, že katalogizační záznam o datové sadě existuje v Národním katalogu otevřených dat (NKOD). Způsob provedení závisí na rozhodnutí o způsobu katalogizace dat. V případě, že poskytovatel provozuje lokální katalog, nemusí s NKOD pracovat přímo. Stačí vytvořit či upravit katalogizační záznam o datové sadě v lokálním katalogu (viz Vytvoření a správa katalogizačního záznamu v lokálním katalogu, jehož obsah je pak při dodržení správného postupu automaticky přenášen do NKOD).

V případě, že poskytovatel neprovozuje vlastní lokální katalog, využívá pro katalogizaci datových sad přímo NKOD. Při práci postupuje podle toho, zda se jedná o katalogizaci nové sady nebo úpravu existujícího katalogizačního záznamu.

Správa katalogizačního záznamu v NKOD je prováděna zasíláním XML dokumentů ve formě přílohy datové zprávy do datové schránky s identifikátorem uur3q2i „Automat PVS

(Ministerstvo vnitra)“ spravované Ministerstvem vnitra ČR. Tyto XML dokumenty lze získat pomocí [formulářů](#) dostupných na stránkách [Rejstříku datových zdrojů](#), kde je také k dispozici [dokumentace](#) jejich struktury.

### 10.19.1 Katalogizace nové datové sady

Postup:

1. Ve formuláři NKOD kliknout na „Zveřejnění datové sady“;
2. V následujících krocích pro datovou sadu a datové zdroje vyplnit katalogizační záznam připravený Kurátorem dat;
3. V posledním kroku je třeba zvolit jednu ze dvou možností:
  - a. přímé odeslání z formuláře do ISDS - pokud jsou k dispozici přihlašovací údaje k datové schránce Poskytovatele dat,
  - b. alternativní způsob odeslání - pak je třeba výsledný XML soubor poslat datovou schránkou (například přes spisovou službu Poskytovatele) jako přílohu zprávy a adresovat ji na datovou schránku Automat PVS (Ministerstvo vnitra) s ID uur3q2i.

### 10.19.2 Úprava katalogizačního záznamu

Pokud poskytovatel používá pro katalogizaci datových sad NKOD a datovou sadu upravuje, tj. v NKOD již pro ni existuje záznam, postupuje takto:

- 1) V rejstříku datových zdrojů NKOD nalézt záznam pro upravovanou datovou sadu a kliknout na operaci Upravit tento záznam;
- 2) V následujících krocích pro datovou sadu a datové zdroje upravit katalogizační záznam připravený Kurátorem dat;
- 3) V posledním kroku je třeba zvolit jednu ze dvou možností:
  - a) přímé odeslání z formuláře do ISDS - pokud jsou k dispozici přihlašovací údaje k datové schránce Poskytovatele dat,
  - b) alternativní způsob odeslání - pak je třeba výsledný XML soubor poslat datovou schránkou (například přes spisovou službu Poskytovatele dat) jako přílohu zprávy a adresovat ji na datovou schránku Automat PVS (Ministerstvo vnitra) s ID uur3q2i.

### 10.19.3 Zrušení katalogizačního záznamu

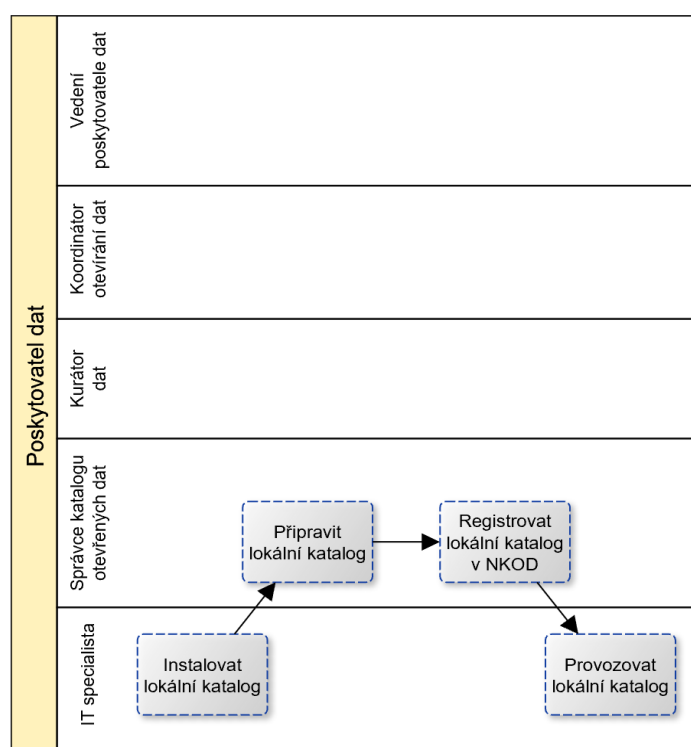
V případě rušení datové sady se postupuje takto:

1. V Rejstříku datových zdrojů nalézt ID záznamu datové sady ke zrušení
2. Ve formuláři NKOD kliknout na „Smazání datové sady nebo odkazu na katalog“;
3. Vyplnit ID záznamu a kliknout na „Další“;
4. V posledním kroku je třeba zvolit jednu ze dvou možností:
  - a. přímé odeslání z formuláře do ISDS - pokud jsou k dispozici přihlašovací údaje k datové schránce Poskytovatele dat,
  - b. alternativní způsob odeslání - pak je třeba výsledný XML soubor poslat datovou schránkou (například přes spisovou službu Poskytovatele dat) jako přílohu zprávy a adresovat ji na datovou schránku Automat PVS (Ministerstvo vnitra) s ID uur3q2i.

## 11 Zprovoznění lokálního katalogu otevřených dat a jeho registrace v NKOD

Datové sady publikované v otevřené podobě je nutno katalogizovat. Pokud je požadováno provádět katalogizaci ve vlastním lokálním katalogu otevřených dat, je nutné jej zprovoznit. Zprovoznění zahrnuje několik činností, které jsou zachyceny v následujícím diagramu. Nedílnou součástí zprovoznění lokálního katalogu je vždy jeho registrace v Národním katalogu otevřených dat (NKOD). Po kliknutí na činnost se dostanete na detailní popis její doporučené náplně. Jednotlivé plavecké dráhy odpovídají rolím, které by se měly procesu otevírání dat účastnit. Po kliknutí na roli se dostanete na její popis.

**Zprovoznění lokálního katalogu je nepovinné.** Alternativně může poskytovatel katalogizovat svoje datové sady přímo v NKOD (viz činnost Správa katalogizačního záznamu).



Legenda:

- Obdélník s plným okrajem značí povinnou činnost.
- Obdélník s přerušovaným okrajem značí nepovinnou činnost.
- Šipky značí návaznost mezi činnostmi.

### 11.1 Instalace lokálního katalogu

Dle rozhodnutí Koordinátora otevírání dat o provozu lokálního katalogu, IT specialista provede instalaci CKAN nebo DKAN na server Poskytovatele dat nebo externí webhosting. V případě, že žádný z výše zmiňovaných nástrojů nevyhovuje, lze implementovat vlastní, pokud alespoň dodrží rozšířené CKAN API.

## 11.2 Příprava lokálního katalogu

Správce (lokálního) katalogu otevřených dat, po instalaci katalogu IT specialistou, připraví katalog pro zveřejnění a pro zadávání záznamů o datových sadách. To zahrnuje:

- identifikaci a vytvoření uživatelů, kteří budou moci spravovat záznamy o datových sadách,
- případnou úpravu vzhledu katalogu dle grafického stylu Poskytovatele.

## 11.3 Provoz lokálního katalogu

V rámci provozu lokálního katalogu IT specialista zajistí zejména jeho aktualizaci při dostupnosti nové verze použitého software, zálohování, kontrolu dostupnosti a obnovení v případě výpadku.

## 11.4 Registrace vlastního katalogu v NKOD

Pokud Poskytovatel provozuje lokální katalog otevřených dat nebo využívá služeb jako Datahub.io s podporou CKAN API v3, může si katalog zaregistrovat v NKOD a tím zajistit publikaci záznamů z lokálního katalogu v NKOD. Pro registraci je nutný přístup k datové schránce Poskytovatele dat nebo možnost datovou schránkou poslat XML soubor jako přílohu.

Správa záznamu lokálního katalogu je prováděna zasíláním XML dokumentů ve formě přílohy datové zprávy do datové schránky s identifikátorem uur3q2i „Automat PVS (Ministerstvo vnitra)“ spravované Ministerstvem vnitra ČR. Tyto XML dokumenty lze získat pomocí formulářů dostupných na stránkách Rejstříku datových zdrojů, kde je také k dispozici dokumentace jejich struktury.

### 11.4.1 Registrace lokálního katalogu

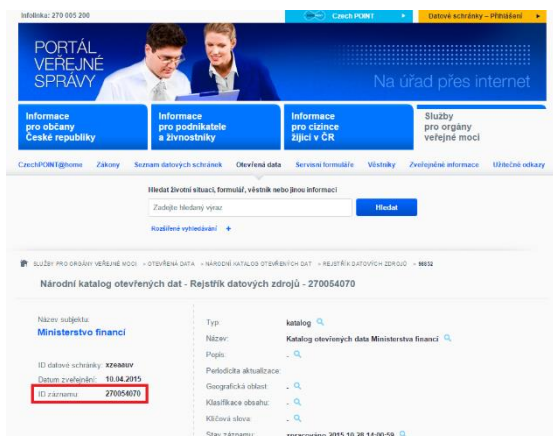
Registrace lokálního katalogu je provedena zasláním registračního XML dokumentu, který lze získat například vyplněním formuláře „Registrace odkazu na lokální katalog“. V tom případě je postup registrace lokálního datového katalogu v NKOD následující:

- 1) Ve formuláři NKOD kliknout na „Registrace odkazu na lokální katalog“;
- 2) Vyplnit název katalogu, jméno a email správce katalogu a kliknout na „Další“;
- 3) Do pole Odkaz na stažení datového zdroje zadat adresu funkce `package_list` CKAN API lokálního katalogu, která vrací seznam názvů obsažených datových sad, například `http://datahub.io/api/3/action/package_list`, a kliknout na „Další“;
- 4) V posledním kroku je třeba zvolit jednu ze dvou možností:
  - a) přímé odeslání z formuláře do ISDS - pokud jsou k dispozici přihlašovací údaje k datové schránce Poskytovatele dat;
  - b) alternativní způsob odeslání - pak je třeba výsledný XML soubor poslat datovou schránkou (například přes spisovou službu Poskytovatele dat) jako přílohu zprávy a adresovat ji na datovou schránku „Automat PVS (Ministerstvo vnitra)“ s ID uur3q2i.

## 11.4.2 Zrušení registrace lokálního katalogu

Registraci lokálního katalogu otevřených dat je možno zneplatnit (zrušit) zasláním XML dokumentu, který lze získat pomocí formuláře „Smazání datové sady nebo odkazu na katalog“. V tom případě je postup zrušení následující:

- 1) V Rejstříku datových zdrojů nalézt ID záznamu s registrací lokálního katalogu, viz obrázek;



- 2) Ve formuláři NKOD kliknout na „Smazání datové sady nebo odkazu na katalog“;
- 3) Vyplnit ID záznamu a kliknout na „Další“;
- 4) V posledním kroku je třeba zvolit jednu ze dvou možností:
  - a) přímé odeslání z formuláře do ISDS - pokud jsou k dispozici přihlašovací údaje k datové schránce Poskytovatele dat;
  - b) alternativní způsob odeslání - pak je třeba výsledný XML soubor poslat datovou schránkou (například přes spisovou službu Poskytovatele dat) jako přílohu zprávy a adresovat ji na datovou schránku „Automat PVS (Ministerstvo vnitra)“ s ID uur3q2i.



## 12 Nástroje pro publikaci otevřených dat

Existuje řada nástrojů, které publikaci otevřených dat usnadňují. Zde uvádíme nejzajímavější kandidáty, jejich popis a způsob použití.

- Národní katalog otevřených dat (NKOD) - katalogizační nástroj a úložiště otevřených dat dostupný na portálu veřejné správy.

*Nástroje pro vytváření lokálních katalogů.* Na odkazovaných stránkách je poskytnut návod na zprovoznění a úpravu příslušného katalogizačního nástroje. Úprava umožní automatický sběr v lokálním katalogizačním nástroji uvedených metadat do Národního katalogu otevřených dat:

- CKAN
- DKAN

### 12.1 Národní katalog otevřených dat (NKOD)

Národní katalog otevřených dat veřejné správy ČR (dále „NKOD“) (<http://portal.gov.cz/portal/ovm/rejstriky/data/97898/>) je centrálním katalogem otevřených dat v ČR. NKOD je dostupný prostřednictvím Portálu veřejné správy ČR (dále „PVS“) v záložce „Otevřená data“. NKOD obsahuje katalogizační záznamy o datových sadách zveřejněných jednotlivými subjekty veřejné správy ČR, včetně odkazů ke stažení dat. Data ke stažení mohou být uložena na webu příslušného Poskytovatele dat nebo přímo na v NKOD.

Národní katalog otevřených dat veřejné správy ČR umožňuje katalogizaci následujícími způsoby:

- vyplnění formuláře na PVS a jeho odeslání přes Informační systém datových schránek, součástí katalogizačního záznamu mohou být i vlastní otevřená data, která pak budou přístupná prostřednictvím NKOD,
- automatické plnění metadaty z lokálních katalogů OVM (po předchozí registraci lokálního katalogu v NKOD). Registrace se opět provádí vyplněním a odesláním formuláře s registrací lokálního katalogu přes Informační systém datových schránek.

Současně je možné zrychlit vyplňování formuláře s metadaty pro NKOD tím, že zvolíte u konkrétní datové sady publikované jinou organizací operaci „Mám podobná data a chci je také zaregistrovat do NKOD“. Pak je možné využít již předvyplněná metadata a doplnit pouze své specifické údaje a přiložit vlastní odkaz na otevřená data nebo případně přímo soubory s otevřenými daty.

NKOD se skládá ze 2 rejstříků:

- **Datové sady** - zobrazení a vyhledávání všech katalogizačních záznamů o datových sadách zveřejněných jednotlivými OVM. Tento rejstřík je viditelný pro všechny skupiny uživatelů PVS.
- **Rejstřík datových zdrojů** - jediné místo pro vkládání, editaci a rušení katalogizačních záznamů (příp. i souborů s otevřenými daty) a registrací lokálních katalogů v NKOD. Tento rejstřík je dostupný pouze v záložce PVS „Služby pro orgány veřejné moci“. Národní katalog otevřených dat (NKOD).

## 12.2 CKAN a jeho rozšíření

Pro instalaci CKAN lze volit ze tří možností podle [návodu pro instalaci \(anglicky\)](#). Instalace CKAN vyžaduje IT specialistu se znalostí administrace OS UNIX.

### Výhody:

- CKAN vyvíjí Open Knowledge Foundation (OKFN) – je zde nízké riziko přerušení podpory open-source verze;
- Běží na [stovkách instancí po světě](#);
- Už má za sebou více jak 5 let vývoje, má vybudovanou komunitu vývojářů;
- Má API i pro zápis dat z jiných aplikací (DKAN jen ručně).

### Nevýhody:

- Požadavky: Linux, PostgreSQL, Python, Java - nestačí tedy běžný webhosting.

### 12.2.1 Instalace

V [návodu pro instalaci \(anglicky\)](#) jsou popsány tři způsoby instalace v závislosti na konkrétním operačním systému.

#### *Instalace na Ubuntu 12.04 64-bitová verze*

Postupujte dle [návodu pro instalaci z balíčku \(anglicky\)](#), který vás provede instalací PostgreSQL, Apache Solr a balíčku CKAN a jednoduchou inicializací.

#### *Instalace pomocí obrazu Docker*

K instalaci pomocí [Docker](#) je potřeba jen stroj, který umí pracovat s Docker obrazy. Postupujte dle [návodu na instalaci pomocí Docker obrazu \(anglicky\)](#).

#### *Instalace ze zdrojového kódu*

Pokud vám nevyhovuje ani instalace pomocí balíčku, ani instalace přes Docker obrazy a máte přístup k serveru s OS Linux a s nainstalovanými technologiemi Python, Java a PostgreSQL, lze CKAN nainstalovat ze zdrojového kódu pomocí [návodu na instalaci ze zdrojového kódu \(anglicky\)](#).

V každém případě je pro kompatibilitu s NKOD třeba po instalaci CKAN ještě doinstalovat [rozšíření metadat](#) dle [návodu na instalaci rozšíření CKAN \(anglicky\)](#).

## 12.3 DKAN a jeho rozšíření

Pro instalaci DKAN je potřeba mít přístup k webovému serveru s podporou technologií PHP a MySQL. To může být i externí webhostingová služba. Dále je potřeba postupovat podle [návodu na instalaci \(anglicky\)](#), kde jsou popsány způsoby instalace, nejčastěji půjde o [instalaci na vlastní server](#) buďto ze zdrojového kódu nebo z předpřipravených balíčků.

Pro DKAN zatím **neexistuje** instalovatelné rozšíření metadatových položek pro kompatibilitu s NKOD.

### Výhody:

- DKAN je modul pro redakční systém Drupal 7 - lze využít jeho rozšíření a grafické styly
- Snadnější instalace, požadavky jsou jen: PHP + MySQL

**Nevýhody:**

- DKAN API je jen pro čtení. To znamená, že uložené katalogizační záznamy lze strojově číst, ale nelze je strojově vkládat, musí se vkládat a upravovat ručně. Není tedy vhodný do prostředí, kde jsou katalogizovaná data generována automatizovaně.
- DKAN vyvíjí NuCivic - komerční firma. Podpora open-source verze v budoucnosti je tedy nejistá, plně v moci této firmy.
- Produkt je mladý, nemá vybudovanou vývojářskou základnu a vyžaduje časté aktualizace.

## 12.4 Použití widgetu NKOD

Bude připraven widget, tj. kód zobrazující datové sady z NKOD pro daného Poskytovatele dat, který lze jednoduše vložit do webové stránky Poskytovatele dat. V současné době tento widget ještě není k dispozici.

## 12.5 Použití NKOD API

NKOD bude poskytovat API - strojové rozhraní pro získání katalogizovaných datových sad Poskytovatele dat. Získané záznamy o datových sadách pak lze přizpůsobit grafickému designu webových stránek Poskytovatele dat a tím je na web Poskytovatele dat začlenit. Toto rozhraní však zatím není k dispozici.

## 12.6 Služba Datahub.io

Datahub.io je veřejná instance CKAN od Open Knowledge Foundation (OKFN). Lze ji snadno využít k publikaci vlastních datových sad s jediným omezením - nelze ji rozšířit o metadata kompatibilní s NKOD. Stačí pro Poskytovatele dat založit uživatele, organizaci a pod ní vytvářet záznamy pro datové sady.

**Výhody:**

- Není třeba nic instalovat, provozovat ani nic platit. Služba Datahub.io je zdarma;
- Je to celosvětově známé místo pro katalogizaci otevřených dat.

**Nevýhody:**

- Je to služba zdarma. Neexistuje záruka, že služba nepřestane fungovat;
- Nelze ji rozšířit o podporu metadat pro NKOD.

**Příklady:**

Statutární město Brno publikuje svá otevřená data na Datahub.io, pod svou organizací Statutární město Brno.

## 13 Nástroje pro publikaci otevřených prostorových dat

- Národní geoportál INSPIRE (<http://geoportal.gov.cz>) - Katalog datových sad a služeb.
- PyCSW (<http://pycsw.org>) - Lehká serverová aplikace pro tvorbu katalogu prostorových dat podporující INSPIRE specifikace.
- GeoNetwork open source (<http://geonetwork-opensource.org/>) - Komplexní serverové řešení pro tvorbu katalogu prostorových dat podporující INSPIRE specifikace.
- MapServer (<http://mapserver.org>) - Mapový server umožňující publikaci prostorových dat pomocí webových služeb OGC (OGC OWS) s konfigurací pomocí textových souborů.
- GeoServer (<http://geoserver.org>) - Mapový server umožňující publikaci prostorových dat pomocí webových služeb OGC (OGC OWS) s webovým konfiguračním rozhraním.

## 14 Technické standardy

### 14.1 Sdílené datové typy

V doporučených datových schématech používáme primitivní datové typy inspirované [datovými typy XML Schema](#).

#### 14.1.1 Datový typ: AnoNe

Tento datový typ reprezentuje binární hodnotu. Může nabývat hodnot *true* nebo *false*.

Předpokládá se, že bude odpovídat specifikaci datového typu [xsd:boolean](#) z jazyka XML Schema, a to při použití v jiných formátech než XML, tj. CSV, XML, RDF, JSON.

#### 14.1.2 Datový typ: Čas

Tento datový typ reprezentuje časový údaj (bez data).

Předpokládá se, že bude odpovídat specifikaci datového typu [xsd:time](#) z jazyka XML Schema, a to při použití v jiných formátech než XML, tj. CSV, XML, RDF, JSON.

#### 14.1.3 Datový typ: Celé číslo

Tento datový typ reprezentuje celočíselný údaj (bez desetinných míst).

Předpokládá se, že bude odpovídat specifikaci datového typu [xsd:integer](#) z jazyka XML Schema, a to při použití v jiných formátech než XML, tj. CSV, XML, RDF, JSON.

#### 14.1.4 Datový typ: Datum

Tento datový typ reprezentuje datum (bez času).

Předpokládá se, že bude odpovídat specifikaci datového typu [xsd:date](#) z jazyka XML Schema, a to při použití v jiných formátech než XML, tj. CSV, XML, RDF, JSON.

#### 14.1.5 Datový typ: Desetinné číslo

Tento datový typ reprezentuje číselný údaj s desetinnými místy.

Předpokládá se, že bude odpovídat specifikaci datového typu [xsd:decimal](#) z jazyka XML Schema, a to při použití v jiných formátech než XML, tj. CSV, XML, RDF, JSON.

### 14.1.6 Datový typ: Řetězec

Tento datový typ odpovídá kratšímu řetězci znaků (typicky kód, ne souvislý text) v kódování UTF-8.

Předpokládá se, že bude odpovídat specifikaci datového typu [xsd:string](#) z jazyka XML Schema, a to při použití v jiných formátech než XML, tj. CSV, XML, RDF, JSON.

### 14.1.7 Datový typ: URL

Tento datový typ reprezentuje URL - Uniform Resource Locator.

Předpokládá se, že bude odpovídat specifikaci datového typu [xsd:anyURI](#) z jazyka XML Schema, a to při použití v jiných formátech než XML, tj. CSV, XML, RDF, JSON.

## 14.2 Technické standardy pro datové sady na stupni otevřenosti 3

Pro dosažení stupně otevřenosti datové sady 3 je nutno zajistit, aby kompletní obsah datové sady byl dostupný prostřednictvím 1 či více distribucí datové sady, které splňují následující standardy.

Obecným standardem datových sad na stupni 3 je, že každá distribuce datové sady musí být kódována v kódování UTF-8.

### 14.2.1 Standardy pro tabulkové datové sady

Distribuce datové sady, která reprezentuje tabulku, musí splňovat následující standardy:

- V jedné distribuci je reprezentována právě jedna tabulka.
- V distribuci neexistuje žádný jiný záznam, než je řádek tabulky nebo její hlavička.
- Všechny záznamy v distribuci, tj. jednotlivé řádky tabulky, mají stejnou strukturu, která odpovídá hlavičce.
- Celý obsah datové sady je vyjádřen v distribucích v alespoň jednom z následujících formátů:
  - Comma Separated Values (CSV, IETF RFC4180), kde jednotlivé řádky tabulky jsou zapsány jako řádky textového souboru a údaje v řádku tabulky jsou zapsány jako řetězce oddělené čárkou (znak s UTF-8 kódem U+002C), přičemž hodnota může být uzavřena v uvozovkách (znak s UTF-8 kódem U+0022), především pokud hodnota sama obsahuje jednu nebo více čárek,
  - eXtensible Markup Language (XML, <http://www.w3.org/TR/xml/>), kde jednotlivé řádky tabulky jsou zapsány jako samostatné XML elementy strukturované pomocí zanořených XML elementů a XML atributů reflektujících hlavičku reprezentované tabulky, nebo
  - JavaScript Object Notation (JSON, <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>), kde jednotlivé



řádky tabulky jsou reprezentovány jako samostatné JSON objekty strukturovaných do dle hlavičky reprezentované tabulky.

## 14.2.2 Standardy pro datové sady s hierarchickou strukturou

Distribuce datové sady, která obsahuje hierarchické strukturované objekty (tj. objekty složené z jiných objektů), které není vhodné reprezentovat v podobě tabulky dle předchozí podkapitoly, musí splňovat následující standardy:

- V jedné distribuci je reprezentována kolekce stejného typu objektů
- Objekty v kolekci mohou obsahovat další vnořené typy objektů.
- Celý obsah datové sady je vyjádřen v distribucích v alespoň jednom z následujících formátů:
  - eXtensible Markup Language (XML, <http://www.w3.org/TR/xml/>), jehož struktura vyjádřená v podobě XML elementů a XML atributů umožňuje výběr jednotlivých údajů reprezentovaných v datové sadě pomocí prostředků jazyka XPath či CSS selektorů bez nutnosti parsování získaných atomických údajů.
  - JavaScript Object Notation (JSON, ECMA-404, nebo <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>), jehož struktura vyjádřená v podobě JSON objektů a primitivních hodnot umožňuje výběr jednotlivých údajů reprezentovaných v datové sadě pomocí prostředků programovacích jazyků používaných pro tvorbu WWW aplikací (JavaScript, Java, apod.) bez nutnosti parsování získaných atomických údajů.

## 14.2.3 Standardy pro datové sady v podobě kolekce textových dokumentů

Distribuce datových sad, které mají charakter textových dokumentů, musí splňovat následující pravidla a standardy:

- V jedné distribuci je reprezentována kolekce dokumentů, kde každý dokument je jednoznačně identifikovatelný v podobě souboru a lze jej z kolekce získat pomocí běžných programovacích prostředků.
- Všechny dokumenty v datové sadě jsou vyjádřeny alespoň v jednom z následujících formátů:
  - Formát odpovídající MIME typu text/plain obsahující pouze znaky textového dokumentu.
  - Libovolný jiný formát pro textové dokumenty, který je strojově čitelný, má otevřenou specifikaci a existuje volně dostupná knihovna, umožňující automatizovanou strojovou extrakci textového obsahu dokumentů v datové sadě (až do úrovně jednotlivých znaků).
    - Např. HTML, DOCX, ODT, DocBook či TeX.

## 14.2.4 Standardy pro datové sady v podobě kolekce obrázků

Distribuce datových sad, které mají charakter obrázků, musí splňovat následující pravidla a standardy:

- V jedné distribuci je reprezentována kolekce obrázků, kde každý dokument je jednoznačně identifikovatelný v podobě souboru a lze jej z kolekce získat pomocí běžných programovacích prostředků.
- Každý obrázek je vyjádřen ve formátu, který je strojově čitelný a má otevřenou specifikaci.

- V případě bitmapových obrázků musí formát umožnit přistupovat pomocí běžných programovacích prostředků k jednotlivým pixelům tvořícím obrázek.
- V případě vektorových obrázků musí formát umožnit přistupovat pomocí běžných programovacích prostředků k jednotlivým vektorovým objektům tvořícím obrázek.

## 14.2.5 Standardy pro prostorové datové sady

Podle velikosti obcí a organizací - resp. přítomnosti oddělení GIS v rámci jejich organizační struktury - mohou být použity postupy a datové formáty dostupné v jejich IT infrastruktuře.

U malých obcí či organizací je celkem zbytečné budovat komplexní prostorovou infrastrukturu (Spatial Data Infrastructure - SDI). V případě, že obec či organizace nějakým způsobem spravuje data s prostorovou složkou (tj. geografická data, geodata), ať už bodová (jako např. seznam památných stromů nebo pomníků), liniová (lokální cestní síť) nebo plošná (přehled obecní zeleně) a chce tyto datové sady otevřít, měla by k tomu zvolit vhodný souborový formát:

- [Comma Separated Values](#) (CSV, IETF RFC4180) - jednotlivé záznamy atributové tabulky jsou zapsány jako řádky textového souboru a hodnoty záznamů jako řetězce oddělené čárkou (znak UTF-8 s kódem U+002C), přičemž hodnota může být uzavřena v uvozovkách (znak UTF-8 s kódem U+0022), především pokud hodnota sama obsahuje jednu nebo více čárek. Položky s prostorovou informací mohou být vyjádřeny souřadnicemi, pokud se jedná o bodová data. U ostatních typů dat jako jsou liniová anebo polygonová data může být použit zápis WKT (viz *Poznámka k zápisu geometrie*). Dále by měl být tento zápis opatřen informací o souřadnicovém systému ve formě EPSG kódu (viz *Poznámka k souřadnicovým systémům*), např. v hlavičce souboru.
- [ESRI Shapefile](#) - vlastnosti jednotlivých prvků dodržují strukturu pro datovou sadu.
- [Geography Markup Language](#) (OGC GML) - jedná se formát XML pro prostorová data. Atributy prvků reflektují strukturu datové sady.
- [GeoJSON](#) - formát, kde jsou jednotlivé záznamy atributové tabulky reprezentovány jako samostatné JSON objekty strukturované dle hlavičky reprezentované tabulky a navíc obsahují prostorovou informaci.
- [OGC GeoPackage](#) - je nad souborovou databází [SQLite](#) postavený standardizovaný formát pro prostorová data. Lze do něj uložit stejnou strukturu jako do CSV a složitou geometrickou strukturu jako samostatný atribut.

Datové sady jsou publikovány tak, že všechny prvky (prostorové objekty) v datové sadě mají stejnou strukturu (stejné atributy).

Pokud jsou některá data organizaci nebo obci poskytována externím subjektem - např. firmou zpracovávající územní plán - měli by jednatelé trvat na odevzdání práce v *otevřeném formátu pro prostorová data* (viz seznam níže). Takovou datovou sadu lze pak bez dalšího úsilí otevřít. Podstatné je v této souvislosti se pokud možno vyhnout uzavřeným formátům (např. DWG, DGN, ESRI GeoDatabase, ...). Případně datové sady v těchto formátech převést před otevřením do výše doporučených otevřených formátů.

Je také vhodné udržovat systém metadat k datovým sadám (jejich slovní popis, označení původu, datum vzniku, licenci a tak dále) např. v samostatném souboru. Datové sady lze zveřejnit „tak jak jsou“ (pokud možno v otevřeném formátu, jak je zmíněno výše) na webových stránkách ve formě odkazů na soubory. K tomu není ani nutné pořizovat často nákladné webové mapové portály.

V případě publikace rozsáhlejší datové sady formou souborů je vhodné doplnit tyto soubory o [Atom](#) feed nebo je publikovat prostřednictvím webových služeb OGC. Tyto způsoby publikace umožňují jednodušší vyhledávání konkrétních dat v rámci datové sady.

U obcí s rozšířenou působností (ORP) a větších, je možné publikovat data v souladu se směrnicí INSPIRE (zákon [č. 380/2009 Sb. http://inspire.gov.cz/](#)) pomocí webových služeb [OGC](#):

- [OGC WFS](#) (Web Feature Service)
- [OGC WCS](#) (Web Coverage Service)

Data poskytovaná těmito službami lze použít pro další GIS analýzy, neboť se jedná o geodata. Pro vektorová data (passporty zeleně, silniční síť, mapu veřejných záchodků apod.) je vhodné použití OGC Web Feature Service, pro rastrová data (např. letecké snímky) potom OGC Web Coverage Service.

Dále lze doporučit i další OGC webové služby [WMS](#) (Web Map Service) a [WMTS](#) (Web Map Tile Service). *Nejedná se však o otevřený způsob distribuce dat, neboť tyto služby neposkytují geodata, ale pouze jejich otisk ve formě obrázku. Takto poskytovaná data nelze dále analyzovat prostřednictvím GIS nástrojů. Služby OGC WMS a WMTS jsou vhodné pro tvorbu webových mapových aplikací a pro využití desktopovými GIS programy jako podkladové mapy.*

Při použití webových služeb OGC je třeba dodržet vhodné nastavení serveru s ohledem na podporované souřadnicové systémy, na jeho očekávanou vytíženost či bezpečnostní rizika apod.

Větší organizace veřejné správy s pokročilou infrastrukturou pro prostorová data (SDI) publikující metadata k prostorovým datovým sadám v souladu se standardem [OGC CSW](#) propojí tato metadata do portálu otevřených dat.

#### **Poznámka k proprietárním službám**

Některé nástroje GIS obsahují proprietární služby pro publikaci dat. Data publikovaná prostřednictvím těchto služeb nelze považovat za dostatečně otevřená. Nebývají podporována existujícími softwary třetích stran, licenční podmínky jejich využití nejsou často transparentní. Jedná se například o tzv. REST API v softwarech firmy ESRI.

#### **Poznámka k souřadnicovým systémům**

**EPSG** - je uznávaný registr mimo jiné pro popis souřadnicových systémů, podporovaný všemi programy pro práci s prostorovými daty (GIS). Relevantní hodnoty pro Českou republiku jsou

- 5514 - Systém jednotné trigonometrické sítě katastrální (S-JTSK) - používaný např. v katastru nemovitostí. Pro tento systém je mimo jiné v GIS typické, že souřadnice nabývají negativních hodnot (jsou uvozeny znaménkem minus).
- 4326 - World Geodetic System 1984 (WGS84), někdy zvaný též "souřadnice GPS". Souřadnice bývají uvedeny ve stupních s desetinným číslem. Některé publikační formáty (např. GPX nebo KML) počítají přímo s tím, že data jsou uložena v tomto souřadnicovém systému.
- 3857 - tzv. Web (nebo též Google) Mercator - je systém používaný na většině globálních webových aplikací. Jeho nevýhodou je mimo jiné zkreslení významné velikosti ploch směrem k pólům.

Další lze dohledat např. na stránce <http://epsg.io>

#### **Poznámka k zápisu geometrie**

**WKT** - Well Known Text - je standardní formát zápisu vektorové geometrie. Popis je mimo jiné k nalezení ve specifikaci OGC [Simple Features](#).

Typ	Příklad
Bod	POINT (30 10)
Linie	LINESTRING (30 10, 10 30, 40 40)
Polygon	POLYGON ( (30 10, 40 40, 20 40, 10 20, 30 10) )

Pomocí WKT lze samozřejmě uložit i komplikovanější datové typy, jako jsou MULTIPOINTS či CompoundCurve a další.

Pokud uvažujete o použití WKT pro popis atributu geometrie u objektů, zvažte publikaci otevřené datové sady v některém k tomu určeném otevřeném formátu (ESRI Shapefile, KML, OGC GeoPackage nebo GeoJSON).

Problematikou otevírání prostorových dat se podrobně zabývá publikace [Otevírání prostorových dat - případová studie](#).

## 14.2.6 Standardy pro ostatní typy datových sad

Jiný formát či jazyk může poskytovatel dat použít pouze v případě datové sady, jejíž charakter neumožňuje použití některého z formátů uvedených v předchozích podkapitolách nebo datová sada spadá do oblasti, ve které jsou běžně používány jiné formáty (které splňují podmínku strojové čitelnosti a otevřenosti specifikace).

Ať je zvolen jakýkoliv formát, musí být zajištěno, že celý obsah datové sady je vyjádřen v distribucích v alespoň jednom formátu, který splňuje podmínku strojové čitelnosti a otevřenosti specifikace a je v dané oblasti běžně používán.

## 14.3 Technické standardy vytváření datových schémat pro datové sady na stupni otevřenosti 3

Na stupni otevřenosti 3 by měly mít distribuce datových sad přiřazeno datové schéma, které popisuje požadovanou syntaktickou strukturu distribucí. Pro vyjádření datového schématu je nutno zvolit vhodný jazyk pro jeho vyjádření. Jazyk závisí na datovém formátu, který byl zvolen pro vyjádření distribuce.

- V případě formátu CSV je nutno schéma vyjádřit v jazyku JSON Table Schema (<http://dataprotocols.org/json-table-schema/>) nebo Metadata Vocabulary for Tabular Data (<http://www.w3.org/TR/tabular-metadata/>).
- V případě formátu XML je nutno schéma vyjádřit v jazyku XML Schema (<http://www.w3.org/XML/Schema>).
- V případě formátu JSON je nutno schéma vyjádřit v jazyku JSON Schema (<http://json-schema.org/>).

Při návrhu datových schémat navíc dodržujte následující standardy:

- Pro primitivní datové typy (řetězec, číslo, datum, ...) používejte datové typy jazyka XML Schema (i v případě CSV a JSON souborů)
- Řada složených datových typů (např. adresa) se často opakuje ve schématech různých datových sad napříč různými poskytovateli. Pro ně využijte ve svém datovém schématu [předdefinovanou strukturu](#) místo vymýšlení vlastní ad hoc struktury.
- Schéma každé distribuce musí být volně dostupné ke stažení v síti WWW.
- Distribuce se stejnou strukturou mají společné schéma.



## 14.4 Technické standardy pro datové sady na stupni otevřenosti 4 a 5

Standardem reprezentace distribucí datových sad na stupni otevřenosti 4 a 5 je datový model RDF. Distribuce jsou pak zapisovány v jednom z formátů pro zápis datového modelu RDF v kódování UTF-8.

### 14.4.1 Datový model RDF

Resource Description Framework (RDF, <http://www.w3.org/TR/rdf11-concepts/>) je datový model pro reprezentaci dat v síti WWW vytvořený a spravovaný konsorciem WWW. Data reprezentuje v podobě trojic

subjekt predikát objekt

Trojice reprezentuje hodnotu (zapsanou v místě objektu) vlastnosti (zapsanou v místě predikátu) nějaké entity (zapsané v místě subjektu). Subjektem je vždy entita. Objektem může být buď primitivní datová hodnota (řetězec, číslo apod.) nebo jiná entita. Predikátem je vlastnost subjektu, pro kterou chceme v trojici reprezentovat hodnotu.

Entity (jak na místě subjektu tak objektu) jsou identifikovány pomocí identifikátoru ve tvaru IRI. Vlastnosti jsou také identifikovány pomocí IRI. V trojicích jsou entity i vlastnosti zapisovány pomocí svého identifikačního IRI. Primitivní hodnoty jsou zapisovány jako řetězce znaků.

Entity navíc mohou být zařazeny to tzv. třídy. Do jedné třídy jsou zařazeny entity se stejnou charakteristikou. Např. všechny knihy v datové sadě mohou být zařazeny do třídy Kniha. Třídy jsou podobně jako entity a vlastnosti identifikovány pomocí IRI. Entita je zařazena do třídy pomocí RDF trojice, kde je v místě predikátu uvedeno klíčové slovo *a* (vychází z anglického neurčitého členu).

Třídy a vlastnosti jsou definovány pomocí tzv. ontologií. K ontologiím více viz kapitola „Chyba! Nenalezen zdroj odkazů.“.

### 14.4.2 Standardy pro zápis dat v datovém modelu RDF

Standardními formáty pro zápis dat v datovém modelu RDF jsou následující formáty:

- Turtle (<http://www.w3.org/TR/2014/REC-turtle-20140225/>)
- N-Quads (<http://www.w3.org/TR/2014/REC-turtle-20140225/>)
- N-Triples (<http://www.w3.org/TR/2014/REC-n-triples-20140225/>)
- JSON-LD (<http://www.w3.org/TR/2014/REC-json-ld-20140116/>)
- RDF/XML (<http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/>)

Pokud je distribuce datové sady zapsána v některém z těchto formátů a splňuje pravidla datového modelu RDF, je nazývána RDF distribuce.



### 14.4.3 Standardy tvorby IRI

Poskytovatel dat je povinen definovat IRI pro identifikaci entit, o kterých v distribucích svých datových sad na stupni otevřenosti 4 a 5 publikuje nějaké údaje. IRI musí být v rámci DNS domény poskytovatele.

V případě, že poskytovatel vlastní doménu druhého řádu je základem všech jeho IRI:

```
http://data.{doména-druhého-řádu}.{doména-prvního-řádu}
```

Prvky {doména-prvního-řádu} a {doména-druhého-řádu} jsou DNS doménou prvního a druhého řádu poskytovatele. Alternativně lze místo protokolu HTTP volit protokol HTTPS.

V případě, že poskytovatel nevlastní doménu druhého řádu, může využít doménu druhého řádu svého zřizovatele (pokud má zřizovatele a ten vlastní doménu druhého řádu). Potom je základem všech jeho IRI

```
http://data.{doména-druhého-řádu}.{doména-prvního-řádu}/{určení-  
poskytovatele}
```

Prvky {doména-prvního-řádu} a {doména-druhého-řádu} jsou DNS doménou prvního a druhého řádu zřizovatele poskytovatele. Prvek {určení-poskytovatele} volí poskytovatel dat tak, aby jej výsledné IRI jednoznačně identifikovalo mezi IRI ostatních poskytovatelů se stejným zřizovatelem. Alternativně lze opět místo protokolu HTTP volit protokol HTTPS.

Není-li možné využít doménu druhého řádu zřizovatele, může poskytovatel využít URL svého WWW portálu či prezentace. Potom je základem všech jeho IRI:

```
{homepage-URL}/data
```

Prvek {homepage-URL} je URL domácí stránky WWW portálu či prezentace poskytovatele.

IRI entit pak mají tvar

```
{základ-IRI}/resource/{určení-entity}
```

Prvek {základ-IRI} je základ všech IRI poskytovatele. Prvek {určení-entity} volí poskytovatel dat libovolně tak, aby výsledné IRI bylo validním IRI a identifikovalo entitu jednoznačně. Tvar tohoto prvku není dále upravován a je tak plně v kompetenci poskytovatele dat.

### 14.4.4 Standardy propojených dat

Standardy propojených dat vyžadují, aby pro IRI entit byly navíc splněny následující podmínky:

- IRI entit jsou dereferencovatelná, tj. při požadavku na přístup k IRI entity prostřednictvím HTTP(S) protokolu jsou vráceny údaje, které se entity týkají. Údaje jsou reprezentovány v datovém modelu RDF a zapsány ve formátech určených pro zápis tohoto datového modelu (zejména Turtle, TriG, nebo JSON-LD). Vráceny jsou všechny údaje o entitě vyskytující se ve všech datových sadách poskytovatele.
- Mezi údaji vrácenými při požadavku na přístup k IRI entity jsou v datovém modelu RDF vráceny i RDF trojice, v jejichž subjektech či objektech se vyskytují IRI entit jiného poskytovatele.

V souvislosti s bodem 2 je standardem explicitně vyjadřovat ekvivalenční vazbu mezi různými IRI vytvořenými různými poskytovateli, které však označují stejnou entitu. Pokud poskytovatel zjistí, že vytváří IRI pro entity, pro které již vytvořil IRI i jiný poskytovatel, zveřejní datovou sadu na stupni otevřenosti 5, která obsahuje ekvivalenční vazby mezi jeho IRI a IRI jiného poskytovatele. Ekvivalenční vazba mezi entitou A a B je vyjádřena v podobě následující RDF trojice:

- subjekt : URI entity A
- predikát : <http://www.w3.org/2002/07/owl#sameAs>
- objekt : URI entity B

## 14.5 Technické standardy vytváření datových schémat pro datové sady na stupni otevřenosti 5

Obdobou datového schématu pro datové sady na stupni otevřenosti 5 jsou tzv. ontologie (někdy též zvané slovníky). Zatímco však ve světě tabulkových či XML dat popisuje datové schéma striktní syntaxi např. CSV či XML distribucí datových sad, ontologie popisuje datovou sadu na konceptuální úrovni. Konkrétně definuje třídy a vlastnosti, které lze v RDF distribucích datových sad použít.

### 14.5.1 Standardy pro vyjádření a zápis ontologií

Standardem pro zápis ontologií jsou jazyky RDF Schema (RDFS, <http://www.w3.org/TR/rdf-schema/>) a Web Ontology Language (OWL, <http://www.w3.org/TR/owl/>). Oba jazyky předpokládají ontologii reprezentovanou v datovém modelu RDF a umožňují využít pro zápis ontologie libovolný formát pro zápis RDF datového modelu (viz [Technické standardy pro datové sady na stupni otevřenosti 4 a 5](#)).

### 14.5.2 Standardy pro tvorbu ontologií

Primárním cílem při návrhu tříd a vlastností použitých v RDF distribuci datové sady musí být maximální znovupoužitelnost údajů v distribuci napříč různými SW aplikacemi. Proto je standardem co nejvíce využívat tříd a vlastností z již existujících ontologií. Datový model RDF je připraven k tomu, aby bylo možné v jedné RDF distribuci kombinovat třídy a vlastnosti z několika různých ontologií. V rámci různých iniciativ již vznikla v mezinárodním kontextu celá řada ontologií, které jsou katalogizovány např. v projektu Linked Open Vocabularies (LOV, <http://lov.okfn.org>).

Při návrhu tříd a vlastností použitých v RDF distribuci je proto standardem postupovat následovně:

- Identifikujeme typy entit vyskytujících se v datové sadě a jejich atributy a vztahy mezi nimi. (Jinými slovy tvoříme konceptuální schéma datové sady.)
- Vybereme třídy a vlastnosti definované existujícími ontologiemi, jejichž sémantika přímo odpovídá sémantice typů entit a jejich atributům a vztahům mezi nimi, které jsme identifikovali v kroku 1.
- V případě, že se v kroku 2 nepodařilo pro některé typy entit, resp. jejich atributy nebo vztahy mezi nimi nalézt odpovídající třídy či vlastnosti z existujících ontologií, vybereme třídy a vlastnosti definované existujícími ontologiemi, jejichž sémantika je obecnější než sémantika těchto zbylých typů entit, resp. jejich atributů nebo vztahů mezi nimi. Pro tyto zbylé typy entit, resp. jejich atributy a vztahy mezi nimi vytvoříme nové třídy, resp. vlastnosti v naší vlastní ontologii a propojíme je s vybranými třídami a vlastnostmi z existujících ontologií s obecnější sémantikou pomocí ISA vazby (pro technické vyjádření ISA vazby viz dále).
- V případě, že se v kroku 3 nepodařilo pro některé typy entit, resp. jejich atributy nebo vztahy mezi nimi nalézt třídy či vlastnosti z existujících ontologií s obecnější sémantikou, vytvoříme pro ně nové třídy, resp. vlastnosti v naší vlastní ontologii.
- V případě, že po vy publikování RDF distribuce zjistíme, že jsme v kroku 2 udělali chybu a zavedli jsme třídu či vlastnost pro typ entity, resp. atribut či vztah, pro níž existuje sémantický ekvivalent v existující ontologii, nebo zjistíme, že se objevila nová ontologie s takovým sémantickým ekvivalentem, propojíme naši třídu či vlastnost s jejím nalezeným sémantickým ekvivalentem pomocí ekvivalenční vazby (pro technické vyjádření ekvivalenční vazby viz dále).
- V případě, že po vy publikování RDF distribuce zjistíme, že jsme v kroku 3 udělali chybu a neidentifikovali ISA vazbu, nebo se objevila nová ontologie pro kterou ISA

vazba existuje, tuto ISA vazbu vytvoříme (pro technické vyjádření ISA vazby viz dále).

Standardem je využívání především následujících ontologií:

- DCMI Metadata Terms (<http://purl.org/dc/terms/>)
- Simple Knowledge Organization System (<http://www.w3.org/2004/02/skos/core#>)
- Schema.org (<http://schema.org/>)
- Data Cube Vocabulary (<http://www.w3.org/TR/vocab-data-cube/>)
- Friend-of-a-Friend (<http://xmlns.com/foaf/0.1/>)
- The organization ontology (<http://www.w3.org/TR/vocab-org/>)
- FRBR-aligned bibliographic ontology (<http://purl.org/spar/fabio/>)

### 14.5.3 Standardy propojování prvků ontologií

Mezi prvky ontologií (tj. mezi třídami nebo mezi vlastnostmi) mohou existovat sémantické vazby. Může se jednat o vazby mezi prvky stejné ontologie nebo mezi prvky dvou různých ontologií. Tyto vazby je nutno v ontologiích vyjádřit. Vazbu vždy vyjádříme v ontologii, která definuje prvek, pro který chceme vazbu vyjádřit. Pro vyjádření vazby postupujeme podle následujícího standardu:

- pro vyjádření ISA vazby mezi třídami A a B, kdy třída A má specifičtější sémantiku a třída B má obecnější sémantiku, vytvoříme v zápisu ontologie následující trojici
  - subjekt : IRI třídy A
  - predikát : `http://www.w3.org/2000/01/rdf-schema#subClassOf`
  - objekt : IRI třídy B
- pro vyjádření ISA vazby mezi vlastnostmi A a B, kdy vlastnost A má specifičtější sémantiku a vlastnost B má obecnější sémantiku, vytvoříme v zápisu ontologie následující trojici
  - subjekt : IRI vlastnosti A
  - predikát : `http://www.w3.org/2000/01/rdf-schema#subPropertyOf`
  - objekt : IRI vlastnosti B
- pro vyjádření ekvivalenční vazby mezi třídami A a B, kdy sémantika třídy A je ekvivalentní sémantice třídy B, vytvoříme v zápisu ontologie následující trojici
  - subjekt : IRI třídy A
  - predikát : `http://www.w3.org/2002/07/owl#equivalentClass`
  - objekt : IRI třídy B
- pro vyjádření ekvivalenční vazby mezi vlastnostmi A a B, kdy sémantika vlastnosti A je ekvivalentní sémantice vlastnosti B, vytvoříme v zápisu ontologie následující trojici
  - subjekt : IRI vlastnosti A
  - predikát : `http://www.w3.org/2002/07/owl#equivalentProperty`
  - objekt : IRI vlastnosti B

### 14.5.4 Standardy tvorby IRI tříd a vlastností definovaných v ontologiích

V případě, že poskytovatel dat definuje vlastní predikáty či třídy v podobě ontologie, potom pro ně musí volit IRI ve tvaru URL v následující podobě:

- v případě, že poskytovatel zavádí pouze jednu ontologii, má ontologie IRI ve tvaru {základ-IRI}/ontology, kde {základ-IRI} je základ IRI poskytovatele dat (viz [Technické standardy pro datové sady na stupni otevřenosti 4 a 5](#))
- v případě, že poskytovatel zavádí více ontologií, mají ontologie IRI ve tvaru {základ-IRI}/ontology/{určení-ontologie} kde {určení-ontologie} je

validní část IRI, která jednoznačně určuje ontologii v kolekci ontologií definovaných poskytovatelem

- třídy mají IRI ve tvaru {IRI-ontologie}/{NázevTřídy} kde {IRI-ontologie} je IRI ontologie, ve které je třída definována a {NázevTřídy} je název třídy unikátní v rámci kolekce všech tříd zavedených v ontologii zapsaný v CamelCase notaci (tj. jednotlivá slova tvořící název třídy jsou spojena za sebe, první znak každého slova je velký a všechny ostatní znaky jsou malé)
- predikáty mají IRI ve tvaru {IRI-ontologie}/{názevPredikátu} kde {IRI-ontologie} je IRI ontologie, ve které je třída definována a {názevPredikátu} je název predikátu unikátní v rámci kolekce všech predikátů zavedených v ontologii zapsaný v camelCase notaci (tj. jednotlivá slova tvořící název třídy jsou spojena za sebe, první znak každého slova kromě prvního slova je velký a všechny ostatní znaky jsou malé)

## 14.6 Technické standardy pro aplikační rozhraní lokálních katalogů

Aby mohl být lokální katalog registrován v NKOD, musí nabízet aplikační rozhraní (API) splňující následující specifikaci. API je webová služba obsahující funkce **package\_list** a **package\_show** ze specifikace [CKAN API verze 3](#). Výstup funkce `package_show` je rozšířen o dodatečná metadata. Výstupy obou funkcí jsou ve formátu [JSON \(ECMA-404\)](#).

Obě funkce musí být přístupné na adrese lokálního katalogu pod pevně stanovenými URL.

Příklad: lokální katalog běží na adrese <http://data.organizace.cz>. Funkce **package\_list** je pak přístupná na adrese [http://data.organizace.cz/api/3/action/package\\_list](http://data.organizace.cz/api/3/action/package_list) a funkce **package\_show** je pak přístupná na adrese [http://data.organizace.cz/api/3/action/package\\_show?id=datova-sada-1](http://data.organizace.cz/api/3/action/package_show?id=datova-sada-1) kde „datova-sada-1“ je ID konkrétní datové sady v lokálním katalogu sestávající z malých písmen bez diakritiky, čísel, podtržítka a pomlček.

### 14.6.1 package\_list

Odpověď na volání funkce **package\_list** obsahuje v poli **result** seznam ID všech datových sad obsažených v katalogu. Tato ID se pak použijí pro volání funkce **package\_show**. Odpověď má tvar:

```
{
  "success": true,
  "result": [
    "datova-sada-1",
    "datova-sada-2",
    "datova-sada-3"
  ]
}
```

### 14.6.2 package\_show

Odpověď na volání funkce **package\_show** s parametrem `id` obsahujícím ID datové sady obsahuje mimo jiné [kompletní metadata datové sady a všech jejích distribucí](#). Pole **result** obsahuje mimo jiné následující parametry:

Název pole	Název položky	Datový typ	Povinnost	Příklad hodnoty
title	Název datové sady	Řetězec	Ano	Kontroly
notes	Popis datové sady	Řetězec	Ano	Seznam kontrol od 1.1.2012
frequency	Periodicita aktualizace	Řetězec dle ISO 8601	Ano	R/P3M
ruian_type	Související geografické území - typ prvku RÚIAN	Typ prvku RÚIAN	Ano	ST
ruian_code	Související geografické území - kód prvku RÚIAN	Řetězec	Ano	1
theme	Klasifikace dle EuroVoc	Seznam URL oddělený mezerou	Ne	<a href="http://eurovoc.europa.eu/171">http://eurovoc.europa.eu/171</a>
maintainer_email	Email Kurátora dat	E-mail	Ne	verejnadata@INSTITUTE.cz
maintainer	Jméno Kurátora dat	Řetězec	Ne	Jan Novák
schema	Odkaz na dokumentaci datové sady	URL	Ne	<a href="http://www.INSTITUTE.cz/open-data/dokumentace-ds1.html">http://www.INSTITUTE.cz/open-data/dokumentace-ds1.html</a>
tags	Klíčová slova	pole objektů	Ano	viz níže
temporal_start	Dotčené časové období od	datum ve formátu xsd:date	Ne	2012-01-01
temporal_end	Dotčené časové období do	datum ve formátu xsd:date	Ne	2014-12-31
publisher_uri	URL poskytovatele dat	URL	Ano	<a href="http://www.INSTITUTE.cz">http://www.INSTITUTE.cz</a>
publisher_name	Název poskytovatele dat	Řetězec	Ano	INSTITUTE
resources	Datové zdroje	pole objektů	Ano	viz níže



Navíc pole result obsahuje:

- položku name obsahující ID datové sady použité jako parametr funkce **package\_show** a zobrazené v seznamu vráceném funkcí **package\_list**.

Příklad:

```
"name": "coi-kontroly"
```

- Pole klíčových slov, kde každé je reprezenováno objektem o dvou položkách se stejným obsahem **name** a **display\_name**.

Příklad:

```
"tags": [{
  "display_name": "kontroly",
  "name": "kontroly"
}],
```

- Pole obsahující metadata jednotlivých datových zdrojů **resources**, viz popis distribuce.

V poli resources má každá distribuce (datový zdroj) mimo jiné následující parametry:

Název pole	Název položky	Datový typ	Povinnost	Příklad hodnoty
url	Odkaz na stažení datového zdroje	URL	Ano	<a href="http://www.INSTITUCE.cz/open-data/kontroly.trig">http://www.INSTITUCE.cz/open-data/kontroly.trig</a>
format	Formát datového zdroje	MIME typ	Ano	application/trig
license_link	Odkaz na podmínky užití	URL	Ano	<a href="https://portal.gov.cz/portal/ostatni/vo-lny-pristup-k-ds.html">https://portal.gov.cz/portal/ostatni/vo-lny-pristup-k-ds.html</a>
name	Název datového zdroje	Řetězec	Ne	Kontroly
describedBy	Odkaz na strojově čitelné schéma datového zdroje	URL	Ne	<a href="http://www.INSTITUCE.cz/open-data/ontologie.trig">http://www.INSTITUCE.cz/open-data/ontologie.trig</a>
describedByType	Formát schématu datového zdroje	MIME typ	Ne	application/trig
temporal_start	Dotčené časové období	datum ve formátu	Ne	2015-01-01

Název pole	Název položky	Datový typ	Povinnost	Příklad hodnoty
	od	xsd:date		
temporal_end	Dotčené časové období do	datum ve formátu xsd:date	Ne	2015-12-31

## 16 Přílohy

### 16.1 Vzor směrnice "Publikace a katalogizace otevřených dat"

SMĚRNICE Č. ?SUBJEKT?

## Směrnice

### Publikace a katalogizace otevřených dat

#### (vzor k volnému použití)

Poznámky k použití a přizpůsobení vzoru směrnice:

- Vzorová směrnice je úředním dílem a s jako takovým je možné s ním volně nakládat;
- vzor směrnice je nutno přizpůsobit k použití v konkrétním orgánu veřejné správy;
- je třeba minimálně nahradit (nebo vypustit) formulace uzavřené dvěma „?“;
- role uvedené ve směrnici je možné sloučit nebo obsadit jedním útvarem nebo zaměstnancem.
- Národní katalog otevřených dat veřejné správy ČR (NKOD), ve kterém je nutno každou publikovanou datovou sadu katalogizovat, umožňuje katalogizaci následujícími způsoby:
  - vyplnění formuláře na PVS a jeho odeslání přes Informační systém datových schránek, součástí katalogizačního záznamu mohou být i vlastní otevřená data, která pak budou přístupná prostřednictvím NKOD,
  - automatické plnění metadat z lokálních katalogů OVM (po předchozí registraci lokálního katalogu v NKOD).

## Čl. 1

### Vymezení působnosti směrnice

- 1) Účelem směrnice je stanovit jednotný postup při publikaci a katalogizaci otevřených dat v ?SUBJEKT\_CELÝM\_NÁZVEM? (dále jen „?SUBJEKT?“). Směrnice vymezuje základní pravidla, kompetence a odpovědnosti při publikaci a katalogizaci otevřených dat v podmínkách ?SUBJEKT?.
- 2) Směrnice je závazná pro zaměstnance ?SUBJEKT?.

## Čl. 2

### Definice pojmů

- 1) Za otevřená data se považují informace poskytované tak, že jsou poskytovány:
  - a) v maximálním možném rozsahu,
  - b) způsobem umožňujícím dálkový přístup,
  - c) ve strojově čitelném a otevřeném formátu,
  - d) za jasně definovaných podmínek užití neomezujících způsob ani účel následného využití poskytovaných informací včetně umožnění jejich využití pro přímý a nepřímý hospodářský prospěch,
  - e) dostupné 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 jednou nebo více distribucemi, tj. datovými soubory nebo prostřednictvím definovaného programového rozhraní (API).
- 3) Distribucí datové sady se rozumí datový zdroj v podobě elektronického souboru, který zpřístupňuje data datové sady nebo jejich část.
- 4) Otevřená datová sada je datová sada, která splňuje podmínky otevřených dat.
- 5) Stupně otevřenosti dat vyjadřují míru otevřenosti dat. Míra je vyjádřena pomocí následujících stupňů:
  - a) stupeň 0, při kterém datová sada existuje v elektronické podobě, ale není dostupná způsobem umožňujícím dálkový přístup nebo nemá specifikovány podmínky užití otevřených dat,
  - b) stupeň 1, při kterém datová sada existuje v elektronické podobě, je dostupná způsobem umožňujícím dálkový přístup a má specifikovány podmínky užití otevřených dat,
  - c) stupeň 2, při kterém jsou splněny požadavky písmene b) a distribuce datové sady jsou reprezentovány ve strojově čitelném formátu,
  - d) stupeň 3, při kterém jsou splněny požadavky písmene c) a distribuce datové sady jsou poskytovány v otevřeném formátu,
  - e) stupeň 4, při kterém jsou splněny požadavky písmene d) a na identifikaci entit v distribuci datové sady se používají identifikátory ve tvaru Internationalized Resource Identifier (IRI),
  - f) stupeň 5, při kterém jsou splněny požadavky písmene e) a distribuce datové sady splňuje standard propojených dat.

- 6) Transformace dat je změna obsahu a struktury datové sady, např. anonymizace, agregace či jiná úprava původních dat.
- 7) Katalogizační záznam tvoří strukturovaná data o určité otevřené datové sadě (metadata) v katalogu otevřených dat.
- 8) Metadaty datové sady se rozumí data popisující datovou sadu, zejména její věcný obsah, časové, územní a další souvislosti,
- 9) Metadaty distribuce datové sady se rozumí data popisující distribuci datové sady, zejména její strukturu, časové a další souvislosti,
- 10) Katalog otevřených dat je softwarová aplikace umožňující zobrazení, vytváření, úpravu, správu a případné rušení (zneplatnění) katalogizačních záznamů. Katalog otevřených dat umožňuje katalogizační záznamy prezentovat v plném rozsahu uživatelům sítě Internet, včetně funkcionality vyhledávání a získání přehledového seznamu všech katalogizačních záznamů.
- 11) Národní katalog otevřených dat veřejné správy ČR (dále „NKOD“) je centrálním katalogem otevřených dat v ČR. NKOD je dostupný prostřednictvím Portálu veřejné správy ČR (dále „PVS“). NKOD obsahuje katalogizační záznamy o datových sadách zveřejněných jednotlivými subjekty veřejné správy ČR, včetně odkazů ke stažení dat. Data ke stažení mohou být uložena na webu příslušného Poskytovatele dat nebo přímo na PVS v NKOD.
- 12) Lokální katalog otevřených dat je softwarová aplikace, která je volitelně implementována pro potřeby katalogizace otevřených dat ?SUBJEKT?.
- 13) Publikační plán vymezuje datové sady plánované k publikaci v podobě otevřených dat. Obsahem publikačního plánu je seznam datových sad. Publikační plán je aktualizován v roční periodě. Pro každou datovou sadu jsou zejména uvedeny:
- název datové sady;
  - popis datové sady;
  - Kurátor dat zodpovědný za analýzu a přípravu datové sady;
  - přínosy a rizika publikace datové sady ve formátu otevřených dat;
  - periodicita publikace (aktualizace) datové sady;
  - podmínky užití datové sady;
  - plánovaný termín první publikace datové sady.
  -

### ČI. 3 Role a obsazení

- 1) Pro práci s otevřenými daty jsou rozlišovány role uvedené v následující tabulce:

Role	Obsazení
Vedení poskytovatele dat	Vedení ?SUBJEKT?
Koordinátor otevírání dat	?ÚTVAR POVĚŘENÝ KOORDINACÍ OTEVÍRÁNÍ DAT?
Kurátor dat	Určený zaměstnanec příslušného útvaru ?SUBJEKT?, který zodpovídá za správu příslušné datové sady, resp. datových sad



	určených k publikaci ve formě otevřených datových sad Pozn.: Obsazení role je definováno v publikačním plánu.
Správce katalogu otevřených dat	?ÚTVAR POVĚŘENÝ SPRÁVOU KATALOGU OTEVŘENÝCH DAT?
IT specialista	?ÚTVAR POVĚŘENÝ PODPOROU PUBLIKACE A KATALOGIZACE OTEVŘENÝCH DAT?

#### Čl. 4

##### Obecná povinnost

Všichni zaměstnanci určení do některé z rolí uvedených v článku 3 jsou povinni pro práci s otevřenými daty zohlednit platnou metodiku Ministerstva vnitra uveřejněnou na <http://opendata.gov.cz>.

#### Čl. 5

##### Vedení poskytovatele dat

Vedení poskytovatele dat zodpovídá za:

- rozhodnutí o otevírání dat;
- obsazení role Koordinátora otevírání dat;
- obsazení dalších rolí v kontextu otevřených dat;
- schválení publikačního plánu a jmenování Kurátorů dat;
- periodické vyhodnocení plnění publikačního plánu.

#### Čl. 6

##### Koordinátor otevírání dat

Koordinátor otevírání dat zodpovídá za:

- přípravu publikačního plánu;
- návrh Kurátorů dat pro jednotlivé datové sady;
- poskytování metodické podpory Kurátorovi dat při plnění povinností plynoucích z této směrnice;
- zveřejnění publikačního plánu;
- rozhodnutí o způsobu katalogizace (NKOD nebo i lokální katalog otevřených dat);
- zajištění finální kontroly formátu a struktury dat před publikací;
- zajištění publikace datových sad a katalogizačních záznamů (metadat);
- komunikaci připomínek uživatelů otevřených dat a reportování o publikaci otevřených dat vedení poskytovatele a příp. tiskovému oddělení Poskytovatele dat.

#### Čl. 7

##### Kurátor dat

1) Kurátor dat zodpovídá za:

- zajištění analýzy datové sady předcházející zveřejnění dat;
- zajištění, že otevřené datové sady budou publikovány v souladu s platnými právními předpisy ČR a standardy publikace otevřených dat Ministerstva vnitra (uvedené na <http://opendata.gov.cz>);
- určení případné nutnosti transformace datových sad, např. anonymizaci, agregaci nebo jinou změnu struktury datových sad;

- d) správu a kontrolu kvality datových sad (tj. mj. kontrolu správnosti, úplnosti, konzistentnosti) před jejich publikací a katalogizací;
  - e) spolupráci při procesu zveřejnění datové sady s Koordinátorem otevírání dat a IT specialistou;
  - f) předání datových souborů s datovými sadami a návrhu obsahu katalogizačních záznamů (metadat) Správci katalogu otevřených dat.
- 2) Kurátor dat je povinen poskytnout IT specialistovi a Správci katalogu otevřených dat potřebnou součinnost při:
    - a) transformaci datové sady a její kontrole;
    - b) určení pracnosti a přínosů plynoucích z publikace otevřených datových sad;
    - c) vytváření a aktualizaci katalogizačních záznamů.
  - 3) Kurátora dat za příslušný útvar a příslušnou datovou sadu, resp. datové sady, určí Vedení poskytovatele dat na základě návrhu Koordinátora otevírání dat.

### Čl. 8

#### Správce katalogu otevřených dat

- 1) Správce katalogu otevřených dat zodpovídá za:
  - a) formální kontrolu a správu katalogizačních záznamů (metadat);
  - b) vkládání katalogizačních záznamů a příp. i datových souborů do NKOD;
  - c) správu katalogizačních záznamů v NKOD.
- 2) ?VOLITELNĚ? Správce katalogu otevřených dat dále zodpovídá za:
  - a) zajištění výběru softwarového nástroje pro lokální katalog otevřených dat;
  - b) zajištění implementace lokálního katalogu otevřených dat;
  - c) registraci lokálního katalogu v NKOD;
  - d) archivaci katalogizačních záznamů a dat uložených v lokálním katalogu otevřených dat.

### Čl. 9

#### IT specialista

IT specialista zodpovídá za:

- a) poskytování technické podpory a součinnosti Kurátorovi dat při plnění povinností plynoucích z této směrnice, zejména při analýze předcházející zveřejnění dat a přípravě datových sad k otevření;
- b) ?VOLITELNĚ? zajištění instalace a provozu lokálního katalogu otevřených dat.
- c)

### Čl. 10

#### Postup publikace a katalogizace otevřených dat

- 1) Postup tvorby publikačního plánu:
  - a) Koordinátor otevírání dat zajišťuje přípravu a aktualizaci publikačního plánu a navrhuje Vedení poskytovatele dat obsazení Kurátorů dat pro jednotlivé datové sady.
  - b) Koordinátor otevírání dat provádí ve spolupráci s Kurátory dat a IT specialisty analýzu pro potřeby sestavení publikačního plánu, která zahrnuje:
    - i) popis datových sad ve struktuře definované publikačním plánem, včetně návrhu podmínek užití jednotlivých datových sad;
    - ii) ověření, že publikace datových sad bude v souladu s platnou legislativou ČR.

- c) Koordinátor otevírání dat předkládá publikační plán ke schválení Vedení poskytovatele.
  - d) Vedení poskytovatele dat schvaluje publikační plán.
  - e) Koordinátor otevírání dat ve spolupráci s IT specialistou rozhoduje o cílovém stupni otevřenosti dat, formátech publikace dat a způsobu katalogizace otevřených dat.
  - f) Koordinátor otevírání dat zveřejňuje aktuální publikační plán.
  - g) Koordinátor otevírání dat periodicky (minimálně jednou ročně) vyhodnocuje plnění publikačního plánu a informuje Vedení poskytovatele dat.
- 2) Postup zveřejnění datové sady v otevřené podobě:
- a) Kurátoři dat provádějí ve spolupráci s Koordinátorem otevírání dat a IT specialisty analýzu datové sady pro potřeby publikace dat, která zahrnuje:
    - i) stanovení maximálního možného rozsahu publikovaných dat, který neklade nepřiměřené náklady na proces zveřejnění;
    - ii) specifikaci požadavků na kvalitu dat, včetně vyloučení překryvu datových sad;
    - iii) ověření kvality datové sady, zejména z hlediska úplnosti, konzistence a správnosti dat, příp. popis zjištěných nedostatků včetně jejich četnosti a závažnosti;
    - iv) navržení opatření pro zvýšení kvality datové sady na úroveň specifikovaných požadavků na kvalitu dat, včetně možných opatření pro zvýšení kvality dat v primární evidenci nebo databázi;
    - v) určení případné nutnosti transformace dat a určení optimálního technického zajištění publikace datové sady;
    - vi) návrh změn ve struktuře datové sady tak, aby data byla vhodně strukturovaná pro publikaci v cílovém formátu;
    - vii) návrh obsahu katalogizačního záznamu (metadat).
  - b) Koordinátor otevírání dat provádí ve spolupráci s IT specialistou přípravu datové sady k publikaci, která zahrnuje:
    - i) návrh, implementaci a testování skriptů pro převod dat do cílového formátu;
    - ii) technické zajištění samotné publikace, zejména transformace datové sady (pokud je nutno ji provádět) a převod datové sady (v rozsahu její stávající velikosti) do cílového formátu;
    - iii) kontrolu transformace a kvality datové sady;
    - iv) dokumentaci postupu publikace;
    - v) přípravu katalogizačního záznamu.
  - c) Koordinátor otevírání dat provádí finální kontrolu formátu a struktury datové sady a katalogizačního záznamu před zveřejněním.
  - d) Koordinátor otevírání dat ve spolupráci s IT specialistou a Správcem katalogu otevřených dat zajistí zveřejnění a katalogizaci datové sady zahrnující zejména:
    - i) předání datových souborů s metadaty Správci katalogu otevřených dat;
    - ii) předání datového souboru s otevřenými datovými sadami k publikaci;
    - iii) ověření správnosti a klasifikace katalogizačního záznamu, případné opravy katalogizačního záznamu;
    - iv) zajištění zaevidování datové sady v NKOD;
    - v) **?VOLITELNĚ?** zaevidování katalogizačního záznamu v lokálním katalogu otevřených dat.

- 3) ?VOLITELNĚ? Postup zajištění lokálního katalogu otevřených dat zajišťuje Správce katalogu otevřených dat ve spolupráci s IT specialistou:
- instalace lokálního katalogu otevřených dat;
  - příprava lokálního katalogu otevřených dat;
  - provozování lokálního katalogu otevřených dat.

### **Čl. 11**

#### **Katalog otevřených dat**

- Všechna publikovaná otevřená data ?SUBJEKT? musí být katalogizována v NKOD.
- ?VOLITELNĚ? Pro zajištění snadnějšího přístupu k otevřeným datům ?SUBJEKT? je na jeho webových stránkách zpřístupněn lokální katalog otevřených dat. Obsah lokálního katalogu otevřených dat ?SUBJEKT?, tj. katalogizační záznamy, je publikován též jako otevřená data. Obsah lokálního katalogu otevřených dat ?SUBJEKT? je vedle formátu HTML dostupný také ve strojově čitelném formátu založeném na formátu DCAT-AP.

### **Čl. 12**

#### **Závěrečná ustanovení**

Tato směrnice nabývá účinnosti dnem vydání.