



# **Metodika identifikace, publikace, katalogizace a využívání otevřených dat pro samosprávu a pro státní správu**

Vytvořeno v rámci projektu

**Rozvoj datových politik v oblasti zlepšování kvality a  
interoperability dat veřejné správy**

**CZ.03.4.74/0.0/0.0/15\_025/0013983**

**Klíčová aktivita:** 02 Rozvoj Národního katalogu otevřených dat

**Verze výstupu:** 01



<b>1</b>	<b>5</b>
1.1	5
1.2	6
1.3	6
1.4	7
1.5	9

## **ČÁST 1. METODIKA**

**11**

<b>2</b>	<b>11</b>	
2.1	11	
2.1.1	Národní katalog otevřených dat (NKOD)	11
2.1.2	Veřejný datový fond	12
2.1.3	Registr práv a povinností (RPP)	12
2.1.4	Nástroje pro monitorování kvality publikace datových sad	12
2.1.5	Otevřené formální normy (OFN)	13
2.1.6	Vzdělávací materiály a e-learningová školení	13
2.2	14	
2.2.1	Organizační zabezpečení publikace poskytovatelem	15
2.2.2	Lokální katalog otevřených dat (LKOD)	18
2.2.3	Zajištění údržby publikovaných otevřených dat	20
<b>3</b>	<b>21</b>	
3.1	21	
3.2	22	
3.3	23	
3.4	25	
3.5	25	
<b>4</b>	<b>25</b>	
4.1	25	
4.1.1	Sémantická a syntaktická interoperabilita publikovaných dat.	25
4.1.2	Kontext tvorby datového schématu	27
4.2	30	
4.2.1	Vytvoření konceptuálního modelu datové sady.	29
4.2.2	Výběr způsobu publikace	32
4.2.3	Přiřazení rolí k činnostem přípravy konceptuálního modelu datové sady	34
4.3	36	
4.3.1	Výběr způsobu reprezentace strukturovaných údajů – datových modelů	35
4.3.2	Obecné charakteristiky publikované datové sady	35
4.3.3	Tabulková data (relační datový model)	36
4.3.4	Hierarchicky strukturovaná data (hierarchický datový model)	36
4.3.5	Propojená data (grafový datový model)	43
4.4	45	
4.4.1	Publikace obrázků	45
4.4.2	Publikace textových dokumentů	45
4.4.3	Publikace prostorových údajů	46
4.5	47	
4.5.1	Doplnění vazeb do SSP (sémantika datové sady na úrovni metadat)	47
4.5.2	Příprava katalogizačního záznamu datové sady	47



<b>5</b>	<b>49</b>	
5.1	50	
5.1.1	Příprava distribucí se strukturovanými daty	48
5.1.2	Příprava distribucí s nestrukturovanými daty	50
5.1.3	Příprava distribucí k publikaci do VDF	50
5.2	52	
5.3	55	
5.3.1	Zveřejnění distribucí datové sady a souborů s definicemi datového schématu každé distribuce	53
5.3.2	Finalizace katalogizačních záznamů	54
5.3.3	Katalogizace v LKOD	55
5.4	57	
5.5	57	
<b>6</b>	<b>58</b>	
6.1	59	
6.1.1	Publikace registrovaných veřejných údajů do VDF	57
6.1.2	Publikace ostatních údajů ve formátu otevřených dat	58
6.1.3	Publikace High Value Datasets (HVDS)	58
6.2	62	
6.2.1	Publikace otevřených dat zaměřených na zlepšení kvality života občanů	61
6.2.2	SMART koncepty	62
6.2.3	Publikace otevřených dat vyplývající z legislativy	64

---

## **ČÁST 2. KOMUNIKACE A PROPAGACE** **65**

<b>7</b>	<b>67</b>	
7.1.1	Komunikační strategie	65
7.1.2	Propagace otevřených dat	65
7.1.3	Prezentace otevřených dat	67
<b>8</b>	<b>69</b>	
8.1	70	
8.2	70	
8.3	71	
8.4	71	
8.5	72	
8.6	72	
8.7	73	

---

## **ČÁST 3. ZNALOSTNÍ** **72**

<b>9</b>	<b>74</b>	
9.1	74	
9.1.1	Základní charakteristiky otevřených dat.	72
9.1.2	Zastřešující koncept / filozofie otevřených dat – otevřenost.	72
9.1.3	Interoperabilita otevřených dat.	73
9.1.4	Principy otevřených dat (Mezinárodní charta otevřených dat, 2015).	74
9.1.5	Kvalita otevřených dat a kvalita jejich publikace	75
9.1.6	FAIR principy	75



9.1.7	Způsoby vzniku hodnoty na základě otevřených dat (zhodnocování).	76
9.1.8	Význam a role otevřených dat	76
9.2	78	
9.2.1	Rozdělení údajů z pohledu jejich sdílení	77
9.2.2	Přístupy k údajům veřejné správy	78
9.2.3	Registrované veřejné údaje	79
9.3	81	
9.3.1	Určení přínosů datových sad	80
9.3.2	Určení rizik datových sad	81
9.3.3	Ošetření rizik	82
9.4	84	
9.4.1	Výklad problematiky vymezení věcného obsahu datové sady	83
9.4.2	Stanovení stupně otevřenosti	85
9.5	88	
9.5.1	OFN	87
9.6	98	
9.6.1	Vytvoření RDF modelu	96
9.6.2	Serializace RDF modelu	97
9.6.3	Pravidla pro výběr slovníků pro vybrané typy dat	98
9.6.4	Zpřístupnění publikovaných propojených dat	99
9.7	102	
9.7.1	Slovník pojmů (Glosář, Tezaurus).	100
9.7.2	Konceptuální model.	100
9.7.3	Sémantický slovník	101
9.7.4	Analýza agendy a odvození údajů	103
9.7.5	Doplnění entit, atributů a vazeb mezi entitami dle datové sady	103
9.7.6	Správa Sémantického slovníku veřejné správy.	103
9.7.7	Podpora nástroji	103
9.7.8	Doplňující zdroje	104
9.8	107	
9.8.1	Metadata záznamu datové sady a pravidla pro jejich vyplňování.	105
9.8.2	Metadata záznamu distribuce datové sady	107
9.8.3	Metadata záznamu datové série	107
9.9	111	
9.9.1	Implementace Lokálního katalogu otevřených dat	108
9.9.2	Registrace lokálního katalogu	109
9.9.3	Zrušení registrace lokálního katalogu	109
9.9.4	Otevřená formální norma "Rozhraní katalogů otevřených dat: DCAT-AP-CZ"	109
9.10	113	
9.10.1	Základní charakteristiky VDF	110
9.10.2	Základní logické komponenty VDF AIS a jejich význam	110
9.10.3	Společné logické komponenty VDF související s AIS Působnostním (RPP)	111
9.10.4	Význam metodiky pro publikaci do VDF	112
9.10.5	Publikace distribucí datové sady do datového úložiště agendy	113
9.10.6	Katalogizace v LKOD agendy	113
9.10.7	Informování notificačního hubu o změně publikovaného veřejného údaje	113
9.10.8	Doplnění evidence údajů v RPP	114



# 1 Úvod

Otevřená data již nejsou pouze nástrojem určeným k zajištění transparentnosti veřejné správy, ale staly se neoddelitelnou součástí budování eGovernmentu, tvoří základní stavební kameny celosvětově přijímaného konceptu otevřeného vládnutí (Open Governance), konceptů budování chytrých měst (SMART City), jsou součástí architektury veřejné správy ČR a základem budovaného jednotného digitálního trhu EU.

Dnešní svět a jeho prostředí je charakteristické svou komplexní provázaností, ve kterém často samotný jediný údaj postrádá větší vypovídající hodnotu a skutečná informace je skrytá až ve spojení s řadou dalších údajů.

Z pohledu OD je tedy „izolovaná“ publikovaná datová sada nedostačující, k zajištění skutečné transparentnosti a vytvoření správných informací s velkou hodnotou je nutné data uvádět v souvislostech, ve správných kontextech a významech. Do popředí se tak dostává požadavek na zajištění plné interoperability dat, což jednoduše znamená publikovat data takovým způsobem, aby bylo možné data navzájem sdílet a propojovat, a tak vytvářet správné kontexty zachycující realitu.

Naplnění uvedeného požadavku je prioritním cílem této metodiky.

Metodika zastřešuje všechny oblasti spojené s publikací otevřených dat ve veřejné správě ČR připravenými standardy, definovanými postupy, nástroji, a dalšími souvisejícími metodikami a dokumenty.

## 1.1 Struktura metodiky

Metodika je rozdělena do 3 částí:

- **Část první – Metodika**, představuje vlastní metodiku a obsahuje všechny postupy od prvotní identifikace témat k publikaci až po konečné zpřístupnění a správu publikovaných datových sad.
  - Řešené oblasti:
    - Prostředí publikace otevřených dat veřejné správy ČR.
    - Identifikace témat otevřených dat.
    - Příprava datové sady.
    - Publikace a katalogizace datové sady.
    - Specifika oblastí veřejné správy.
- **Část druhá – Komunikace, propagace a využití OD**, nemá přímý metodický charakter, neboť se již nezaměřuje na publikaci otevřených dat, ale na komunikaci, prezentaci publikovaných otevřených dat a možné přístupy při jejich využití. Tuto oblast je obtížné nějak standardizovat, proto jsou v této části uvedeny spíše pouze doporučení, rady a možné inspirace.
  - Řešené oblasti:
    - Prezentace a propagace dat.
    - Využití otevřených dat.
- **Část třetí – Znalostní**, je zaměřena na doplnění znalostí o otevřených datech. Jsou zde k dispozici doplňující a rozšiřující informace k postupům a standardům uvedeným v části první (metodice).
  - Řešené oblasti:
    - Rekapitulace informací o otevřených datech.
    - Údaje veřejné správy ČR.
    - Příklad ošetření navržených témat otevřených dat.
    - Datové schéma.
    - Datová specifikace.
    - Propojená data.



- Popis sémantiky veřejné správy.
- Katalogizační záznam.
- Stručný popis veřejného datového fondu.

## 1.2 Uživatelé metodiky

Metodika je primárně určena všem organizacím a pracovníkům veřejné správy ČR, zejména ale těm, kteří jsou pověřeni vykonáváním rolí uvedených v kapitole „Prostředí poskytovatele OD“.

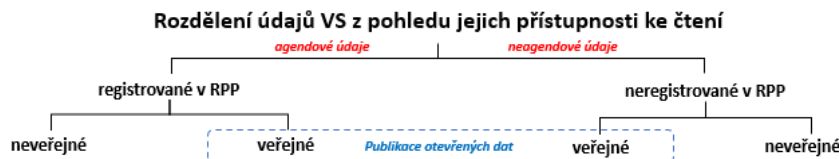
Postupy přípravy a publikace otevřených dat jsou pro všechny typy organizací veřejné správy shodné. Rozdíly se týkají pouze typů publikovaných témat a kategorie dat (agendové/neagendové), které vyplývají z různých pravomocí a odpovědností státní správy a samosprávy.

- **Státní správa** má za úkol zajišťovat plnění úkolů a povinností stanovených zákonem na celostátní úrovni, má zpravidla větší objem a rozmanitost dat, která jsou důležitá pro celostátní politiku a řízení, provozuje národní registry a databáze, a tak se využití otevřených dat zaměřuje na plnění legislativních povinností publikace veřejných údajů a zajištění transparentnosti a efektivity procesů na celostátní úrovni.
- **Samospráva** se zaměřuje na zajištění kvality života občanů na místní úrovni, na plnění místních potřeb a priorit, může mít menší objem dat, ale tyto data mohou být důležitá pro konkrétní region, obec nebo město, a tak se využití otevřených dat soustředí na zlepšení kvality života občanů v konkrétním regionu a na poskytování lepších místních služeb.
- **Jakýkoliv poskytovatel mimo veřejnou správu**, metodika je samozřejmě využitelná i v těchto případech (ideální cílová představa je založena na sdílení stejných standardů a postupů), neboť zejména v konceptech SMART se veřejná správa a soukromý sektor potkávají a integrace potřebných, avšak různorodých dat bývá častým řešeným problémem. Otevřená data představují ideální integrační platformu. Zatím neřešenou otázkou je případná katalogizace datových sad publikovaných mimo veřejnou správu v NKOD.

Některým specifikám publikace otevřených dat souvisejícími s typy organizací veřejné správy (státní správa / samospráva) se věnuje kapitola 6 „[Specifika oblastí veřejné správy](#)“.

## 1.3 Zaměření metodiky

Zaměření metodiky souvisí se členěním údajů veřejné správy a jejich [sdílením](#).



- Pro sdílení registrovaných údajů (veřejných i neveřejných) mezi OVM a zpřístupnění těchto údajů SPUÚ pro účely výkonu jejich agend, je určen propojený datový fond ([PPDF](#)). Ke sdílení i zpřístupnění údajů je nutná evidence oprávnění pro čtení nebo zápis v Registru práv a povinností (RPP) související s výkonem příslušných agend.
- Pro sdílení veřejných registrovaných údajů mezi OVM a zpřístupnění těchto údajů SPUÚ bez omezení přístupu je určen Veřejný datový fond ([VDF](#)). V tomto případě se jedná pouze o čtení údajů v podobě otevřených dat a není nutná evidence oprávnění přístupu v RPP vyplývající z výkonu souvisejících agend.
- Pro sdílení veřejných registrovaných i neregistrovaných údajů veřejné správy s veřejností bez omezení je určen Otevřený přístup v podobě otevřených dat.



Detailnímu vysvětlení rozdělení údajů veřejné správy se věnuje kapitola „[Údaje VS](#)“.

Metodika pokrývá:

- přípravu a publikaci neagendových (neregistrovaných v RPP) veřejných údajů veřejné správy ve formátu otevřených dat,
- přípravu agendových (registrovaných) veřejných údajů veřejné správy ve formátu otevřených dat pro publikaci do VDF,
  - specifiky publikace do VDF jsou upraveny dokumenty ([dostupnými na POD](#)):
    - Metodika definice údajů vedených v agendě.
    - Metodika poskytování dat ve veřejném datovém fondu.
    - Metodika tvorby a údržby sémantického slovníku pojmů veřejné správy.

## 1.4 Oblasti otevřených dat a jejich podchycení metodikou

Metodika rozlišuje a metodicky pokrývá následující oblasti otevřených dat.

**Prostředí publikace otevřených dat**, které je určeno vytvoření podmínek pro publikaci otevřených v prostředí veřejné správy ČR a dělí se na:

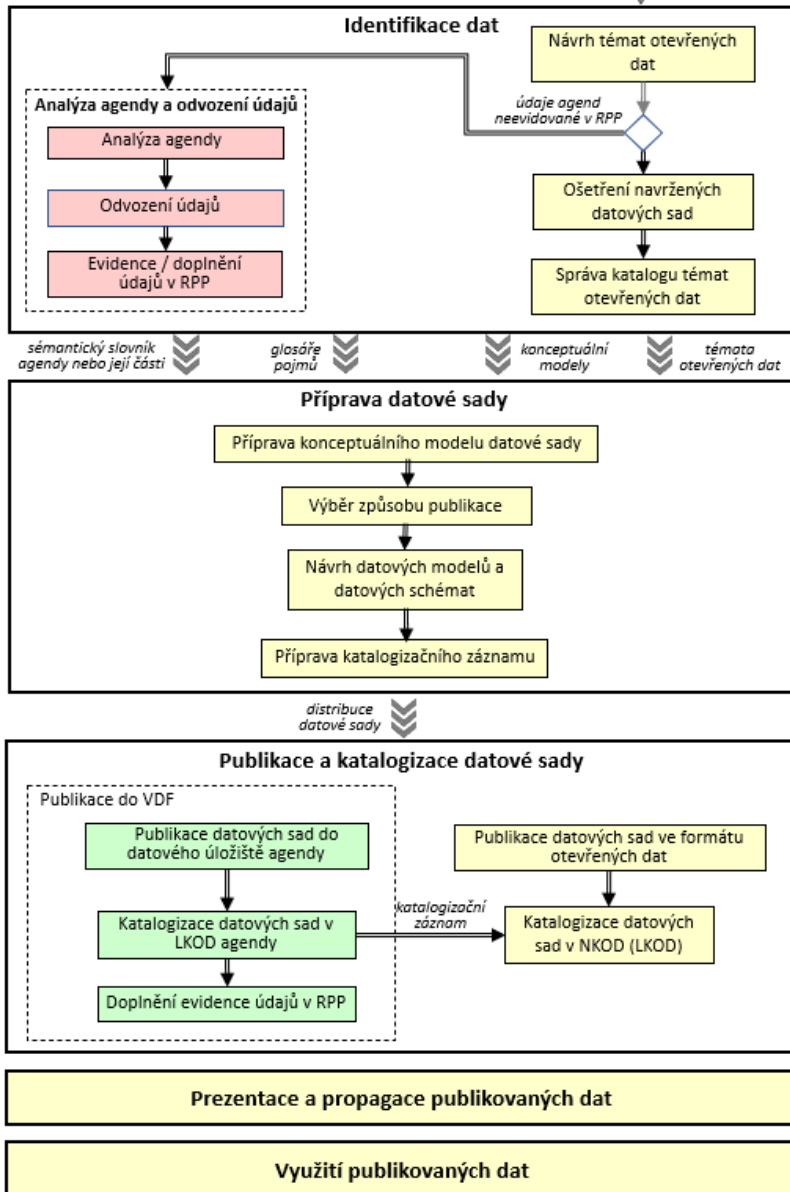
- **Sdílené prostředí publikace OD.**
  - Zahrnuje metodiky, nezbytnou sdílenou infrastrukturu, připravené a publikované standardy a vzdělávací materiály.
  - Sdílené prostředí je vytvářeno a spravováno centrálně agenturou DIA a je určeno pro všechny poskytovatele otevřených dat veřejné správy ČR.
  - Detailnímu popisu je věnována kapitola 2.1 „[Sdílené prostředí publikace OD](#)“.
- **Prostředí poskytovatele OD.**
  - Věnováno organizačnímu zabezpečení publikace dat v činnostech poskytovatele a případnému doplnění jeho lokální infrastruktury pro publikaci OD.
  - Detailnímu popisu je věnována kapitola 2.2 „[Prostředí poskytovatele OD](#)“.

**Publikace otevřených dat** je rozdělena do čtyř dílčích oblastí:

- **Identifikace dat.**
  - Zaměřuje se na
    - identifikování a shromažďování témat vhodných k publikaci formou otevřených dat,
    - rozlišení mezi agendovými a neagendovými údaji,
    - posouzení rizik a přínosů vyplývajících z jejich publikace
    - a správu vybraných témat k publikaci a informací o jejich publikování.
  - Detailnímu popisu je věnována kapitola 3 „[Identifikace témat otevřených dat](#)“.
- **Příprava datové sady.**
  - Ústředním tématem je zabezpečení sémantické a syntaktické interoperability datové sady.
  - Pozornost je věnována využití sémantiky a zejména dodržení postupu při návrhu struktur distribucí datové sady: „**téma -> konceptuální model -> datový model -> datový formát**“.
  - Uvedeny postupy pro přípravu datových sad se strukturovanými i nestrukturovanými údaji, využití otevřených formálních norem pro přípravu datových struktur distribucí a postupy přípravy katalogizačních záznamů včetně zajištění sémantiky na úrovni metadat.
  - Detailnímu popisu je věnována kapitola 4 „[Příprava datové sady](#)“.



Záměr publikace  
otevřených dat



#### Sdílené prostředí publikace OD

- Připravené metodiky (barva reprezentuje metodické pokrytí publikace otevřených dat):

Metodika identifikace, publikace, katalogizace a využívání otevřených dat pro samosprávu a pro státní správu.

Metodika tvorby a údržby konceptuálních datových modelů agend.  
Metodika definice údajů vedených v agendě.

Metodika poskytování dat ve veřejném datovém fondu.

- Národní katalog otevřených dat (NKOD).
- Veřejný datový fond (VDF) a jeho infrastruktura.
- Registr práv a povinností (RPP).
- Standardy (OFN).
- Nástroje pro monitorování kvality publikace datových sad.
- Vzdělávací materiály a e-learningová školení.
- Prototypy modelovacích nástrojů.

#### Prostředí poskytovatele OD

- Organizační zabezpečení publikace poskytovatelem.
- Přiřazení rolí k činnostem.
- Lokální katalog otevřených dat (LKOD).
- Údržba publikovaných otevřených dat

#### • Publikace datové sady.

- Oblast je zaměřena zejména na fyzickou přípravu distribucí datových sad a jejich zveřejnění. Informace jsou uvedeny formou doporučení pro přípravu fyzických dat k jejich transformaci do konkrétních formátů distribucí.
- Uvedeny jsou také postupy a způsoby rozhodování při návrhu podmínek užití pro jednotlivé distribuce.
- Detailnímu popisu je věnována dílčí část kapitoly 5 „[Příprava distribucí datové sady k publikaci](#)“.

#### • Katalogizace datové sady.

- Zaměřeno na zveřejnění datových sad, finalizaci katalogizačních záznamů a postupy katalogizace datových sad, zejména v NKOD.





- Detailnímu popisu je věnována dílčí část kapitoly 5 „[Publikace otevřených dat](#)“.

#### **Komunikace, propagace a využití otevřených dat.**

- Tato oblast nemá přímý metodický charakter a jsou zde shromážděny pouze doporučení, rady a možné inspirace k uvedeným tématům.

Obsah metodiky je zaměřen na metodické pokrytí všech oblastí a souvisejících postupů publikace otevřených dat ve veřejné správě ČR v jejich [cílové \(správné, úplné a kvalitní\) podobě](#).

## **1.5 Omezení metodiky**

Obsah metodiky je zaměřen na metodické pokrytí všech oblastí a souvisejících postupů publikace otevřených dat ve veřejné správě ČR v jejich **cílové (správné, úplné a kvalitní) podobě**.

Publikace dat ve veřejné správě ČR má za sebou ale již několikaletou historii a hodně dat bylo již publikováno, často ale bez dodržování požadovaných standardů na jejich publikaci. Rovněž i samotný obsah metodiky byl vytvářen a precizován postupnými kroky v souladu s tím, jak byl rozšiřován význam a rozsah pojmu „otevřená data“, a to:

- v informační architektuře veřejné správy ČR,
- v legislativě ČR a
- v kontextu role otevřených dat v EU.

Přirozeným výsledkem činností prováděných „za pochodu“ je zcela zákonitě určitý nesoulad mezi aktuálním stavem publikace datových sad a „nově“ požadovaným stavem.

Snaha a požadavek na postupné zavádění nových postupů v souladu s postupy uváděnými v metodice může místy vést k omezení její použitelnosti, například z důvodu prozatímní absence některých nástrojů, nebo částí potřebné infrastruktury prostředí otevřených dat.

Základní omezení je ale zatím spojeno s požadavkem na zajištění sémantické interoperability publikovaných dat. Cílem digitalizace veřejné správy je vytvoření jednotného sdíleného datového fondu, jehož předpokladem je také podchycení jednoznačného významu všech sdílených údajů, a tedy i publikovaných datových sad, které jsou jeho součástí. K identifikaci a popisu významu údajů je určeno:

- na straně správy a evidence agendových údajů modelování sémantiky agend a postupné vytváření „Sémantického slovníku pojmů veřejné správy“ (SSP),
- v oblasti publikace otevřených dat je obdobně vyžadováno modelování témat datových sad (prostřednictvím tvorby jejich konceptuálních modelů na základě pojmů spravovaných v SSP) a případné doplňování nových identifikovaných pojmů do SSP, pokud vzniknou jako výstup zmíněného modelování.

Stávající omezení modelování sémantiky (a tedy také zajištění sémantické interoperability):

- nedostatečný počet namodelovaných agend a s tím související omezený počet pojmů v SSP,
- podpora modelovacími nástroji usnadňujícími analýzu je zatím na úrovni vytvořených prototypů, které jsou sice zcela funkční, ale zatím nejsou plně implementovány do prostředí veřejné správy (schází



- dořešení administrace nástrojů, jejich správa a také správa a administrace vytvářeného slovníku SSP),  
• zatím omezený počet vyškolených pracovníků veřejné správy pro práci v oblasti sémantického modelování a pro práci s připravenými nástroji.

Význam a role zajištění sémantické interoperability dat bude postupně narůstat s postupnou implementací podpůrných modelovacích nástrojů.

Metodika bude postupně doplňována dalšími dokumenty a postupy, které budou upravovat používání podpůrných nástrojů, upravovat procesy při vytváření sémantických modelů, a také procesy spojené s jejich následným využíváním, identifikací nových pojmů, rozšiřováním slovníku pojmů a využívání sémantiky při zajištění sémantické interoperability publikovaných datových sad.

Prakticky to znamená, že až do ukončení implementace nástrojů budou (bohužel) akceptovatelné i publikace datových sad bez požadovaného zajištění sémantické interoperability dat, avšak požadavek na zajištění syntaktické interoperability se nemění a platí v plném rozsahu.



# Část 1. Metodika

## 2 Prostředí publikace otevřených dat veřejné správy ČR

Základní podmínkou **publikace kvalitních a interoperabilních otevřených dat** je existence vytvořeného prostředí pro jejich publikaci, a to jak u samotného poskytovatele, tak i v celé veřejné správě ČR.

### 2.1 Sdílené prostředí publikace OD

Na straně veřejné správy ČR a zejména garanta otevřených dat VS se zejména jedná o:

- Ukotvení otevřených dat a souvisejících povinností v legislativě ČR.
  - [Zákon č. 106/1999 Sb.](#) Zákon o svobodném přístupu k informacím
  - [Zákon č. 123/1998 Sb.](#) Zákon o právu na informace o životním prostředí
- Metodické pokrytí všech oblastí otevřených dat:
  - Metodika identifikace, publikace, katalogizace a využívání otevřených dat pro samosprávu a pro státní správu (tato metodika).
  - [Metodika tvorby a údržby sémantického slovníku pojmů veřejné správy.](#)
  - [Metodika definice údajů vedených v agendě.](#)
  - [Metodika poskytování dat ve veřejném datovém fondu.](#)
  - [Metodika pro vypracování analýzy dopadu otevřených dat do prostředí ČR.](#)
- Zajištění infrastruktury nutné pro publikaci a využívání otevřených dat:
  - Národní katalog otevřených dat ([NKOD](#)),
  - Veřejný datový fond (VDF) a jeho komponenty,
  - Registr práv a povinností ([RPP](#)),
  - Repositář Otevřených formálních norem ([OFN](#)),
  - Repositář sémantických modelů.
- Vydané a publikované standardy pro publikaci otevřených dat:
- [Nástroje pro monitorování kvality](#) publikace datových sad a zpracovávání zpětné vazby.
- Vzdělávací materiály a [e-learningové kurzy](#) na Portálu otevřených dat ([POD](#)).
- Připravené [prototypy nástrojů](#) pro podporu sémantického modelování dat a vytváření datových schémat.
  - Termit – tvorba slovníků pojmů, nástroj pro správu odborného pojmosloví a jeho propojení na legislativu,
  - Ontographer – tvorba konceptuálních modelů, nástroj pro správu konceptuálních modelů využívajících vytvořené pojmosloví,
  - Dataspecer – nástroj pro automatizovanou tvorbu datových specifikací a OFN přímo z vytvořených konceptuálních modelů,
  - ShowIT – webový prohlížeč a vyhledávač pojmů v sémantickém slovníku pojmů (SSP).

#### 2.1.1 Národní katalog otevřených dat (NKOD)

Definice ze Zákona o svobodném přístupu k informacím - [§ 4c, Zákon č. 106/1999 Sb.](#):

*(2) Národní katalog otevřených dat je informační systém veřejné správy přístupný způsobem umožňujícím dálkový přístup sloužící k evidování informací zveřejňovaných jako otevřená data a dalších informací zveřejňovaných způsobem umožňujícím dálkový přístup, zejména rozhraním pro programování aplikací a informací o produktech*



a službách, které otevřená data a další informace využívají. Informace obsažené v národním katalogu otevřených dat jsou poskytovány způsobem umožňujícím dálkový přístup jako otevřená data.

(3) Správcem národního katalogu otevřených dat je Digitální a informační agentura.“

### 2.1.2 Veřejný datový fond

Veřejný datový fond (VDF) je metoda sdílení veřejných agendových údajů v prostředí VS, která je tvořena a zajišťována:

- provázanými již existujícími, ale i zatím neexistujícími technologickými komponentami architektury VS ČR, které se dělí do dvou skupin:
  - společné komponenty VDF – jsou zaměřeny především na správu a organizaci zpřístupňovaných údajů a souvisejících procesů, jsou plně zajišťovány garantem otevřených dat,
  - komponenty jednotlivých agend – jsou zaměřeny především na zpřístupnění a užívání údajů jednotlivých agend, jsou zajišťovány jednotlivými OVM, správci agend, správci ISVS,
- všemi veřejnými agendovými údaji VS publikovanými ve formátu otevřených dat,
- pravidly a postupy vytvářejícími a zabezpečujícími metodu sdílení údajů v duchu její definice v [Informační koncepci ČR](#) v [dílčím cíli 5.10](#),
- standardy zajišťujícími plnou interoperabilitu sdílených údajů napříč VS.

Podrobnější informace o VDF jsou uvedeny v kapitole „[Stručný popis VDF](#)“.

### 2.1.3 Registr práv a povinností (RPP)

Registr práv a povinností (AIS působnostní) obsahuje a spravuje informace pro řízení přístupu k údajům základních registrů a k údajům v agendových informačních systémech.

Pro potřeby VDF je RPP rozšířen o evidenci doplňujících informací k agendám a jejich údajům, a také obsahuje nové následující komponenty:

- Úložiště číselníků a evidence RPP,
  - obsahuje a zpřístupňuje datové sady publikovaných veřejných číselníků,
  - obsahuje a zpřístupňuje obsah (údaje) RPP jako datové sady.
- Lokální katalog RPP,
  - obsahuje katalogizační záznamy pro:
    - datové sady zastřešující verze číselníků,
    - datové sady s verzemi číselníků,
    - datové sady s údaji, které jsou evidovány a spravovány v RPP (obsah RPP).
  - zpřístupňuje katalogizační záznamy k jejich katalogizaci v NKOD.
- Systém pro zpřístupnění údajů RPP,
  - připravuje obsah RPP a veřejné číselníky do podoby datových sad a dávkově je předává do úložiště datových sad k publikaci.

### 2.1.4 Nástroje pro monitorování kvality publikace datových sad

Za kvalitu dat i katalogizačních záznamů zodpovídají jednotliví poskytovatelé. Ti musí na kvalitu svých dat dbát, pravidelně ji sledovat a dělat kroky k nápravě zjištěných nedostatků. K usnadnění těchto činností Portál otevřených dat obsahuje následující připravené nástroje:



**Statistiky NKOD a dashboardy** – denně je zjišťována dostupnost zdrojů registrovaných v Národním katalogu otevřených dat a také zpracovávány statistiky kvality jejich metadatových záznamů.

Základní **ukazatele** vypočítané z metadat NKOD:

- Kvalita metadatových záznamů v NKOD.
- Statistika dostupnosti distribucí, schémat, podmínek užití a dokumentace.
- Druhy registrace datových sad v NKOD.
- Denní změny v NKOD.
- Měsíční změny v NKOD.

Statistiky jsou pro každého poskytovatele vizualizované pomocí dashboardů, a to jak pro okamžitý aktuální stav, tak i v delším časovém pohledu vypovídajícím o jejich trendu.

**Příklady špatné praxe** – praktický doplněk monitorování kvality publikace datových sad, který shromažďuje často se **opakující problémové situace**. Pro každou problémovou oblast je uveden seznam nejčastějších nedostatků, včetně navrhaných řešení.

## 2.1.5 Otevřené formální normy (OFN)

Definice ze Zákona o svobodném přístupu k informacím - [§ 3a odst. 3, Zákon č. 106/1999 Sb.](#):

*„(3) Otevřenou formální normou se pro účely tohoto zákona rozumí pravidlo, které bylo vydáno písemně a obsahuje specifikace požadavků na zajištění schopnosti různých programových vybavení vzájemně si poskytovat služby a efektivně spolupracovat.“*

Zatím bylo vydáno 34 otevřených formálních norem včetně sdílených specifikací, což jsou otevřené formální normy specifikující části, které jsou společné pro více jiných otevřených formálních norem, datových specifikací a schémat datových sad.

OFN je věnována kapitola „**OFN**“, a doplňující informace (zejména k jejich použití) jsou ještě uvedeny v dalších relevantních kapitolách metodiky,

## 2.1.6 Vzdělávací materiály a e-learningová školení

**Vzdělávací materiály:**

- Analýza existujících informačních systémů veřejné správy vč. jejich příslušné legislativy a identifikace datových sad pro veřejný datový fond (odkaz na [dokument](#) a [přílohu](#)).
- Komunikační strategie pro cílové skupiny otevřených dat (odkaz na [dokument](#) a [přílohu](#)).
- Propagační články a vizualizace ke konkrétním otevřeným datům v NKOD (sekce „[Články ze světa otevřených dat](#)“ na „[Portálu otevřených dat](#)“).
- [Výstupy projektu KODI](#) - Rozvoj datových politik v oblasti zlepšování kvality a interoperability dat veřejné správy – OPZ CZ.03.4.74/0.0/0.0/15\_025/0013983.
- Publikované informace a návody na „[Portálu otevřených dat](#)“.

**E-learningová školení:**

- [Co jsou otevřená data](#) – základní informace o tom, co jsou otevřená data, proč se mají publikovat a jaký je jejich kontext v ČR.
- [Publikační plán a publikace dat](#) – poskytnutí informací k vytvoření publikačního plánu organizace.
- [Katalogizace otevřených dat](#) - základní informace k rozhodnutí o způsobu katalogizace otevřených dat a vytvoření katalogizačních záznamů.
- [Úvod do otevřených formálních norem \(OFN\)](#) – poskytnutí informací k posouzení možností/nutností využití otevřených formálních norem a k výhodám jejich využití.



- [Technické aspekty otevřených dat](#) – poskytnutí informací k získání přehledu o technologiích a technických postupech používaných v otevřených datech.
- [Veřejný datový fond \(VDF\) v architektuře veřejné správy](#) – seznámení s novou metodou sdílení údajů ve veřejné správě prostřednictvím veřejného datového fondu (VDF), který je definován v Informační koncepci ČR (IKČR) jako dílčí cíl 5.10 a je součástí eGovernmentu VS ČR.

#### Záznamy školení v kanálu Youtube „Digitální a informační agentury“:

- [Modelování významu dat ve veřejné správě](#) – seznámení s modelováním významu dat, použití sémantických slovníků pro vyhledávání a dokumentaci dat a způsob tvorby sémantických slovníků s využitím konkrétních softwarových nástrojů.
- [Zpracování a vizualizace otevřených dat](#) – seznámení se způsoby čištění dat, jejich zpracováním, vyhledáváním a způsoby transformace do konkrétních výstupů.
- [Propojená data jako další stádium vývoje OD](#) – seznámení se základy propojených dat, slovníky pro popis významu dat, modelem RDF a dotazovacím jazykem SPARQL.
- [Právní úprava otevřených dat](#) – seznámení s právním zakotvením otevřených dat v kontextu českého právního řádu a práva EU.
- [Práva k informacím a datům: Právní překážky pro otevřená data](#) – seznámení se specifickými právními překážkami, které je třeba vyřešit v průběhu procesu poskytování otevřených dat a během jejich opětovného užití.
- Záznam workshopu "[Publikace úředních desek jako otevřená data](#)" - představení povinného způsobu zveřejnění pomocí otevřené formální normy pro úřední desky (praktická ukázky použití OFN) a jeho doplnění ukázkami „[Ověření správnosti publikace dat Úředních desek](#)“.

#### Záznamy Výročních konferencí „Otevřená data ve veřejné správě“:

- Konference Otevřená data ve veřejné správě 2017–[1. část](#), [2. část](#).
- Konference [Otevřená data 2020](#).
- Konference [Otevřená data 2021](#).
- Konference [Otevřená data 2022](#).

## 2.2 Prostředí poskytovatele OD

Prostředí poskytovatele a jeho příprava je plně v kompetencích konkrétního poskytovatele, a bude zcela určitě vždy závislé na velikosti organizace a množství spravovaných a potenciálně publikovaných dat. Lze připustit i publikaci způsobem **ad hoc** (bez implementace činností spojených s přípravou a publikací otevřených dat do provozních činností organizace), avšak i v takovém případě je nutné naplňovat [požadovanou kvalitu](#) publikovaných dat a kvalitu publikace.

**Bez dodržení požadované kvality publikovaných dat je jakákoliv publikace otevřených dat zbytečná a pouhé plýtvání zdroji.**

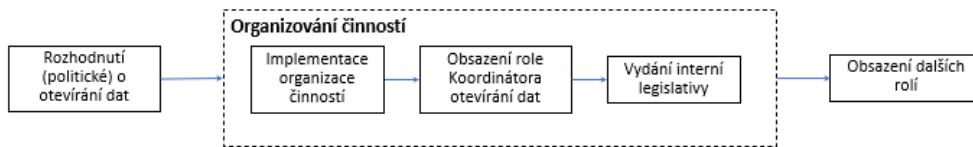
Dále popisované prostředí poskytovatele představuje **systemový přístup** k publikaci, a je vhodný pro jakýkoliv typ organizace jakékoliv velikosti. Přizpůsobení je pouze otázkou konkrétní implementace v konkrétní organizaci.

**Na straně poskytovatele se jedná zejména o:**



- „Politické“ rozhodnutí o otevírání dat a stanovení základních cílů v této oblasti vedením organizace.
- Implementaci činností nutných k publikaci otevřených dat do chodu organizace a jejich zakotvení vhodnou interní legislativou.
- Personální zajištění implementovaných činností prostřednictvím přiřazení požadovaných rolí konkrétním pracovníkům organizace.
- Zajištění (vyhrazení) nezbytné technické infrastruktury pro potřeby publikace:
  - případný Lokální katalog otevřených dat (LKOD),
  - úložiště publikovaných datových sad.
- Zajištění údržby publikovaných otevřených dat.
- Zahrnutí otevřených dat do informační strategie organizace.

## 2.2.1 Organizační zabezpečení publikace poskytovatelem



Z – zodpovídá  
S – spolupracuje  
R – rozhoduje

Vedení organizace

Koordinátor otevírání dat

<i>Rozhodnutí (politické) o otevírání dat</i>	Z	
<i>Implementace organizace činností</i>	Z	
<i>Obsazení role Koordinátora otevírání dat</i>	Z	
<i>Vydání interní legislativy</i>	R	Z
<i>Obsazení dalších rolí</i>	R	Z

### 2.2.1.1 Rozhodnutí (politické) o otevírání dat

Absence „politického“ prohlášení a zapojení vedení organizace do publikace otevřených dat vede k:

- nedostatečné motivaci pracovníků organizace se otevřenými daty seriózně zabývat,
- nízké kvalitě výběru publikovaných témat,
- nedodržování kvalitativních požadavků na publikované datové sady,
- časté publikaci způsobem ad hoc pouze nadšenci z řad pracovníků organizace.

Rozhodnutí („politické prohlášení“) vedení organizace má výrazný vliv na přístup organizace k otevřeným datům.

Není výjimkou, že v NKOD jsou katalogizované datové sady otevřených dat, které jsou až nepoužitelné, bez nějaké hodnoty, a že se v podstatě jedná o plýtvání ze strany organizace.

Přístup vedení organizace k otevřeným datům lze v zásadě klasifikovat následujícím způsobem:

- Žádný, bohužel poměrně dosti běžný. Pro tento přístup platí výše uvedený text.
- **Nesystémový**, publikace probíhá bez vyjasněné strategie, sleduje spíše krátkodobé malé cíle, případně reaguje na nové vznikající povinnosti, vyplývající například ze zákona. Nejedná se o cílový přístup organizací veřejné správy, ale pro zapojení se (zejména počáteční) do publikace otevřených dat se jedná o legitimní způsob.



- **Systémový**, který vychází z plného pochopení významu a role otevřených dat pro veřejnost, privátní sektor, veřejnou správu a celou EU. Jedná se o cílový přístup organizací veřejné správy. Základní informace o systémovém pohledu na otevřená data jsou uvedeny v kapitole „[Rekapitulace informací o otevřených datech](#)“.

### 2.2.1.2 Organizování činností

Organizace veřejné správy (zejména státní správy) se vyznačují liniově štábní organizační strukturou, ve které jsou pracovní pozice a vztahy nadřízenosti a podřízenosti uspořádány a orientovány vertikálně. Řízení je převážně určováno:

- popisy pracovních míst s vymezenými hlavními činnostmi, které jsou v rámci pracovního místa vykonávány,
- specifikovanými vztahy podřízenosti a nadřízenosti,
- směrnicemi a dalšími řídicími dokumenty, zaměřenými na povinnosti a činnosti nepodchycené v popisech pracovních míst.

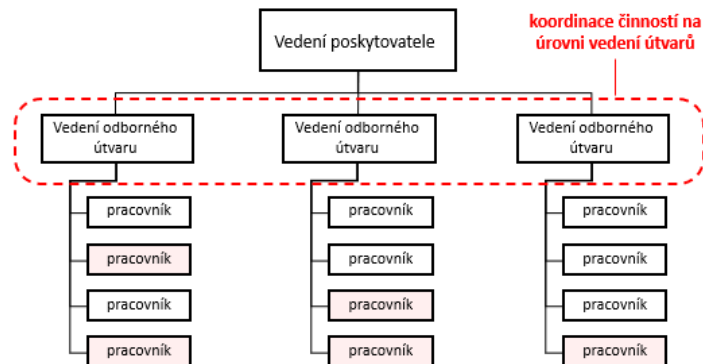
Oblast otevřených dat je charakteristická kombinací procesního a projektového řízení.

- Vytvoření prostředí pro publikaci otevřených dat, příprava každé nové datové sady k publikaci a její prvotní publikace jsou často jedinečným sledem činností, a tedy malým projektem.
- Opakovaná (rutinní, nebo přímo automatická) publikace již jednou publikované datové sady vykazuje charakteristiky procesu, a je tedy již záležitostí procesní.

Metodika je zaměřena výhradně na část „projektovou“, tedy nezabývá se již rutinním (procesním) zabezpečením (často již automatické) publikace aktuálních otevřených dat.

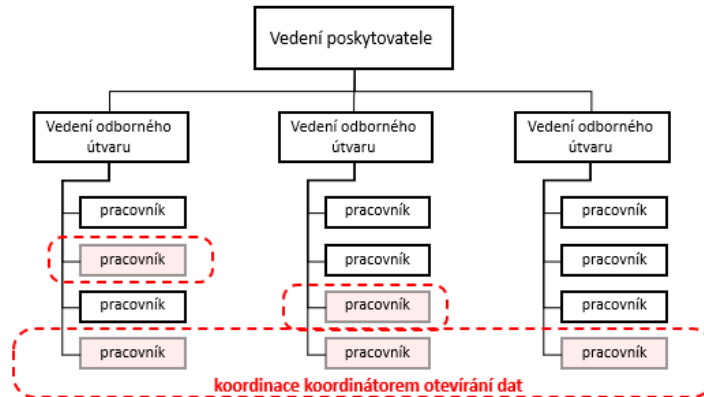
Zavedení prvků projektového řízení do liniových organizačních struktur lze v podstatě realizovat dvěma způsoby:

- Koordinací činností projektu na úrovni vedení útvarů. Nevýhoda takového řízení spočívá v samostatném (izolovaném) provádění činností jednotlivými útvary bez koordinace se zbývajícími členy projektového týmu (a příslušnými dalšími činnostmi) v ostatních útvarech.



- Koordinací aktivit na úrovni přímo zapojených pracovníků do realizace projektu prostřednictvím „maticové struktury“.





- Jedná se o doporučenou formu pro řešení publikace otevřených dat.
- Koordinaci všech zapojených pracovníků zajišťuje „koordinátor otevřených dat“ prostřednictvím definovaných potřebných rolí.
- Způsob implementace a přiřazení rolí konkrétním pracovníkům závisí na konkrétních poskytovatelích.
- Všechny role včetně jejich pravomocí a odpovědností by měly být popsány v interní legislativě (např. [Vzorová směrnice publikace a katalogizace otevřených dat](#)).
- Jednu roli může zastávat více zaměstnanců a jeden zaměstnanec může zastávat více rolí:
  - např. Kurátorů dat bývá typicky více, neboť se jedná zpravidla o věcné, nebo také doménové, experty, přičemž málokdy zná jeden zaměstnanec všechny datové sady organizace.
- Způsoby přiřazení rolí:
  - pomocí směrnice,
  - v rámci pracovní smlouvy či dodatku k ní,
  - vytvořením pracovní skupiny na základě jmenovacího dekretu nejvyšším vedením organizace (zajištění jistého stupně odpovědnosti a součinnosti daného zaměstnance),
  - ad hoc – ústní dohodou, v zápise z porady, ...
- Institucionalizace maticového řízení do činností organizace:
  - nejvhodnější formou je vydaná směrnice,
  - pro její vytvoření je možné použít připravený a na POD publikovaný návrh [Vzorové směrnice publikace a katalogizace otevřených dat](#) se strukturou:
    - vymezení působnosti směrnice,
    - pojmy,
    - přehled rolí,
    - práva a povinnosti rolí,
    - postup publikace a katalogizace otevřených dat,
    - katalogy otevřených dat,
    - závěrečná ustanovení.

### 2.2.1.3 Role zabezpečující činnosti otevření dat v organizaci

- Vedení organizace
  - Nese „politickou“ zodpovědnost za celou agendu otevření dat organizace.
  - Zodpovídá za vytvoření potřebné organizační infrastruktury otevření dat v organizaci.
  - Obsazuje roli Koordinátora otevření dat, prostřednictvím něhož komunikuje a činí další rozhodnutí v oblasti otevřených dat.



- Schvaluje potřebnou interní legislativu.
- Schvaluje obsazení dalších rolí na základě návrhu koordinátora otevírání dat.
- Schvaluje témata publikace / publikační plán.
- **Koordinátor otevírání dat**
  - Je přímo podřízen vedení organizace.
  - Zodpovídá za koordinaci všech činností spojených s otevíráním dat v organizaci.
  - Poskytuje metodickou podporu všem rolím.
  - Reportuje o publikaci otevřených dat Vedení poskytovatele.
- **Kurátor dat**
  - Odborník na příslušnou věcnou doménu.
  - Vhodní kandidáti se rekrutují např.:
    - ze správců/gestorů dat,
    - doménových expertů,
    - datových analytiků.
  - Zajišťuje činnosti analytického charakteru spojené s přípravou datových sad.
- **Správce katalogu otevřených dat**
  - Přípravuje a technicky spravuje katalogizační záznamy, případně provozuje a udržuje lokální katalog otevřených dat organizace.
  - Zajišťuje činnosti spojené s katalogizací otevřených dat.
  - Jedná se převážně o práci s metadaty datových sad a prací s katalogizačním systémem.
- **IT specialista**
  - Technicky zaměřená role, poměrně často zastoupena dodavatelem informačního systému organizace.
  - Pracuje převážně se samotnými daty, připravuje je po technické stránce a zajišťuje chod příslušné IT infrastruktury.
  - Zajišťuje činnosti související s technickými aspekty otevírání dat.
- **Ohlašovatel agendy**
  - [Ohlašovatel agendy](#) je podle zákona č. 111/2009 Sb., o základních registrech zodpovědný za řádné ohlášení agendy její aktualizace, a především za správnost a pravdivost informací o údajích vedených v agendě.
  - Ověřuje evidenci agendových údajů v RPP a její správné provedení dle Metodiky definice údajů vedených v agendě.
  - Určuje přesné vymezení agendových údajů (kdy se jedná o agendový údaj a kdy již ne) na základě věcné znalosti agendy a souvisejících agendových informačních systémů.

### **Role zapojené do sémantického modelování**

- **Specialista na modelovanou doménu (které se slovník publikovaného tématu týká).**
  - Má přehled o doméně a shromažďuje relevantní dokumentaci k publikované oblasti (doméně).
  - Přípravuje a spravuje tezaurus modelované (publikované) domény.
  - Spolupracuje na tvorbě konceptuálního modelu.
- **Datový architekt**
  - Spolupracuje na tvorbě konceptuálního modelu.
  - Přípravuje sémantický slovník modelované (publikované) domény.
  - Přípravuje datová schémata a datové specifikace.



## 2.2.2 Lokální katalog otevřených dat (LKOD)

Provozování lokálního katalogu otevřených dat (LKOD) je jedním ze způsobů zajištění katalogizace dat v Národním katalogu otevřených dat (NKOD), což je jedna ze zákonných podmínek pro to, aby data mohla být nazývána otevřenými daty. Druhým způsobem je přímá katalogizace v NKOD bez provozování vlastního lokálního katalogu.

Způsob katalogizace s využitím LKOD je vhodný pro organizace, u kterých se dá předpokládat katalogizace většího množství datových sad či častější aktualizace jejich metadat.

LKOD je tedy volitelně implementován pro potřeby katalogizace otevřených dat konkrétního poskytovatele dat. Jeho rozhraní musí odpovídat vydané Otevřené formální normě "[Rozhraní katalogů otevřených dat](#)" a vytvářet podmínky nutné pro harvestování obsahu LKOD Národním katalogem otevřených dat. Tato otevřená formální norma v čase prochází vývojem, je proto nutné ji sledovat a udržovat rozhraní LKOD v souladu s OFN. Automatická publikace záznamů z lokálního katalogu do NKOD je zajištěna jeho registrací v NKOD poskytovatelem.

### 2.2.2.1 Registrace lokálního katalogu

Registrace LKOD je provedena zasláním registračního dokumentu, který lze získat vyplněním formuláře po kliknutí na odkaz „Registrovat nový lokální katalog“ v patičce NKOD ([POD](#)).

Postup registrace lokálního katalogu v NKOD je následující:

- Otevřít [registrační formulář](#).
- Registrační údaje a jejich popis ve formuláři.
  - Název lokálního katalogu otevřených dat (povinný atribut).
    - Tento název rozlišuje mezi jednotlivými katalogy otevřených dat dané instituce. Pokud má jeden, může to být například „Katalog otevřených dat XXX“
  - Jméno a email správce katalogu (povinné atributy).
    - Jméno a email správce lokálního katalogu otevřených dat, se kterým lze komunikovat o případných problémech s katalogem (například s formátem API, přístupností rozhraní apod).
  - Typ API lokálního katalogu (povinný atribut).
    - Stanovuje typ API (rozhraní) registrovaného katalogu. Více viz Otevřené formální norma [Rozhraní katalogů otevřených dat](#).
  - URL API lokálního katalogu (povinný atribut).
    - Stanovuje URL API katalogu dle Otevřené formální normy [Rozhraní katalogů otevřených dat](#).
  - Domácí stránka katalogu (volitelný atribut)
    - URL domácí stránky katalogu, pokud existuje. Lokální katalog může obsahovat větší množství informací a souvislostí, než se dostane do Národního katalogu otevřených dat. Proto je vhodné, aby se uživatelé po nalezení datové sady v NKOD mohli podívat i do lokálního katalogu.
- Posledním krokem je odeslání výsledného dokumentu datovou schránkou (například přes spisovou službu Poskytovatele) v podobě přílohy datové zprávy s předmětem „NKOD“ a její adresování na datovou schránku Národního katalogu otevřených dat (NKOD) s ID m3hp53v.



### 2.2.2.2 Zrušení registrace lokálního katalogu

Registraci LKOD je možné zrušit zasláním dokumentu, který lze získat pomocí formuláře pro smazání lokálního katalogu.

Postup zrušení je následující:

- V [Seznamu lokálních katalogů v NKOD](#) je nutné nalézt rušený lokální katalog a v jeho detailu kliknout na symbol smazání.
- Výsledný dokument je nutné poslat datovou schránkou (například přes spisovou službu Poskytovatele) jako přílohu datové zprávy s předmětem „NKOD“ a adresovat ji na datovou schránku Národního katalogu otevřených dat (NKOD) s ID m3hp53v.

### 2.2.2.3 Varianty LKOD vyskytující se ve VDF

V případě publikace otevřených dat do veřejného datového fondu jsou k dispozici dva typy lokálních katalogů.

Lokální katalog otevřených dat RPP – je součástí „AIS Působnostní“ a obsahuje katalogizační záznamy pro:

- datové sady zastřešující verze číselníků,
- datové sady s verzemi číselníků,
- datové sady veřejných údajů vedených v RPP.

Lokální katalog otevřených dat (jednotlivých agend):

- Obsahuje a spravuje katalogizační záznamy datových sad publikovaných agendových veřejných údajů.
- K dispozici jsou dvě možné varianty jeho realizace:
  - LKOD pro každý konkrétní AIS a jeho publikované údaje, LKOD je přímo komponentou AIS.
  - LKOD konkrétní organizace VS, který katalogizuje všechny údaje publikované organizací (včetně publikovaných agendových údajů).

Pro všechny lokální katalogy platí povinnost dodržovat Otevřenou formální normu [Rozhraní katalogů otevřených dat](#), postupy jejich registrace i smazání v NKOD jsou shodné.

## 2.2.3 Zajištění údržby publikovaných otevřených dat

Kvalita publikovaných datových sad je určována samozřejmě již jejich publikací, ale také jejich navazující správou a údržbou v čase. Je nutné zajistit jejich trvalost, aby data a metadatové informace byly udržovány a uchovávány na dlouhodobém základě, a aby byly dostupné a použitelné i v budoucnosti.

Jedná se zejména o jejich:

- dostupnost.
- integritu,
- aktualizaci a
- kompatibilitu.

Více k tématu v kapitole „[Kvalita otevřených dat a kvalita jejich publikace](#)“.

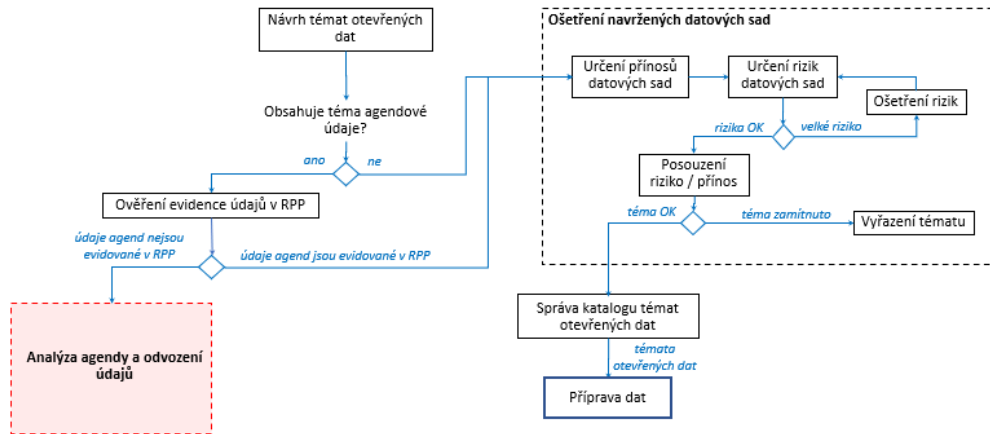
K zajištění kvality publikovaných otevřených dat je vhodné a doporučené:

- Využívat „[Nástroje pro monitorování kvality publikace datových sad](#)“, které sledují základní nedostatky, jako jsou:
  - nedostatečná kvalita metadatových záznamů v NKOD,
  - nedostupnost distribucí, schémat, podmínek užití a dokumentace,
  - neshody mezi formátem distribuce v NKOD a formátem indikovaným serverem,
  - nedostupnost přístupových bodů distribucí ve formě datové služby,



- o nedostupnost techniky CORS u přístupových bodů distribucí ve formě datové služby,
- o nedostupnost specifikací datových sad,
- o nedostupnost specifikací datových služeb.
- Zajistit způsob přijímání zpětné vazby od uživatelů publikovaných dat a její zpracování.
- Sledování změn legislativy a patřičně na reagovat.

### 3 Identifikace témat otevřených dat



	Vedení organizace	Koordinátor otevírání dat	Kurátor otevřených dat	Správce katalogu OD	IT specialista	Ohlašovatel agendy
Z – zodpovídá						
S – spolupracuje						
R – rozhoduje						
<i>Návrh témat otevřených dat</i>	S	Z	S	S	S	S
<i>Ověření evidence údajů v RPP</i>					S	Z
<i>Ošetření navržených témat otevřených dat</i>		S	Z		S	
<i>Správa katalogu témat otevřených dat</i>		Z	S	S	S	
<i>Analýza agendy a odvození údajů</i>		S	S		S	Z

#### 3.1 Návrh témat otevřených dat

Prvním krokem předurčujícím kvalitu a hodnotu publikovaných otevřených dat (pokud se nejedná o agendové veřejné údaje), je správný výběr témat publikovaných údajů a oblastí

##### Seznam možných způsobů a inspirací identifikace témat potenciálních otevřených dat:

- Datový audit organizace, z něhož je zřejmé:
  - o jaká data organizace spravuje,
  - o jaký je jejich význam,
  - o kde jsou uložena,
  - o jaké jsou jejich vzájemné vztahy (kontext),
  - o jakým způsobem jsou sdílena mezi informačními systémy...
- Legislativní požadavky na povinnou publikaci otevřených dat, např. [v § 5a zákona č.106/199 Sb.](#)
- Publikace veřejných údajů agend (údajů registrovaných v RPP) do VDF.
- Požadavky na publikaci HVDS (High Value Datasets) uvedené [v § 5b zákona č.106/199 Sb.](#) a ve „[Směrnici Evropského parlamentu a Rady \(EU\) 2019/1024 ze dne 20. června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru](#)“, v příloze č.1. V případě HVDS je nutné věnovat



- pozornost jejich specifikám, které jsou uvedeny v kapitole „[Publikace High Value Datasets](#)“.
- Vytěžení plánovaných cílů publikace – návrhy vychází ze strategického rozhodnutí organizace o cílech publikace otevřených dat.
  - Vytěžení z motivací k publikaci, týká se zejména začátků práce s otevřenými daty.
  - Vytěžení (inspirace) z potenciální hodnoty publikovaných otevřených dat. Otevřená data při svém zhodnocování mohou vystupovat ve dvou rolích:
    - nástroj pro utváření společenských hodnot (oblast Open Governmentu),
    - zdroj pro tvorbu ekonomických hodnot (oblast opakovaného použití dat a eGovernmentu).
  - Identifikace informací vyměňovaných s jinými OVM.
  - Identifikace oblastí nedostatečně komunikovaných s veřejností.
  - Inspirace [vzorovými publikačními plány](#), které jsou k dispozici na POD pro různé typy organizací.
  - Oslovení vhodné skupiny zaměstnanců Koordinátorem otevírání dat se žádostí, aby navrhované datové sady zaznamenávali do sdíleného dokumentu.
  - Interní workshop iniciovaný Koordinátorem za účasti Kurátorů dat, na kterém seznam vznikne.
  - Podněty ze strany veřejnosti či jiných institucí shromažďované dostupnými prostředky (Wish listy).
  - Analýza webu organizace.
  - Konceptualizace zveřejňovaných dokumentů.
  - Průzkumy a analýzy zahraničních katalogů se zaměřením na obdobné publikující organizace.
  - Inspirace vytvořenými aplikacemi nad otevřenými daty v ČR.
  - Inspirace příklady úspěšných aplikací (zhodnocení dat státní správy) ze zahraničí
  - Seznam stávajících OFN a jejich použití v publikovaných datových sadách
  - Průzkum datových sad využitých u úspěšných konceptů SMART City
  - Domény Smart City a jejich datové pokrytí.

Každé navržené téma by mělo být zpravidla definováno svým názvem, odpovědným kurátorem a stručným popisem.

V této části se ještě nemusí jednat o přesné vymezení rozsahu publikovaného kontextu tématu, a tedy ani rozsahu publikovaných údajů. Vždy je nutné a výhodné v dalších krocích kontext tématu zachytit v podobě konceptuálního modelu, a až následně dle zvoleného typu publikace a použitého datového modelu vymezit rozsah publikovaných údajů, a také konkrétní datové sady pro publikaci. Bližší vysvětlení v kapitole „[Výklad problematiky vymezení věcného obsahu datové sady](#)“.

## 3.2 Ověření evidence údajů v RPP

Na základě navržených témat otevřených dat je nutné nejprve posoudit, zda se v navržených tématech vyskytují údaje agend.

Pokud se vyskytují, je nutné ověřit:

- zda jsou tyto agendové údaje již evidované v RPP,
- zda jejich evidence je provedena správně a obsahuje všechny požadované údaje dle [Metodiky definice údajů vedených v agendě](#).

V případě, že kvalita evidence není dostatečná, nebo agendové údaje nejsou evidované, je nutné jejich evidenci



nejprve zajistit dle [Metodiky definice údajů vedených v agendě](#) a následně se k publikaci navržených témat vrátit.

Role s odpovídající věcnou znalostí agendy a agendových informačních systémů:

Přesné vymezení agendových údajů (kdy se jedná o agendový údaj a kdy již ne) není možné explicitně stanovit. Určení je vždy závislé na věcné znalosti agendy a souvisejících agendových informačních systémů, a rozsahu údajů subjektů a objektů práva zpracovávaných agendou.

- Ohlašovatel agendy [§ 54 zákona č. 111/2009 Sb.](#) o základních registrech,
  - je povinen uvést do Registru práv a povinností jaké údaje se v agendovém informačním systému o daném subjektu nebo objektu práva vedou (dle souvisejících kontextů).

V případě publikace otevřených dat do VDF se jedná také o:

- Správce AIS, [zákon č. 111/2009 Sb. o základních registrech](#), [§ 51a odst. 1 zákona č. 111/2009 Sb.](#) o základních registrech,
  - udržuje datový fond AIS v aktuálním tvaru s využitím procesů vyrozumívání o změnách údajů,
  - spolupracuje s věcnými správci agend.
- Věcný správce agendy.

Podrobnější informace o agendových údajích, jejich evidenci a publikaci do veřejného datového fondu jsou uvedeny v kapitolách [Údaje VS](#) a [Stručný popis VDF](#).

### 3.3 Ošetření navržených témat otevřených dat

Navržená témata otevřených dat je nutné posoudit a ošetřit z pohledu potenciálních rizik vyplývajících z jejich publikace, a také ohodnotit z pohledu jejich potenciálních přínosů vyplývajících z jejich následného užití.

Pokud poskytovatel tento krok nepodstoupí, nedokáže zhodnotit a docenit možnosti využití publikovaných dat (jejich hodnotu), a také dohlédnout všechny potenciální problémy spojené s jejich publikací.

Způsob posouzení, návrh a výběr hodnotících kritérií, výběr metody jejich ocenění apod., závisí plně na přístupu a rozhodnutí poskytovatele. Neexistuje předepsaný způsob hodnocení ani předepsaná kritéria a jejich počet, jediné, co je jasně a konkrétně definované jsou zákony, které přesně vymezují hranice veřejnosti údajů.

**Kvalitní a důsledná evaluace témat je zárukou kvality obsahu publikovaných otevřených dat a přípravou pro určení adekvátních podmínek jejich užití.**

Příklad možného konkrétního způsobu posouzení přínosů, rizik i možností ošetření potenciálních rizik publikovaných datových sad na základě hodnoceného tématu, je uveden v kapitole „[Příklad ošetření navržených témat otevřených dat](#)“.

**Příklad kritérií pro hodnocení přínosů datových sad:**

- Posílení transparentnosti – dostupnost a přístup k informacím o fungování orgánů veřejné správy.
- Podpora ekonomického růstu – zdroj pro nové inovace, služby, produkty a podnikatelské příležitosti.
- Zlepšení služeb veřejné správy a zlepšení kvality života – informování veřejnosti o plánovaných změnách a akcích, vytváření větší důvěry mezi občanem a orgánem veřejné správy.
- Podpora opětovného použití dat – strojová čitelnost, dostupnost metadat a právní otevřenost dat



zajišťuje větší opakovanou využitelnost dat veřejného sektoru.

- Zlepšení vnímání veřejné správy veřejností – otevřenější a transparentnější orgány veřejné správy.
- Zlepšení procesů a dat veřejné správy, zlepšení komunikace a spolupráce veřejné správy.
- Zamezení chyb vzniklých při práci s daty – dostupnost dat ve strojově čitelném formátu minimalizuje potřebu manuálního přepisování dat z dokumentů.
- Snížení počtu dotazů dle zák. č. 106/1999 Sb.
- Pořádek ve vlastních datech – publikace umožní (donutí) orgán veřejné správy uspořádat svá vlastní data a optimalizovat vlastní interní procesy,
- Zvýšení hodnoty dat – strojová čitelnost a otevřenost dat spolu s metadaty přidávají datům novou hodnotu,

#### **Příklad kritérií pro identifikaci a ohodnocení rizik datových sad:**

- Zveřejnění dat v rozporu se zákonem – ověření, zda publikaci nebrání související legislativa.
- Porušení ochrany obchodního tajemství – zvážení, zda publikace neporušuje obchodní tajemství.
- Porušení ochrany osobních údajů – ověření, zda předmětem publikace nejsou chráněné osobní údaje.
- Zveřejnění nevhodných dat či informací – zvážení, zda publikovaná data nebudou mít nějaké negativní dopady.
- Dezinterpretace dat – posouzení, zda otevřená data nebude možno dezinterpretovat, např. nevysvětlením používaných pojmů,
- Absence konzumentů dat – existence zájmu o data ze strany budoucích uživatelů.
- Překrývání dat – potenciální překryv s jinými již publikovanými datovými sadami.
- Ohrožení bezpečnosti státu / majetku / osob.

#### **Doporučené způsoby ošetření rizik:**

- Anonymizace/agregace dat (doporučený způsob i pro publikaci neveřejných údajů publikovaných na základě principu eGovernmentu P13 stanoveným v [Informační koncepci České republiky](#)).
- Návrh vhodné komunikační strategie.
- Vytvoření (doplnění) vhodných metadat, doplňujících informací, nebo vhodné dodatečné dokumentace.
- Vhodné vymezení datové sady a propojení na související datové sady.

#### **Posouzení riziko / přínos**

Výsledek hodnocení přínosů a rizik tématu otevřených dat může ukázat, že související přínosy jsou nízké a/nebo rizika jsou příliš vysoká.

- V případě „**dobrovolného otevírání dat**“ je vhodné každé téma otevřených dat revidovat dle následujících pravidel:
  - Pokud se nepodařilo identifikovat výraznější přínos spojený s tématem, je vhodné zvážit, zda se na základě takového tématu vyplatí datovou sadu otevírat.
  - Pokud bylo identifikováno významné riziko a není možné nalézt způsob, jak riziko eliminovat, je vhodné zvážit, zda se na základě takového tématu vyplatí datovou sadu otevírat.
  - Hodnocení poměru riziko / přínos je vždy pro každého poskytovatele subjektivní, není možné stanovit přesné obecně platné pravidlo.
- V případě „**povinného otevírání dat**“ na základě povinností stanovených v legislativě, nebo na základě principu P13 „Open data by Default“ platí, že data se publikují vždy a bez ohledu na (ne)identifikované přínosy.





### 3.4 Správa katalogu témat otevřených dat

Jedná se o přehledné zachycení vybraných navržených témat datových sad s cílem:

- časového naplánování postupné publikace navržených datových sad (vytvoření **publikačního plánu**),
- kontinuální správy a průběžné revize všech vlastností již publikovaných datových sad (rizika, přínosy, podmínky užití).

Forma i způsob vedení katalogu témat otevřených dat je plně v kompetenci poskytovatele, vhodné je ale zachytit a spravovat podstatné informace o navržených tématech otevřených dat včetně informací o jejich publikaci.

#### Příklad vedených informací v katalogu témat otevřených dat:

- Schválená témata (datové sady) k publikaci v podobě otevřených dat:
  - název datové sady,
  - popis datové sady,
  - určený kurátor dat zodpovědný za analýzu a přípravu datové sady,
  - časový plán její publikace,
  - periodičita publikace (aktualizace) datové sady,
  - identifikované přínosy,
  - identifikovaná rizika a způsoby jejich případných ošetření,
  - plánovaný datum publikace.
- Informace doplněné k tématům až při přípravě datových sad k publikaci:
  - doplňující informace a doporučení pro publikaci,
  - specifikované podmínky užití
  - informace o jednotlivých distribucích,
  - datum publikace, ...

### 3.5 Analýza agendy a odvození údajů

V případě témat, ve kterých se vyskytují agendové údaje, které nejsou v RPP evidovány, nebo jejich evidence není kvalitní a není provedena dle Metodiky definice údajů vedených v agendě, je nutné analýzu agendu nejprve provést.

Agendu je nutné analyzovat dle [Metodiky tvorby a údržby sémantického slovníku pojmů veřejné správy](#) a následně odvozené údaje subjektů a objektů práva evidovat v RPP dle [Metodiky definice údajů vedených v agendě](#).

## 4 Příprava datové sady

### 4.1 Východiska přípravy datové sady

#### 4.1.1 Sémantická a syntaktická interoperabilita publikovaných dat.

Interoperabilita otevřených dat se týká schopnosti různých datových souborů, formátů a standardů spolupracovat a být vyměňovány mezi různými aplikacemi, systémy a organizacemi bez ztráty informací a údajů.

Pro přípravu datové sady k publikaci je důležité zajištění sémantické a syntaktické interoperability publikovaných



dat.

**Sémantická interoperabilita datové sady** tedy znamená, že data jsou publikována v takovém formátu, který

**Sémantická interoperabilita** zajišťuje, že při výměně a sdílení informací bude správně chápán a zachován jejich význam a smysl (datových prvků a vztahů mezi nimi).

Rozhodujícím způsobem se na tom podílí vytvoření a kvalita konceptuálních modelů datové sady (zachycení kontextu), a zejména jejich napojení na ontologie veřejné správy (Sémantický slovník pojmů veřejné správy).

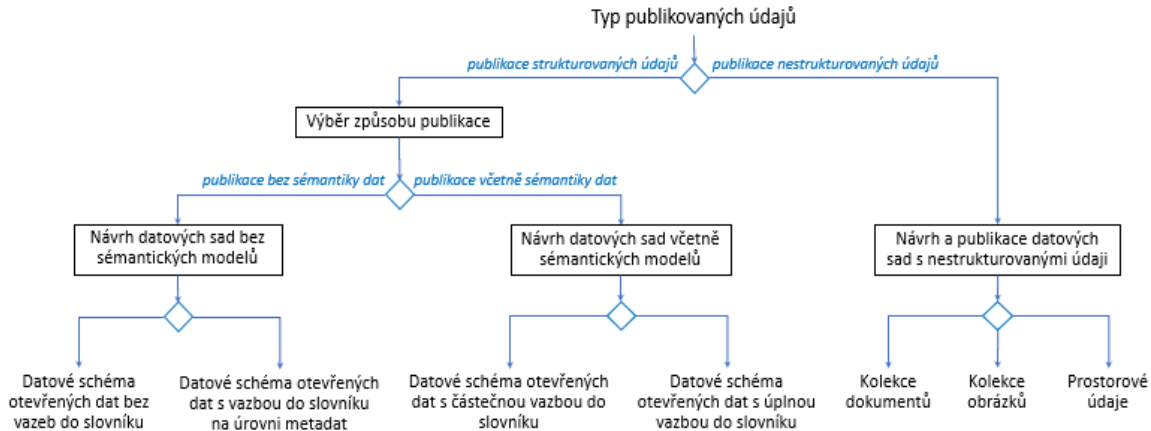
umožňuje jejich správné a konzistentní zpracování v aplikacích a systémech. Možnosti a způsoby zajištění sémantické interoperability datové sady:

- Použití standardů.
  - Standardy pro formáty dat, značkovací jazyky a ontologie usnadňují interpretaci dat mezi různými aplikacemi a systémy.
- Uvedení metadat.
  - Metadata o datové sadě usnadňují porozumění datům a způsobům, jakými jsou data uložena a jak mohou být použita.
- Přesné a konzistentní pojmenování.
  - Používání přesného a konzistentního pojmenování pro data a jejich popis podporuje sémantickou shodu mezi daty.
- Použití odkazů na další zdroje.
  - Odkazy na další zdroje, jako jsou ontologie a slovníky, pomáhají uživatelům interpretovat data a rozumět kontextu datové sady.
- Využívání globálně jednoznačných identifikátorů (IRI).
  - Použití jednoznačných identifikátorů pro data a jejich popis pomáhá při zajištění shody dat mezi různými aplikacemi a systémy.

**Syntaktická interoperabilita** zajišťuje standardizaci formátů vyměňovaných informací (dat), metadat, datových schémat a definice datových typů. Mezi prostředky, které jsou určeny k její dosažení patří vydávané otevřené formální normy, technické standardy pro publikaci otevřených dat, doporučené datové formáty a jazyky pro popis datových schémat. **Zajištění syntaktické interoperability je nezbytným předpokladem publikace otevřených dat.**

Výběr způsobů publikace datových schémat je určován přístupem a možnostmi zajištění sémantické i syntaktické interoperability.

Publikace datových sad z pohledu zajištění sémantické interoperability se v zásadě člení následujícím způsobem:



### Publikace strukturovaných údajů bez sémantiky dat:

- publikace bez existence souvisejících sémantických (konceptuálních) modelů, nebo navržených na základě sémantických modelů vytvořených „izolovaným“ způsobem mimo SSP,
- lze pouze ošetřit některá metadata datové sady jejich sémantickým napojením na SSP.
- Úsilí se zaměřuje především na zajištění syntaktické interoperability datových formátů, případně na doplnění sémantiky některých metadat v katalogizačním záznamu.
- Typy datových schémat:
- Datové schéma otevřených dat bez vazeb do slovníku.
- Datové schéma otevřených dat s vazbou do slovníku na úrovni metadat.

### Publikace strukturovaných údajů včetně sémantiky dat:

- publikace s plnohodnotnou sémantikou dat i metadat, za předpokladu, že vytvořené modely jsou součástí repozitáře sémantických modelů VS ČR a jsou vytvořeny dle příslušné metodiky.
- Úsilí je zaměřeno na sémantickou i syntaktickou interoperabilitu datových sad. To znamená zajištění sémantiky jednotlivých datových položek i sémantiky metadat v katalogizačním záznamu, v oblasti syntaktické interoperability dodržení doporučených datových modelů s povinným použitím standardizovaných otevřených formálních norem, použitím standardizovaných datových formátů a strojově čitelných datových schémat.
- Typy datových schémat:
  - Datové schéma otevřených dat s částečnou vazbou do slovníku.
  - Datové schéma otevřených dat s úplnou vazbou do slovníku.

### Publikace nestrukturovaných údajů:

- publikace dokumentů, obrázků a prostorových údajů,
- úsilí se zaměřuje především na zajištění syntaktické interoperability datových formátů,
- zajištění sémantiky dat je obtížně realizovatelné, lze pouze ošetřit některá metadata datové sady jejich sémantickým napojením na SSP.



## 4.1.2 Kontext tvorby datového schématu

Kontext přípravy datového schématu datové sady přehledově ve zkratce:

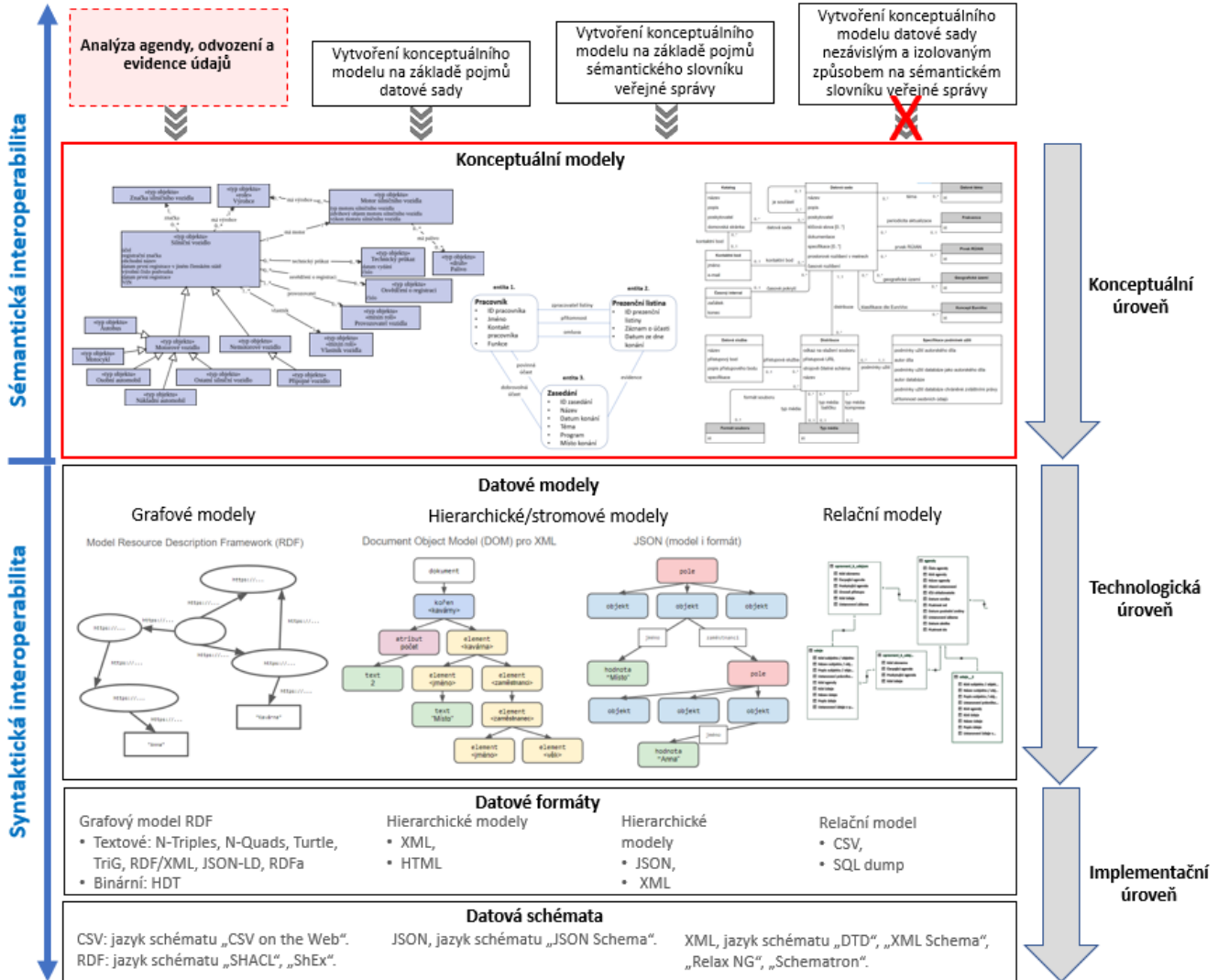
Vytvoření konceptuálního (sémantického) modelu, který zachycuje obraz reality publikovaného tématu datové sady.

Výběr datového modelu pro reprezentaci dat na základě konceptuálního modelu. Datových modelů může být vybráno více a různého typu (volba možných distribucí datové sady).

Jednotlivé oblasti přípravy datové sady se navíc významnou měrou podílí na zajištění:

- Sémantické interoperability (správného chápání a zachování smyslu sdílených informací).
  - Analýza a sémantické modelování.
  - Konceptuální modely, které jsou nezávislé na použitých technologiích,
    - existují různé možnosti jejich tvorby, cílem je ale standardizace nástrojů MV ČR,
    - vždy jsou výstupem analýzy agend a sémantického modelování datových sad (témat),
    - jsou nutným vstupem pro návrh datového modelu datové sady.
- Syntaktické interoperability (zajištění přesného a standardizovaného formátu vyměřovaných informací).
  - Datové modely – výběr způsobu reprezentace dat.
  - Datové formáty – způsob uložení dat modelu do souboru.
  - Datová schémata – v jazyce pro zápis schématu popis předepsané struktury dat čitelný lidmi i strojem.

Pohled na strukturu (kontext) přípravy datových schémat datové sady určené k publikaci strukturovaných údajů zachycuje obrázek 1 „Kontext tvorby datového schématu“.



Obrázek 1 Kontext tvorby datového schématu

Klíčovou oblastí pro přípravu datové sady je konceptuální model.

Prakticky to tedy znamená, že nedostačuje pouze „nějak“ pojmenovat jednotlivé prvky modelu (následně i prvky datové sady) a model „nějak“ vytvořit, ale je nutné se vrátit k významu slova koncept – „pojem“ a modelovat pojmy v jejich plném významu a v kontextu sémantického slovníku pojmů veřejné správy (tedy v kontextu všech dat veřejné správy). Z tohoto důvodu jsou konceptuální modelování a tvorba konceptuálního modelu publikovaných otevřených dat detailněji vysvětleny v kapitole „[Popis sémantiky veřejné správy](#)“.

Úplné informace o problematice sémantického modelování jsou uvedeny v [Metodice tvorby a údržby sémantického slovníku pojmů veřejné správy](#).

Možnosti získání / vytvoření konceptuálního modelu datové sady, nebo tématu otevřených dat:

- Vytvoření konceptuálního modelu na základě pojmů sémantického slovníku veřejné správy.
  - Model v (ne)úplné podobě je již součástí konceptuálního modelu modelované agendy.
- Vytvoření konceptuálního modelu na základě pojmů datové sady.
  - Na základě SSP, pokud obsahuje pojmy datové sady.
  - Na základě témat datových sad:



- tvorba modelu „od nuly“ (pojmy datové sady v SSP nejsou),
- téma odpovídá OFN (Ize nalézt nadřazené pojmy datové sady a využít specializace).
- Vytvoření konceptuálního modelu datové sady nezávislým a izolovaným způsobem na sémantickém slovníku veřejné správy. Model je vytvořen bez zahrnutí nových pojmů do SSP a bez jeho začlenění do celkového konceptuálního modelu veřejné správy.
  - **Jedná se o nepodporovanou variantu.**

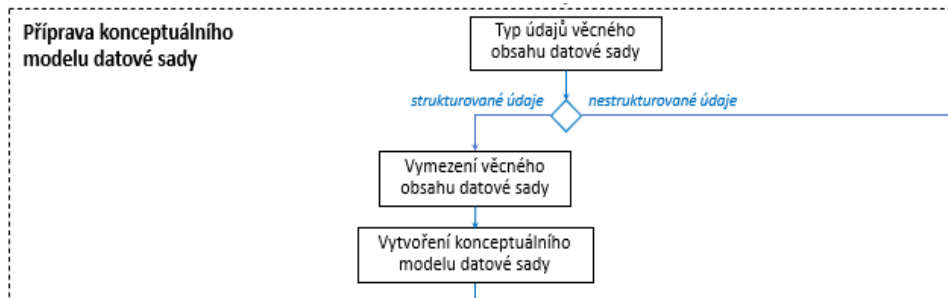
Možnosti využití existujících konceptuálních modelů v sémantickém slovníku pro návrh datových sad:

Modely lze využít bez jakýchkoliv změn, obsah datových sad je omezen pouze na oblast v modelu zachyceného kontextu.

Modely lze doplňovat o nové pojmy a vztahy, a tyto zahrnout do celkového kontextu konceptuálního modelu veřejné správy. Doplněné modely jsou k dispozici dalším poskytovatelům.

## 4.2 Příprava konceptuálního modelu datové sady a způsobu publikace

### 4.2.1 Vytvoření konceptuálního modelu datové sady.



V případě publikace nestrukturovaných údajů se způsob publikace řídí pokyny uvedenými v kapitole „[Publikace nestrukturovaných údajů](#)“.

Při výběru způsobu publikace strukturovaných údajů je nejprve nutné:

upřesnit vymezení věcného obsahu datové sady,  
následně získání /vytvoření odpovídajícího konceptuálního modelu.

#### 4.2.1.1 Vymezení věcného obsahu datové sady

Vstupem do přípravy datové sady jsou navržená témata otevřených dat z oblasti „[Identifikace témat otevřených dat](#)“.

Každé navržené téma je zpravidla definováno svým názvem, odpovědným kurátorem a stručným popisem. Toto ještě ale nemusí znamenat přesné vymezení rozsahu publikovaného kontextu tématu, a tedy ani rozsahu publikovaných údajů. Je nutné si uvědomit, že při použití zejména hierarchických datových modelů model datové



sady je obvykle určován jednou kořenovou entitou, a ta předurčuje účel využití příslušné datové sady. Při komplexnějších tématech (zastřešujících například tematickou oblast) to může odpovídající datovou sadu omezit při jejím využití. Pro komplexnější témata je jednoznačně výhodnější používat pro jejich publikaci grafové modely, které obsahují kompletní provázaný kontext celé tematické oblasti. Více na toto téma v „[Příkladu různého vymezení věcného obsahu datové sady](#)“.

Rozhodnutí o rozsahu publikovaného kontextu tématu je zásadním předpokladem pro zajištění široké využitelnosti publikovaných datových sad.

Doporučená hlediska pro vymezení rozsahu obsahu datových sad:

počet entit (subjektů a objektů práva v případě agendových údajů) s atributy v údajích a jejich vzájemné vazby – složité datové sady vedou ke komplikovaným datovým schémátům a konceptuálním modelům,  
externě stanovená vymezení tematických oblastí, např. v případě agend zákon ustavující příslušnou agendu mohou přesně vymezovat

Pokud se ukáže, že původně definovaná témata datových sad v oblasti identifikace témat nebyla zcela přesně nebo vhodně vymezená a popsána, je vhodné a nutné, témata v katalogu témat datových sad upravit takovým způsobem, aby jejich publikace byla jednoduchá.

#### 4.2.1.2 Vytvoření konceptuálního modelu datové sady

Na základě vymezení obsahu datové sady je nutné zachytit také strukturu obsahu. Ideálním způsobem je zachycení formou konceptuálního modelu, který zachytí nejen strukturu obsahu, ale také jeho kontext (vnitřní vazby prvků reprezentující obsah a jeho význam) a v případě využití pojmů sémantického slovníku veřejné správy také přesný význam jednotlivých prvků konceptuálního modelu v ontologii veřejné správy ČR.

##### Koncept datové sady

V případě, že nelze konceptuální model vytvořit, nebo kdyby vytvořený model nedával smysl (například v případě publikace jednoduché tabulky), bude dostačující obsah zamýšlené datové sady zachytit způsobem, ze kterého bude zřejmé, co přesně bude obsahem publikované datové sady, jaký je význam jednotlivých položek obsahu a jak bude struktura datové sady vypadat (tabulka, seznam pojmů, grafické zobrazení struktury, ...). V dalším textu bude tento způsob zachycení obsahu datové sady označován pojmem „**koncept datové sady**“.

Koncept datové sady lze považovat za prvotní představu vymezení kontextu publikovaného tématu, na jehož základě se v dalších krocích namodeluje plnohodnotný konceptuální model.

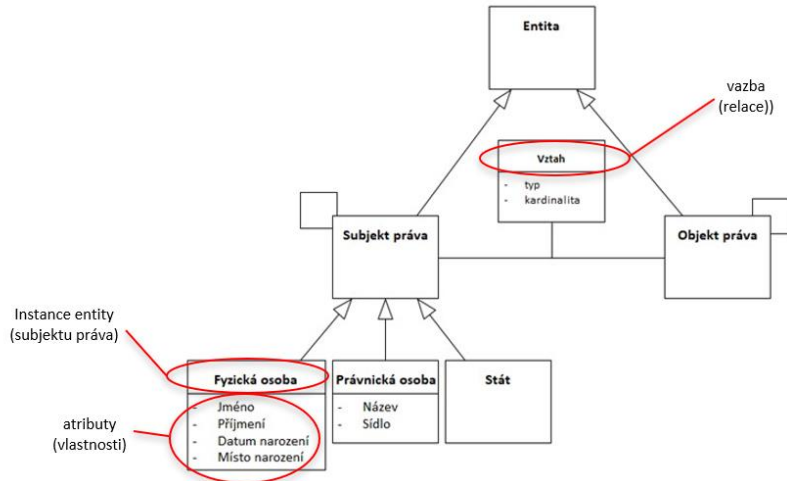
##### Konceptuální model

Konceptuální model je zcela obecný a základní pojmový model, nezávislý na jakékoliv potenciální implementaci jeho obsahu.

Zajištění sémantické interoperability je důležitou podmínkou využitelnosti publikovaných dat. Z tohoto důvodu je problematika konceptuálního modelování a tvorby konceptuálního modelu publikovaných otevřených dat zařazena do kapitoly „[Popis sémantiky veřejné správy](#)“.

Ve zkratce to prakticky znamená, že je nutné se vrátit k významu slova koncept – „pojmem“ a modelovat pojmy v jejich plném významu a v kontextu sémantického slovníku pojmů veřejné správy (a v kontextu všech dat veřejné správy).

Základní prvky konceptuálního modelu.



- Entita je základním pojmem při popisu reálného světa a práci s informacemi o něm. Entita musí být rozlišitelná od ostatních entit a existovat nezávisle na nich.
- Mezi jednotlivými entitami mohou existovat vzájemné vztahy s uvedenými kardinalitami.
- V legislativních dokumentech jsou tyto entity, jejich vlastnosti, vztahy, role a související události popisovány „pojmy“, tyto jsou předmětem sémantického modelování.
- V různých konceptuálních modelech se mohou entity a s nimi související informace vyskytovat s různými názvy, vždy v závislosti na použité metodě modelování a způsobu tvorby konceptuálního modelu. V metodice je pro oblast otevřených dat preferována terminologie i modely vycházející ze sémantického modelování.
- RPP používá pro označení entit při evidenci údajů poskytovaných agendou pojmy subjekty a objekty práva.
- Subjekty práva se rozumí osoby, které mohou být zavázány povinnostmi nebo oprávněny právem, mají vlastní vůli a schopnost konat.
- Objekty práva jsou předměty, věci hmotné či nehmotné, které nemají vlastní vůli a rozum, nejsou jim ukládány povinnosti ani přiznávána práva, se kterými subjekty práva nakládají a mohou k nim nabývat různých vztahů.

### Možnosti vytvoření konceptuálního modelu

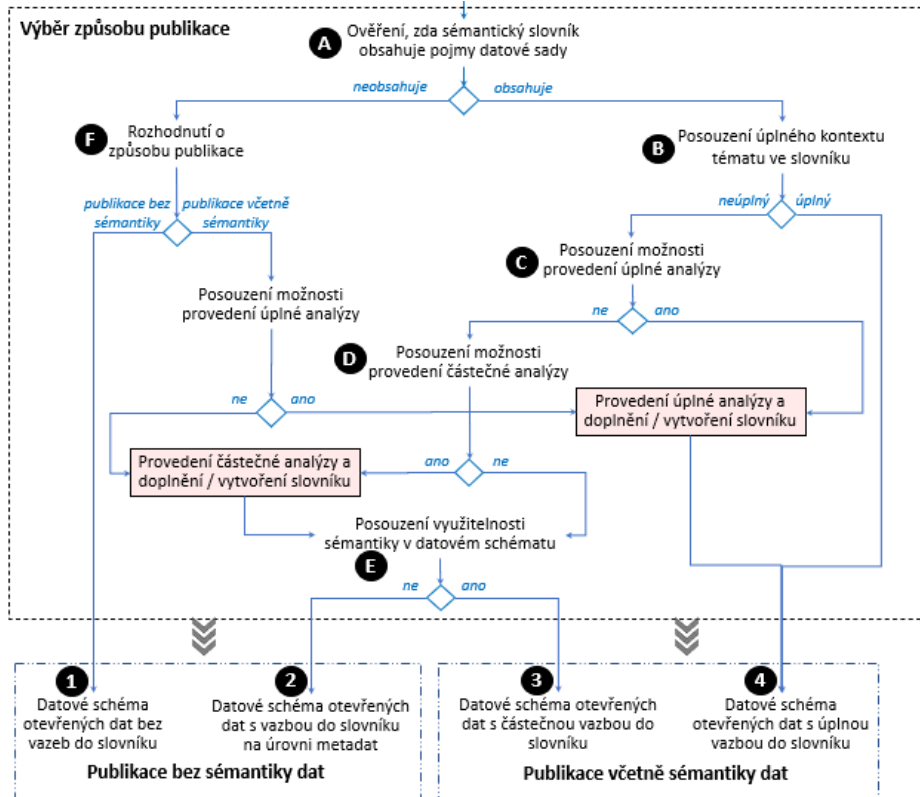
- Vytvoření konceptuálního modelu na základě pojmů sémantického slovníku veřejné správy dle postupů uvedených v „[Metodice tvorby a údržby konceptuálních datových modelů agend](#)“.
  - Pokud model (nebo většina jeho prvků) v nějaké (ne)úplné podobě je již součástí konceptuálního modelu veřejné správy (již namodelované agendy, domény, datové sady, ...) je ho možné převzít a pouze doplnit o chybějící prvky (objekty, subjekty, vazby, vlastnosti).
  - Pokud model neexistuje je nutné ho vytvořit:
    - na základě existujících pojmů SSP, pokud již SSP pojmy datové sady obsahuje,
    - na základě pojmů témat (oblastí) datových sad, ty ale musí být nejdříve zavedeny do glosáře pojmů,
    - na základě existujících OFN, pokud odpovídají tématu datové sady, v tomto případě lze využít pojmy OFN ke specializaci nových pojmů datové sady, a tyto zavést do glosáře pojmů.
- Vytvoření konceptuálního modelu datové sady nezávislým a izolovaným způsobem bez sémantického slovníku veřejné správy. Model je vytvořen bez zahrnutí nových pojmů do SSP a bez jeho začlenění do celkového konceptuálního modelu veřejné správy.





- Tato možnost je přijatelná pouze pro publikaci bez sémantiky dat.
- **Jedná se o nepodporovanou variantu.**

## 4.2.2 Výběr způsobu publikace



Výběr způsobu publikace se týká výhradně publikace strukturovaných údajů. Rozhodování začíná ověřením (A), do jaké míry jsou pojmy vytvořeného konceptuálního modelu datové sady, nebo jeho koncepčního návrhu již zavedeny v sémantickém slovníku pojmů veřejné správy.

Pokud slovník pojmy již obsahuje je důležité, zda je ve slovníku zachycen (B) i úplný kontext datové sady (konceptuální model pojmů veřejné správy obsahuje všechny vazby a všechny pojmy konceptuálního modelu). Pokud ano, tak nejlepší volbou je publikace datové sady s úplnou vazbou významu datových položek do slovníku (4).

Pokud ne, tak je nutné zvážit možnosti (připravenost podkladů a organizace) k dokončení úplné (C) analýzy tématu/tematické oblasti, následně ji také provést a publikovat datovou sadu opět s úplnou vazbou do slovníku (4).

Pokud úplnou analýzu nelze provést, je nutné zvážit provedení alespoň částečné (D) analýzy zaměřené na doplnění slovníku o vybrané klíčové prvky konceptuálního modelu (klíčové entity, použití OFN a specializace jejich pojmů, ...).

Po provedení částečné analýzy je nutné posoudit (E), zda výsledky analýzy (doplnění slovníku sémantických pojmů) jsou dostatečné k publikaci (3) datové sady s částečnou vazbou sémantiky dat do slovníku (3. úroveň otevřenosti s publikací kontextu), nebo je možné sémantiku datové sady zajistit pouze na (2) úrovni metadat.

V případě, že slovník pojmy konceptuálního modelu datové sady neobsahuje, je nutné (F) rozhodnout, zda publikace proběhne bez (1) sémantiky (bez vazeb do slovníku), anebo se sémantikou. Pokud se sémantikou tak



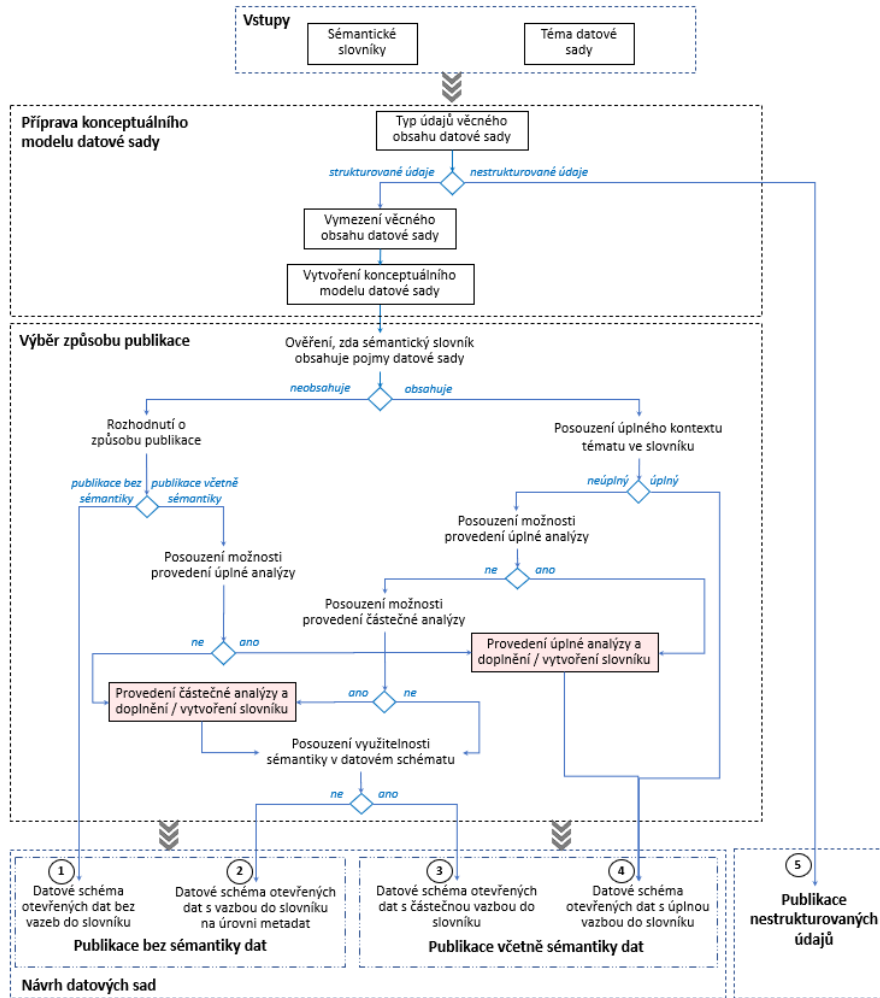
je možné provést úplnou analýzu a získat možnost publikace **(4)** s úplnou vazbou do slovníku, nebo je nutné minimálně provést částečnou analýzu s možnostmi publikace **(2)** a **(3)**.

#### Upřesnění některých činností.

- (A) Ověření, zda sémantický slovník obsahuje pojmy datové sady.
  - Rozhodování o způsobu publikace probíhá na základě konceptuálního modelu datové sady, který může být v této fázi vytvořen různými (i nepodporovanými) způsoby a jeho kvalita i úplnost může být rozmanitá. Rovněž lze případně pro rozhodování využít koncept datové sady, pokud pro publikované téma konceptuální model nelze smysluplně vytvořit.
  - Rozhodování je založeno na zjištění, zda a do jaké míry jsou pojmy konceptuálního modelu datové sady (entity, vazby, atributy) již zavedeny v [SSP](#).
  - Posouzení dostatečnosti je individuální, ale mělo by být pravidlem, že pro rozhodnutí „obsahuje“ by mělo platit, že většina subjektů a objektů vyskytující se v konceptuálním modelu (nebo jejich nadřazených pojmů) by měla být v [SSP](#) již zavedena.
- (B) Posouzení úplného kontextu tématu ve slovníku [SSP](#).
- (C) Posouzení možnosti provedení úplné analýzy.
  - Posouzení možností provedení úplné analýzy je možné provést s detailní znalostí věcné problematiky publikovaného tématu datové sady (kurátor dat, ohlašovatel agendy, datový analytik, ...).
- (D) Posouzení možnosti provedení částečné analýzy.
  - Platí stejné podmínky jako u úplné analýzy.
- (E) Posouzení využitelnosti sémantiky v datovém schématu (kurátor dat, datový analytik).
- (F) Rozhodnutí o způsobu publikace.
- Provedení úplné / částečné analýzy a doplnění / vytvoření slovníku.
  - Vysvětlení a doplňující informace uvedeny v kapitole „[Popis sémantiky veřejné správy](#)“.
  - Doporučené postupy:
    - „[Metodika tvorby a údržby konceptuálních datových modelů agend](#)“.
    - Záznam školení „[Modelování významu dat ve veřejné správě](#)“.



### 4.2.3 Přiřazení rolí k činnostem přípravy konceptuálního modelu datové sady



#### Zapojení rolí ve skupinách činností:

	Koordinátor otevírání dat	Kurátor otevřených dat	IT specialista	Datový	Specialista na modelovanou doménu	Ohlašovatel agendy
Z – zodpovídá S – spolupracuje R – rozhoduje						
Vymezení věcného obsahu datové sady		Z	S	S	S	S
Vytvoření konceptuálního modelu datové sady		Z		S	S	S
Ověření, zda sémantický slovník obsahuje pojmy datové sady		S			Z	
Rozhodnutí o způsobu publikace	S	S			Z	S
Posouzení úplného kontextu tématu ve slovníku		S			Z	S
Posouzení možnosti provedení úplné analýzy		S			Z	S
Posouzení možnosti provedení částečné analýzy		S			Z	S
Posouzení využitelnosti sémantiky v datovém schématu		S		Z	S	
Provedení úplné analýzy a doplnění / vytvoření slovníku	S	S		S	Z	



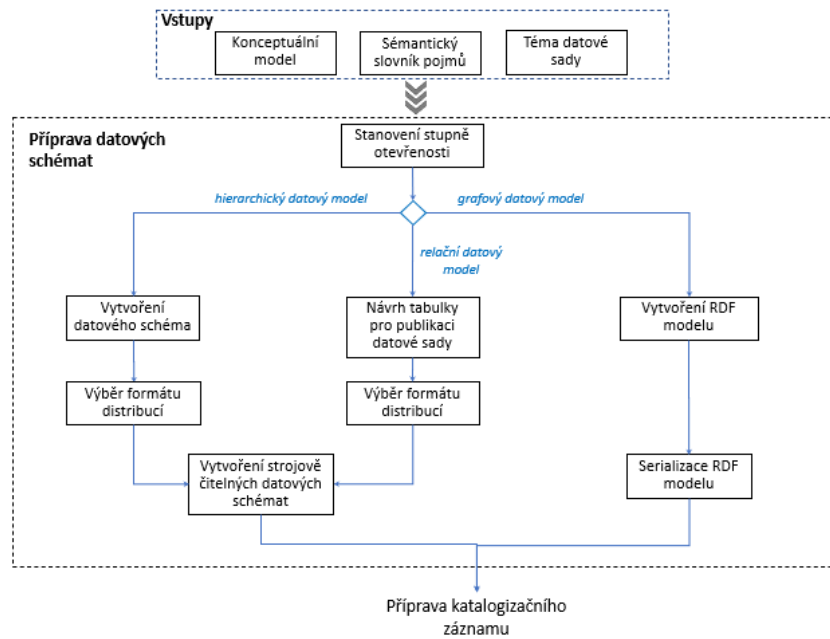
Provedení částečné analýzy a doplnění / vytvoření slovníku	S	S		S	Z	
--	---	---	--	---	---	--

## 4.3 Návrh datových modelů a datových schémat

### 4.3.1 Výběr způsobu reprezentace strukturovaných údajů – datových modelů

Pro přípravu datových schémat distribucí publikované datové sady je nutné mít připravený konceptuální model, nebo vybraný kontext odpovídající rozsahu datové sady.

Možnosti reprezentace konceptuálního modelu datovými modely v závislosti na jejich typech:



Z – zodpovídá

S – spolupracuje

R – rozhoduje

	Koordinátor otevírání dat	Kurátor otevřených dat	IT specialista	Datový architekt	Specialista na modelovanou doménu	Ohlašovatel agenty
--	------------------------------	------------------------------	-------------------	---------------------	---	-----------------------

Stanovení stupně otevřenosti

Vytvoření datového schéma

Výběr formátu distribucí

Vytvoření strojově čitelných datových schémat

Návrh tabulky pro publikaci datové sady

Vytvoření RDF modelu

Serializace RDF modelu

S	Z	S		S	
	S	S	Z	S	
S	Z	S	S		
	S	S	Z		
S	S	S	Z		
	S	S	Z	S	
		S	Z	S	



## 4.3.2 Obecné charakteristiky publikované datové sady

### 4.3.2.1 Stanovení stupně otevřenosti

Prvním rozhodnutím poskytovatele je výběr způsobu reprezentace publikovaného tématu dle zamýšleného stupně otevřenosti publikovaných datových sad.

- V kontextu otevřených dat veřejné správy České republiky je nutné pro každou publikovanou datovou sadu zvolit minimálně 3. stupeň otevřenosti (vysvětlení stupňů otevřenosti je uvedeno v kapitole [“Stanovení stupně otevřenosti”](#)).
  - Stupně otevřenosti 4 a 5 s sebou nesou větší pracnost při přípravě datové sady k publikaci.
  - Je možné zvolit i více stupňů najednou (včetně 1. či 2.), platí však, že alespoň jeden ze stupňů musí být minimálně na úrovni 3. (vyplývá ze [§ 4b odst. 1 zákona č. 106/1999 Sb.](#) o svobodném přístupu k informacím).
- S určením stupně otevřenosti souvisí i technická podoba zveřejnění datové sady, tedy:
  - v jakých distribucích bude obsah datové sady zveřejněn,
  - v jakém datovém formátu budou distribuce vyjádřeny.

### 4.3.2.2 Určení kódování dat datových sad

Při předávání údajů mezi informačními systémy ve veřejné správě musí být vždy použito znakové sady Unicode a kódování UTF-8. Pokud je při předávání údajů využita i jiná znaková sada, pak údaj musí být povinně uveden i ve znakové sadě Unicode v kódování UTF-8 a takto uvedený údaj je referenční. Údaj v jiné znakové sadě je pouze dodatečným tvarem údaje a má pouze informativní charakter.

## 4.3.3 Tabulková data (relační datový model)

### 4.3.3.1 Návrh tabulky pro publikaci 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 ve formátu [Comma Separated Values \(CSV\)](#), (více k formátu CSV e-learning „Technické aspekty otevřených dat“ modul č.7 „[Formáty pro otevřená data – CSV](#)“).

### 4.3.3.2 Tabulka + využití sdílených specifikací

V případě publikace údajů, které jsou specifikovány [sdílenými specifikacemi OFN](#), je nutné strukturu těchto údajů (hlavičku tabulky) vyjadřovat ve struktuře definované sdílenou specifikací.

Seznam vydaných sdílených specifikací je dostupný na stránkách [POD](#).

### 4.3.3.3 Sémantika publikovaných tabulkových dat

Sémantiku datové sady na úrovni metadat je možné zajistit provázáním jejích metadat s SSP. Postup je uveden v kapitole [Příprava katalogizačního záznamu datové sady](#).

Sémantiku datové sady na úrovni dat lze zajistit správným mapováním pomocí formátu „[CSV on the Web](#)“.



#### 4.3.4 Hierarchicky strukturovaná data (hierarchický datový model)

Hierarchicky strukturovaná data jsou data, která jsou organizovaná do stromové struktury, kde každý prvek datové sady je reprezentován uzlem stromu. Každý uzel může mít jednoho nebo více potomků (děti), jednoho nebo více předků, přičemž nejbližší předek je vždy jeho rodičem, kromě kořene stromu, který nemá žádného předka.

Hierarchický datový model je z konceptuálního modelu odvozen následujícím postupem:

- výběr kořenové entity (kořene stromu),
- následné postupné doplňování struktury potomky odvozenými z vazeb mezi prvky konceptuálního modelu.

Distribuce datové sady, která obsahuje hierarchické strukturované objekty (tj. objekty složené z jiných objektů), musí splňovat následující doporučení a 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 (více k formátu CSV e-learning „Technické aspekty otevřených dat“ modul č.6 „[Formáty pro otevřená data – XML](#)“).
  - 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 (více k formátu CSV e-learning „Technické aspekty otevřených dat“ modul č.7 „[Formáty pro otevřená data – JSON](#)“).

Pro výměnu dat je důležité použití **standardizovaných formátů**, které umožňují různým systémům a aplikacím snadnou interpretaci dat a jejich využití. Z tohoto důvodu je nutné při publikaci dat dodržovat určitá pravidla a standardy, aby byla zajištěna správnost a srozumitelnost dat pro všechny zúčastněné strany. To zahrnuje standardizaci názvů a popisů datových prvků, jakož i standardizaci vztahů mezi jednotlivými prvky v hierarchické struktuře.

K zajištění standardizace jsou určeny Otevřené formální normy (OFN), což jsou technická doporučení zaměřená na vybrané datové sady. Zajišťují, že stejná data publikovaná různými poskytovateli budou interoperabilní a bude je možné jednodušeji využívat nezávisle na tom, kdo data poskytuje. Pro poskytovatele otevřených dat, kteří jsou [povinnými subjekty](#) dle zákona č. 106/1999 Sb. o svobodném přístupu k informacím, jsou doporučení vyplývající z OFN závazná dle [§ 4b](#).

[Definice Otevřené formální normy](#) v zákoně č. 106/1999 Sb.

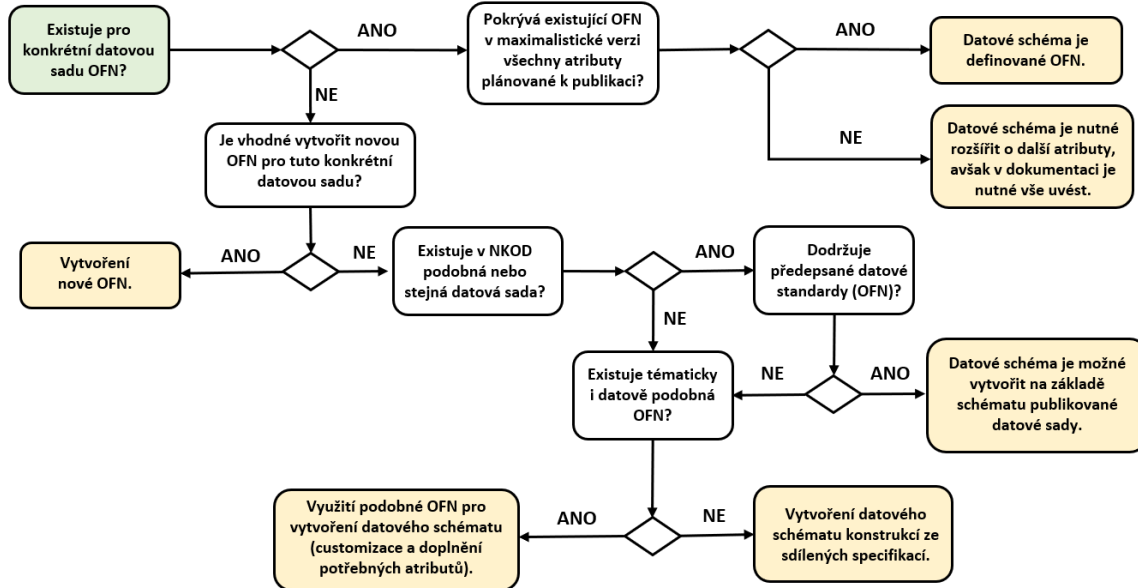
*(3) Otevřenou formální normou se pro účely tohoto zákona rozumí pravidlo, které bylo vydáno písemně a obsahuje specifikace požadavků na zajištění schopnosti různých programových vybavení vzájemně si poskytovat služby a efektivně spolupracovat.*



Podrobnější informace k otevřeným formálním normám jsou k dispozici v kapitole „[OFN](#)“, na stránkách POD v sekci „[Otevřené formální normy \(OFN\)](#)“ a v e-learningovém kurzu „[Úvod do otevřených formálních norem](#)“.

#### 4.3.4.1 Varianty přípravy datového schématu hierarchicky strukturovaných dat

Rozhodovací schéma pro výběr správné varianty přípravy datové sady.



Obrázek 2 Varianty přípravy hierarchicky strukturovaných dat

- Příprava datového schématu na základě OFN.
  - OFN existuje pro konkrétní téma datové sady,
  - OFN definuje datové schéma tématu konkrétní datové sady, katalogizační záznam, sémantiku vazbou do SSP.
- Příprava datového schématu na základě podobné OFN.
  - OFN může být základem pro podobné datové sady (využití specializace pojmů konceptuálního modelu datové sady),
  - OFN konkrétního tématu lze použít s malými úpravami pro návrh datového schématu podobných datových sad, katalogizační záznam i strojově vyjádřené schéma distribuce datové sady je nutné modifikovat dle tématu a sémantiku doplnit specializací pojmů OFN dle pojmů datové sady.
- Příprava datového schématu s využitím sdílených specifikací.
  - OFN pro konkrétní téma datové sady neexistuje a ani nelze využít podobné OFN prostřednictvím její specializace,
  - základem datového schématu je konceptuální model / koncept budoucí publikované datové sady a pro detailní vypracování datového schématu jsou využity sdílené specifikace,
  - sémantiku je nutné řešit zavedením pojmů datové sady do SSP a vytvořením konceptuálního modelu dle předepsané metodiky,
  - strojově vyjádřené schéma distribuce datové sady je nutné připravit dle struktury datového schématu.
- Příprava datového schématu na základě datového schématu již publikované datové sady registrované v NKOD
  - postup využití je stejný, jako v případě použití podobné OFN, pouze s tím rozdílem, že základem



- není OFN,
  - tuto možnost lze využít pouze v těch případech, kdy použitá publikovaná datová sada je publikované některou z výše uvedených možností.
- Vytvoření nové OFN a její následné použití pro přípravu datového schématu.
  - tato možnost vyžaduje příslib širokého využití více poskytovateli, neboť příprava nové OFN vyžaduje spolupráci více poskytovatelů organizovanou garantem otevřených dat VS, je pracnější a časově náročná,
  - vypracovaná nová OFN následně nabídne stejné možnosti využití jako již připravené OFN,
  - při výběru této varianty je nutné mít již předem prodiskutovaný návrh nové OFN s dalšími poskytovateli, kteří by data podle takové OFN měli zájem publikovat, a pak následně kontaktovat garanta otevřených dat VS.

#### 4.3.4.2 Příprava datového schématu na základě OFN

Otevřené formální normy ve smyslu [§ 3a odst. 3 zákona č. 106/1999 Sb., o svobodném přístupu k informacím](#) jsou pro poskytovatele otevřených dat, kteří jsou povinnými subjekty dle [§ 4b zákona č. 106/1999 Sb. o svobodném přístupu k informacím](#), závazné.

Jedná se o technická doporučení zaměřená na vybrané datové sady, která zajišťují, že stejná data publikovaná různými poskytovateli budou interoperabilní. Tím je umožněno taková data jednodušeji využívat nezávisle na tom, od kterého jsou poskytovatele.

Otevřená formální norma (OFN) specifikuje, jak publikovat data o konkrétních věcech (např. sportovištích, akcích apod).

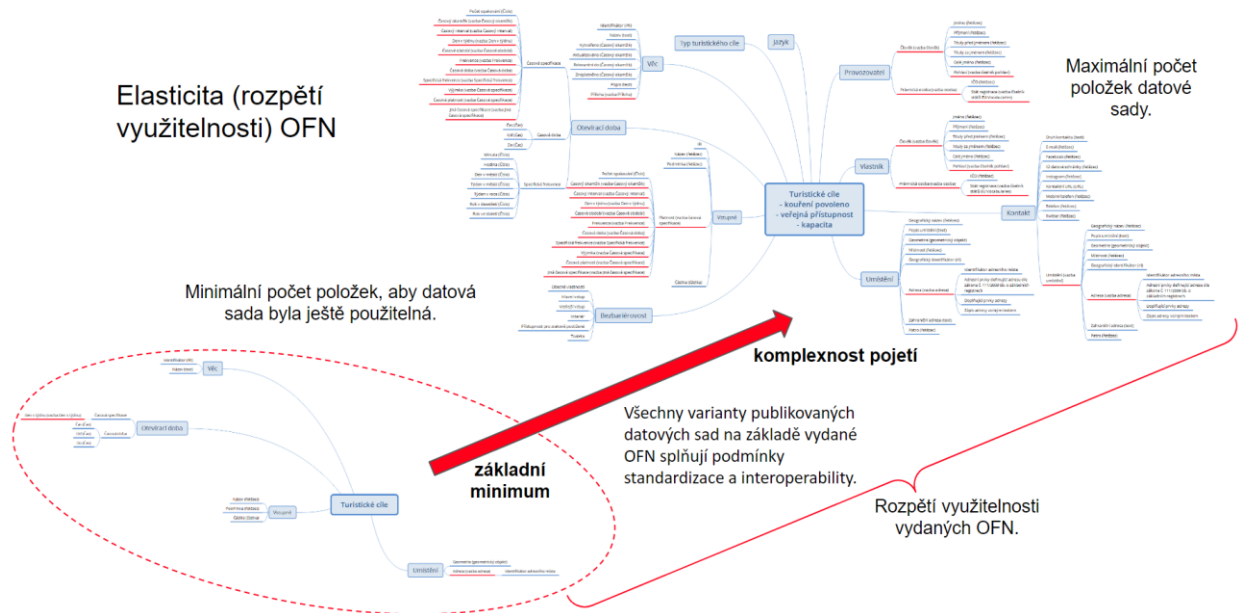
Smysl OFN a zamýšlený způsob použití je ve zkratce následující:

**Pokud se publikují data, která významově odpovídají tomu, co je zachyceno v modelu, použije se specifikace odpovídající OFN.**

Všechny položky jsou nepovinné, tj. pokud některé nejsou k dispozici, a objektivně nejsou nutné, nic se neděje, prostě v datech nebudou.

U každé OFN je uveden "minimalistický" příklad, který vymezuje, pod

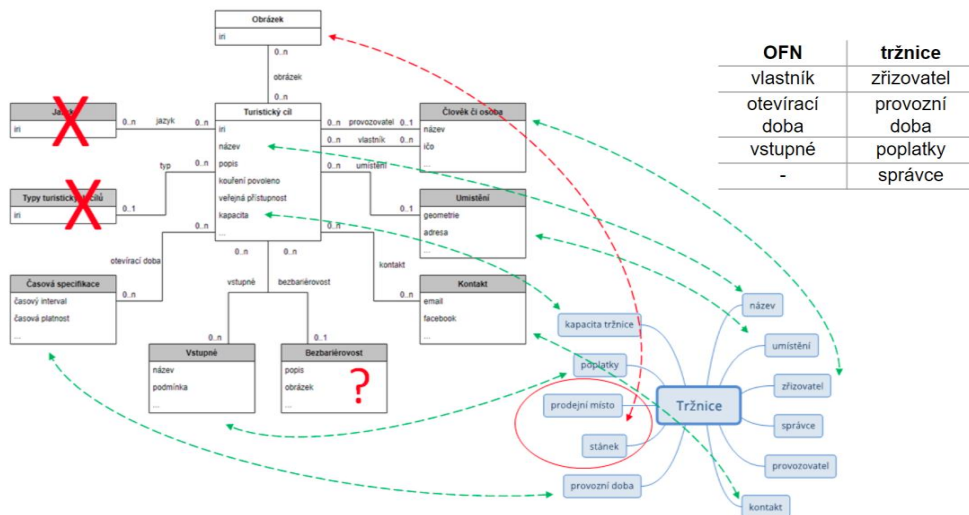




[Návod k použití OFN poskytovateli dat.](#)

### 4.3.4.3 Příprava datového schématu na základě podobné OFN

Postup je založen na podobnosti hierarchických datových modelů. Taková podobnost je přirozená v případech, kdy oba modely jsou v podstatě specializací společného nadřazeného tématu.



V případě zachyceném na obrázku se jedná o datový model turistického cíle a datový model tržnice. Je snadné si představit, že v obou případech se jedná o specializaci nadřazeného datového modelu „Veřejně přístupného místa“ a je tedy přirozené, že většina atributů kořenové entity bude shodná a případné odlišnosti budou souviset pouze s jejich specializací. Využití podobnosti OFN je v takových případech poměrně snadné, a i přizpůsobení připravených strojově čitelných datových schémat není obtížné.

V případě publikace datové sady s částečnou vazbou do sémantického slovníku pojmů úpravy souboru s publikovaným kontextem schématu (soubory typu JSON-LD kontext v OFN) také nejsou obtížné, za předpokladu, že sémantický slovník obsahuje pojmy odpovídající významu jednotlivých datových položek datové sady. Pokud ve slovníku pojmy zavedeny nejsou, a datová sada má být publikována se sémantikou, je



nutné provést analýzu datové sady dle „[Metodiky tvorby a údržby konceptuálních datových modelů agend](#)“ a pojmy do slovníku doplnit.

Podobnosti OFN lze využít pro návrh datového schématu i v těch případech, kdy datové modely neodpovídají specializaci nadřazeného modelu, avšak v těchto případech budou úpravy modelu a strojově čitelného schématu složitější.

Postup přípravy datového schématu na základě podobnosti OFN je podrobněji uveden v kapitole „[Příklad přípravy datového schématu na základě podobné OFN](#)“.

Doplňující informace včetně výkladu jsou k dispozici i v e-learningovém kurzu „[Úvod do otevřených formálních norem \(OFN\)](#)“ na Portálu otevřených dat v sekci vzdělávání. Problematice využití podobných OFN se věnuje modul „4. Využití připravených OFN pro tvorbu datových schémat“.

#### **4.3.4.4 Příprava datového schématu s využitím sdílených specifikací.**

Příprava datového schématu v tematických oblastech, kde nejsou k dispozici připravené kompletní OFN pro konkrétní témata, je založena na:

- datovém modelu (vycházejícím z konceptuálního modelu nebo konceptu) datové sady k publikaci,
- využití prvků repositáře OFN (sdílených specifikací a základních datových typů) k postupnému nahrazování jednotlivých položek datového modelu vhodnými sdílenými specifikacemi.

S výhodou je možné využít skutečnosti, že dekompozice sdílených specifikací je v podstatě základem jejich konstrukce, a že tato dekompozice je plně zachycena v repositáři OFN prostřednictvím vzájemných referencí. Repositář OFN si tak lze představit jako množinu hierarchických stromových modelů dat (reprezentující datová schémata a vyjádřených ve formátu JSON), které navzájem sdílí společné prvky repositáře od těch nejjednodušších (základních datových typů) až po ty nejsložitější (specifické OFN).

Samotný návrh datového schématu ze strany poskytovatele se tak zaměřuje zejména na návrh datového modelu (odvození z konceptuálního modelu) a postupnou náhradu jeho položek prostřednictvím referenčních odkazů na sdílené specifikace a základní datové typy. Při jejich náhradě stačí pro každou položku datového schématu vždy jenom odkazovat na vrchol hierarchického datového modelu odpovídající ošetřované položce datového schématu (odpovídající základnímu datovému typu, nebo sdílené specifikaci).

Návrh datového schématu je tak posunut na logicky vyšší úroveň, kdy není nutné řešit technické a jiné detaily a stačí se plně soustředit na zachycení věcné tematiky publikované oblasti.

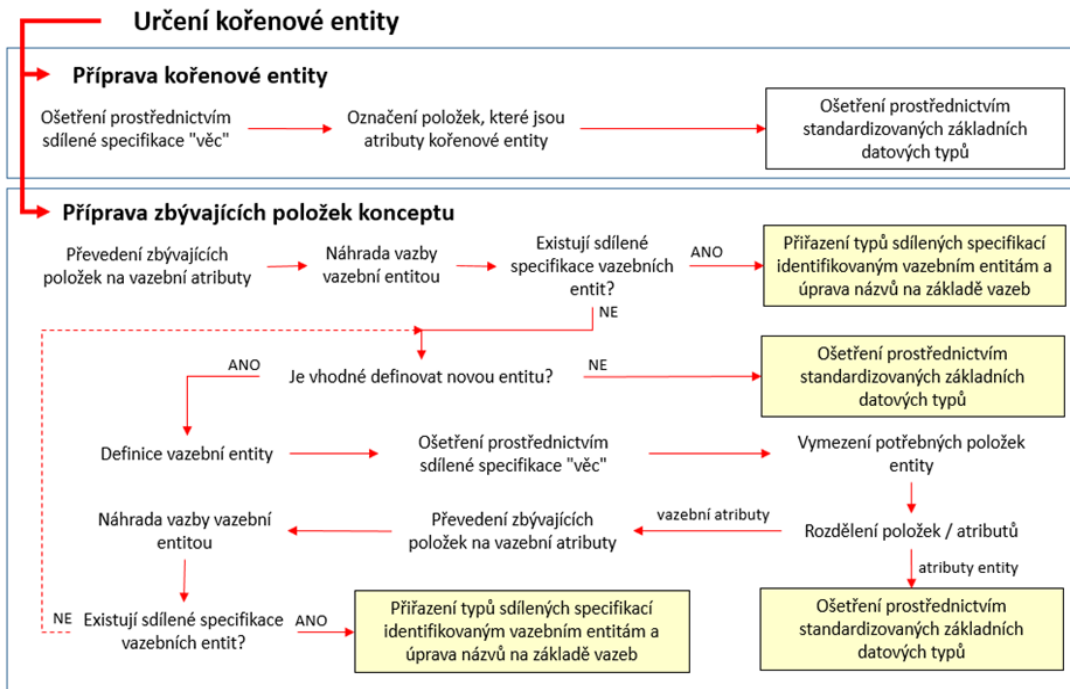
#### **Tvorba datového schématu.**

Postup tvorby datového schématu z datového modelu využitím sdílených specifikací základních datových typů.

- Identifikace a příprava kořenové entity:
  - určení kořenové entity připraveného konceptu (schématu) - určuje téma datové sady,
  - kořenová entita bude většinou vycházet ze subjektů nebo objektů práva agendy, jejíž veřejná data budou publikována,
  - ošetření kořenové entity prostřednictvím sdílené specifikace "věc",
  - označení položek, které jsou atributy kořenové entity,
  - ošetření atributů základními datovými typy z Repositáře OFN.
- Příprava zbývajících položek konceptu:
  - převedení zbývajících položek na vazební atributy a pojmenování vazby,
  - specifikace entity, na kterou vazba odkazuje.
- Ošetření entit navázaných vazbami na kořenovou entitu:
  - nahrazení navázaných entit sdílenými specifikacemi (pokud jsou k dispozici v Repositáři OFN),



- úprava jejich názvů na základě vazeb.
- Ošetření entit navázaných vazbami na kořenovou entitu a neodpovídajících sdíleným specifikacím:
  - ošetření navázané entity prostřednictvím sdílené specifikace "věc",
  - označení položek, které jsou atributy navázané entity,
  - ošetření atributů základními datovými typy z Repositáře OFN.
  - převedení zbývajících položek na vazební atributy a pojmenování vazby,
  - specifikace entity, na kterou vazba směřuje.
  - nahrazení navázaných entit sdílenými specifikacemi (pokud jsou k dispozici v Repositáři OFN),
  - úprava jejich názvů na základě vazeb.
- Datové schéma je nutné vyjádřit strojově čitelným způsobem ve formátu JSON.



Podrobnější výklad problematiky je v kapitole „[Příprava datového schématu s využitím sdílených specifikací](#).“

Doplňující informace včetně výkladu jsou k dispozici i v e-learningovém kurzu „[Úvod do otevřených formálních norem \(OFN\)](#)“ na Portálu otevřených dat v sekci vzdělávání. Problematice využití sdílených specifikací se věnuje modul „5. Tvorba datových schémat s využitím sdílených specifikací“.

#### 4.3.4.5 Příprava datového schématu (konceptuálního modelu) na základě datového schématu (konceptuálního modelu), již publikované a registrované datové sady v NKOD

Návrh datového schématu je v principu shodný, jako v případě „[Přípravy datového schématu na základě OFN](#)“.

Rozdíl spočívá především ve skutečnosti, že v případě OFN se jedná o oficiálně vydanou normu, jejíž kvalita a úplnost je garantovaná garantem otevřených dat veřejné správy a datové schéma je připraveno včetně zajištění sémantiky dat.

V případě využití datového schématu již publikované datové sady si musí poskytovatel nejprve sám ověřit jeho kvalitu, úplnost a správnost použití sdílených specifikací, případně i správné použití celé OFN. Částečně mu v tom může pomoci katalogizační záznam datové sady v NKOD, kde v položce metadat „Specifikace“ by měly být uvedeny odkazy a informace o využitých specifikacích (OFN, datové specifikace, sdílené specifikace, standardy,



...).

#### 4.3.4.6 Vytvoření nové OFN a její následné použití pro přípravu datového schématu.

Nová OFN se v zásadě připravuje v následujících situacích:

- Existuje předpoklad (a zájem) více poskytovatelů publikovat otevřená data zaměřená na stejné tematické oblasti (většinou se týká samosprávy). OFN v těchto případech zajistí syntaktickou i sémantickou interoperabilitu publikovaných dat, a tím připravuje podmínky pro snadné užití publikovaných datových sad. Předpokladem je tedy „masovost“ používání OFN, a také zapojení více potenciálních poskytovatelů do jejich přípravy.
- Jedná se o OFN, které mají „specifický charakter“ (nejsou vázány na zapojení potenciálních poskytovatelů do jejich přípravy, např. „Rozhraní katalogů otevřených dat“), nebo o situace, kdy OFN jsou důležité pro specifikaci požadavků na dodavatele některých ISVS (např. „Registr práv a povinností“). V těchto případech příprava OFN je plně v kompetenci garanta otevřených dat.

Průběh přípravy nové OFN určené pro „masové“ použití:

- Tvorba nové OFN musí být vždy opodstatněná výhledem budoucího hromadného využití připravované OFN.
- Vydání nové OFN zajišťuje a autorizuje vždy garant OD veřejné správy ČR ([§ 3a odst. 3 zákona č. 106/1999 Sb.](#)).
- Nová OFN nemůže vzniknout jako výsledek návrhu jednoho poskytovatele dat, ale vždy se jedná o společný návrh více poskytovatelů za účasti garanta OD. Je tedy nutné zaangažovat více potenciálních poskytovatelů publikace připravovaného tématu a celou přípravu koordinovat garantem OD.
- Klíčovým vstupem pro přípravu je úplný kontext tématu otevřených dat vyjádřený technologicky nezávislým konceptuálním modelem, vytvořeným na základě slovníku pojmů veřejné správy. Konceptuální model i vymezení kontextu tématu musí být ve shodě s představami všech do tvorby zapojených poskytovatelů, a také jimi musí být schválen.
- Pro všechny navržené datové modely musí být vytvořeny strojově čitelná datová schémata a připravena metadata pro katalogizaci datových sad vytvořených na základě vytvořené OFN.
- Dokumentace nové OFN musí být úplná a následně publikovaná na stránkách POD.

V případě, že se nějaký konkrétní poskytovatel rozhodne připravit si podle uvedených postupů „vlastní OFN“, tak se již nejedná o OFN, ale o tzv. „[datovou specifikaci](#)“. Tyto hrají významnou roli při publikaci veřejných agendových údajů.

Vlastní využití nové OFN pro přípravu datového schématu a datové sady konkrétním poskytovatelem již probíhá shodným způsobem jako v případě „[Přípravy datového schématu na základě OFN](#)“.

#### 4.3.4.7 Výběr formátu distribucí

Pro hierarchicky strukturovaná data jsou k dispozici k výběru datový formát XML a datový formát JSON. Pro formát XML existuje řada standardů pro popis schémat, dotazování se, transformování dat a jejich ukládání. Formát XML se používá zejména ve světě tradičních informačních systémů a při výměně dat mezi nimi. Formát JSON se používá spíše ve světě webových aplikací pro poskytování dat na Webu. Data ve formátu JSON lze standardním způsobem obohatit o mapování na sémantické slovníky. Z tohoto důvodu existující OFN a Sdílené specifikace používají právě formát JSON.

#### 4.3.4.8 Vytvoření strojově čitelných datových schémat

Příprava strojově čitelných datových schémat se řídí technickými standardy pro datové sady na stupni otevřenosti 3, které jsou dostupné na stránkách POD prostřednictvím [odkazu „Technické standardy vytváření datových](#)



[schémat pro datové sady na stupni otevřenosti 3](#)“.

### 4.3.5 Propojená data (grafový datový model)

Podrobnější informace k problematice otevřených dat jsou uvedeny v [kapitole Propojená data](#) a otevřené formální normě [OFN Propojená data](#).

#### 4.2.6.1 Vytvoření RDF modelu

Pro reprezentaci propojených dat se používá datový model RDF – Resource Description Framework [[rdf11-concepts](#)]. Aktuální verze 1.1 byla vydána konsorciem W3C v roce 2014.

Jedná se o grafový datový model, tj. data jsou reprezentována jako uzly a hrany v grafu, kde uzly reprezentují entity a datové hodnoty, a hrany reprezentují jejich propojení.

Takový graf se dá popsat pomocí množiny trojic <uzel 1, hrana, uzel 2>, které říkají, že existuje entita uzel 1, existuje entita uzel 2 a jsou propojeny hranou hrana.

V RDF se jednotlivým částem každé trojice říká **subjekt**, **predikát** a **objekt**, a trojice říká, že objekt je hodnotou vlastnosti predikát nějaké entity subjekt.

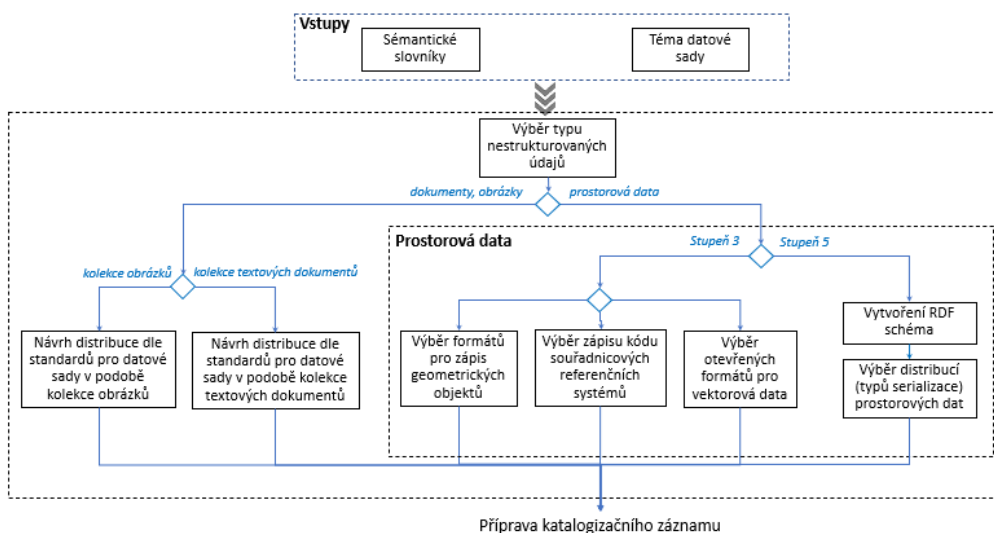
Praktické postupy, doporučení, pravidla a seznam vhodných slovníků pro přípravu RDF modelu obsahuje [OFN Propojená data](#).

#### 4.3.5.1 Serializace RDF modelu

Datový model RDF má 7 standardních serializací. Liší se podporou pro ukládání RDF trojic či RDF čtveřic a využíváním jiných hostitelských datových formátů. Přírozenými trojicovými serializacemi jsou N-Triples a doporučený Turtle, čtveřicovými N-Quads a doporučený TriG. Dále lze RDF trojice serializovat do RDF/XML a RDFa. Pro vývojáře webových aplikací je pak populární serializace JSON-LD, která kombinuje výhody RDF a formátu JSON. JSON-LD je základem Otevřených formálních norem.

Praktická doporučení a bližší informace k jednotlivým způsobům serializace jsou uvedena v [OFN Propojená data](#).

## 4.4 Publikace nestrukturovaných údajů





S – spolupracuje  
R – rozhoduje

otevírání dat otevřených specialisty dat

Výběr typu nestrukturovaných údajů	S	Z	S
Návrh distribuce dle standardů pro datové sady v podobě kolekce obrázků		S	Z
Návrh distribuce dle standardů pro datové sady v podobě kolekce textových dokumentů		S	Z
Výběr formátů pro zápis geometrických objektů		Z	S
Výběr zápisu kódu souřadnicových referenčních systémů		Z	S
Výběr otevřených formátů pro vektorová data		Z	S
<a href="#">Vytvoření RDF modelu</a>		S	Z
Výběr distribucí (typů serializace) prostorových dat		S	Z

#### 4.4.1 Publikace obrázků

Pokud datová sada obsahuje osobní údaje, jako jsou například fotky lidí nebo citlivé informace, musí být tyto údaje řádně zabezpečeny a ochráněny v souladu s platnými právními předpisy.

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.

#### 4.4.2 Publikace textových dokumentů

Pro publikaci nestrukturovaných údajů v podobě dokumentů existují následující možnosti:

- Převod nestrukturovaných údajů na strukturované údaje.
  - Pokud jsou nestrukturované údaje uloženy v textových souborech nebo jiných nestrukturovaných formátech, lze je převést na strukturované formáty, jako jsou např. CSV, XML, nebo JSON. Lze využít řadu nástrojů a technologií, které konverzi usnadní.
- Využití strojového učení a dalších technologií k extrahování strukturovaných dat.
  - Z textových dokumentů lze využitím strojového učení extrahovat strukturovaná data. Existuje řada technologií pro zpracování přirozeného jazyka (Natural Language Processing - NLP), které mohou pomoci při extrahování dat z textových dokumentů.
- Publikace nestrukturovaných údajů jako sémantická data.
  - Sémantické technologie umožňují vytvoření ontologií a mapování nestrukturovaných dat na konkrétní entity. Výsledkem jsou sémantická data, která jsou vhodná pro další zpracování a analýzy. Existuje řada nástrojů a technologií, které mohou pomoci s publikací sémantických dat.
  - Jedná se o stejné postupy jako v případě analýzy agend na základě legislativních dokumentů.
- **Publikace v podobě textových souborů (kolekce textových dokumentů).**
  - Pokud není možné nestrukturované údaje konvertovat na strukturované formáty, lze je publikovat jako textové soubory (například „obyčejný text .txt“, „HTML .html“, „Markdown .md“ nebo formáty pro stránkované dokumenty .docx, .odt apod.).



Distribuce datových sad, které jsou publikované v podobě textových souborů (nejčastější varianta), 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.

### 4.4.3 Publikace prostorových údajů

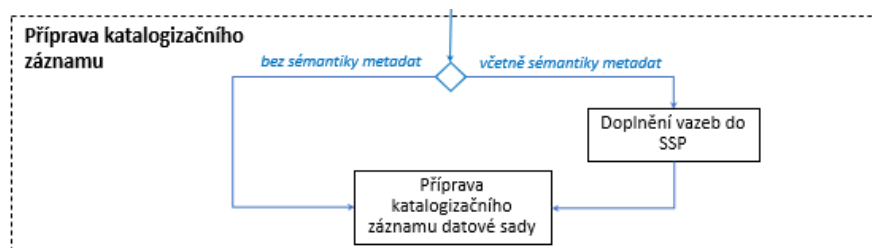
Prostorová data zahrnují všechny typy dat, které jsou prostorově určeny, tedy že přímo nebo nepřímo odkazují na určitou polohu nebo zeměpisnou oblast. Zpravidla jsou určena geometrickým objektem a prostorovými vztahy s okolím (topologií). Obecně lze prostorová data rozdělit na vektorová a rastrová.

Vektorová data by měla být publikována ve vhodném otevřeném formátu za použití standardizovaných geometrických objektů.

Publikace prostorových dat je řešena otevřenou formální normou „[Prostorová data](#)“, která je zaměřena na:

- Výběr formátů pro zápis geometrických objektů.
- Výběr otevřených formátů pro vektorová data.
- Výběr zápisu kódu souřadnicových referenčních systémů.
- Rastrové formáty.
- Propojená prostorová data.
- Výběr distribucí (typů serializace) prostorových dat.
- Způsoby publikace prostorových dat.

### 4.5 Příprava katalogizačního záznamu



Z – zodpovídá

S – spolupracuje

R – rozhoduje

Koordinátor otevírání dat  
Kurátor otevířených dat  
Správce katalogu OD  
IT specialista  
Datový architekt

*Doplnění vazeb do SSP*

*Příprava katalogizačního záznamu datové sady*

	S	S	S	Z
	S	Z	S	

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

Příprava katalogizačního záznamu je usnadněna [průvodcem registrace datové sady](#) v POD, který interaktivním



způsobem provede poskytovatele všemi kroky registrace, a navíc mu v závěru poskytnu informaci o správnosti vytvořeného záznamu.

Struktura záznamu o datové sadě musí odpovídat evropskému standardu [DCAT-AP](#). Ten rozlišuje datovou sadu a distribuci datové sady.

- Datová sada je množina souvisejících dat, vztahujících se k jednomu věcnému tématu. Datová sada je poskytována formou jedné či více distribucí, tj. datových souborů či prostřednictvím definovaného programového rozhraní (API).
- Distribuce představuje datový zdroj v podobě elektronického souboru, který 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.

Záznam o datové sadě se dle uvedeného rozdělení rovněž dělí na 2 typové části:

- První část obsahuje [metadata o datové sadě](#). Jedná se např. o název a popis datové sady.
- Druhá část obsahuje [metadata o distribucích datové sady](#) (datových zdrojích). Jedná se např. o specifikaci datového formátu datového zdroje, jeho datové schéma apod. Každá distribuce má sv vlastní metadata.

Stejným způsobem je i rozdělena samotná příprava metadat katalogizačního záznamu.

- Specifikace metadat datové sady probíhá v této fázi přípravy datové sady a je popsána v této kapitole.
- Doplnění metadat o distribucích datové sady probíhá až při samotné přípravě distribucí a je popsána v kapitole „[Finalizace katalogizačních záznamů](#)“.

Struktura jednotlivých částí záznamu a popis metadat jsou detailně uvedeny v kapitole „[Katalogizační záznam](#)“:

- [Metadata záznamu datové sady](#).
- [Metadata záznamu distribuce datové sady](#).
- [Metadata záznamu datové série](#).

Doplňující informace včetně výkladu jsou k dispozici i v e-learningovém kurzu „[Katalogizace otevřených dat](#)“ na Portálu otevřených dat v sekci vzdělávání.

#### 4.5.1 Doplnění vazeb do SSP (sémantika datové sady na úrovni metadat)

Zajištění sémantiky datové sady na úrovni vazeb metadat do slovníku pojmů veřejné správy ([SSP](#)) představuje:

- identifikaci relevantních pojmů uváděných v metadatech katalogizačního záznamu, pro které je vhodné (potřebné) jejich význam přesně specifikovat pomocí pojmů uvedených ve slovníku pojmů veřejné správy,
- vyhledání identifikovaných pojmů v SSP,
- nalezení a zkopírování jejich odpovídajících identifikátorů IRI,
- doplnění odkazů IRI k příslušným pojmům v metadatech katalogizačního záznamu.

#### 4.5.2 Příprava katalogizačního záznamu datové sady

##### 4.5.2.1 Katalogizační záznam pro katalogizaci do NKOD:

Příprava katalogizačního záznamu pro katalogizaci do NKOD se provádí prostřednictvím interakce s [průvodcem registrace datové sady](#) v POD.

V případě návrhu datového schéma na základě OFN, průvodce umožňuje také využít předpřipravená metadata v samotné OFN, a tyto použít jako základ pro přípravu katalogizačního záznamu.

Postup přípravy katalogizačního záznamu:





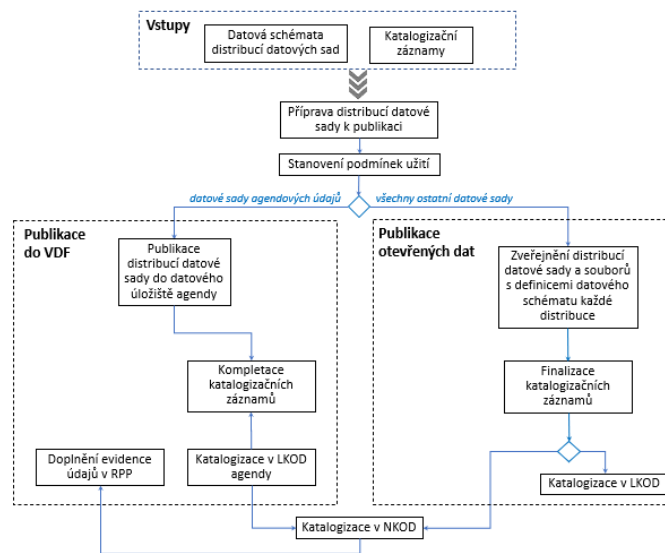
- (Případné stažení předpřipravených metadat z OFN).
- Nahrání předpřipravených metadat do průvodce (pouze v případech publikace dle OFN, nebo využití již hotových katalogizačních záznamů).
- Postupné vyplnění všech zbývajících metadat ([včetně případných souvisejících IRI](#) k zajištění sémantiky) nutných pro katalogizaci datové sady, průvodce tuto činnost usnadňuje nabídkou výběru předpřipravených informací, pokud je to možné.
- Kontrola vytvořeného katalogizačního záznamu průvodcem.
- Stažení rozpracovaného katalogizačního záznamu z průvodce pro jeho doplnění metadaty distribucí až na základě jejich publikace.

Po publikaci distribucí datové sady bude katalogizační záznam doplněn stejným způsobem o metadata jednotlivých distribucí.

#### 4.5.2.2 Katalogizační záznam pro katalogizaci do LKOD

Příprava katalogizačního záznamu pro katalogizaci do lokálního katalogu otevřených dat (LKOD) se řídí postupy vyplývajícími z konkrétní (specifické) implementace lokálního katalogu u konkrétního poskytovatele.

## 5 Publikace a katalogizace datové sady



Z – zodpovídá

S – spolupracuje

R – rozhoduje

Koordinátor otevírání dat  
Kurátor otevřených dat  
Správce katalogu OD  
IT specialista  
Ohlašovatel agendy

*Příprava distribucí datové sady k publikaci*

*Stanovení podmínek užití*

*Zveřejnění distribucí datové sady a souborů s definicemi datového schématu každé distribuce*

*Finalizace katalogizačních záznamů*

*Katalogizace v LKOD*

*Katalogizace v NKOD*

*Publikace do VDF*

S	S		Z	
S	Z			
S	S		Z	
S	S	Z		
S	S	Z	S	
S	S	Z		
S	S	S	Z	Z



## 5.1 Příprava distribucí datové sady k publikaci

### 5.1.1 Příprava distribucí se strukturovanými daty

Při přípravě distribucí datové sady otevřených dat k publikaci je doporučeno postupovat systematicky dle dále uvedených postupných kroků:

- **Identifikace zdrojů** s požadovanými daty, případně sběr dat.
- **Získání dat** pro přípravu distribucí. Možné způsoby:
  - Export dat z informačního systému.
    - Mnoho informačních systémů umožňuje export dat v různých formátech, jako jsou CSV, XML nebo JSON. Tyto soubory mohou být následně zpracovány a převedeny do požadovaného formátu pro přípravu distribuce datové sady.
  - Přímý přístup k databázi.
    - Pokud je k dispozici přístup k databázi informačního systému, lze data získat pomocí dotazovacího jazyka, jako je například SQL. Tento přístup umožňuje získat přesně ta data, která jsou potřebná a lze je také transformovat a integrovat podle potřeb.
  - API rozhraní.
    - Některé informační systémy poskytují API rozhraní, které umožňuje přístup k datům pomocí programování. API rozhraní mohou být použity k získání dat v reálném čase a také k automatizaci procesu získávání dat.
  - Webové stránky.
    - Pokud jsou data publikována na webových stránkách, mohou být získána pomocí webového skrapingu. Tento postup zahrnuje použití speciálních nástrojů nebo skriptů k extrakci dat z webových stránek a převodu do požadovaného formátu.
  - Sběr dat pomocí dotazníků.
    - Pokud data nejsou k dispozici v informačním systému, může být použit dotazník k získání dat od uživatelů nebo subjektů, kteří data vlastní nebo s nimi pracují.
  - Sledování datových toků v případě publikace „[dynamických dat](#)“.
    - Data mohou být získána sledováním datových toků, například pomocí senzorů, měření nebo jiných technologií, které generují data.
  - Formou dohod s externími subjekty (dodavateli veřejné správy).
    - V některých případech mohou být data získána prostřednictvím dohody s externími subjekty, kteří data vlastní nebo s nimi pracují pro potřeby veřejné správy. Tyto subjekty mohou poskytnout data v přímo požadovaném formátu nebo data mohou být převedena do požadovaného formátu během přípravy distribuce datové sady.
- **Posouzení kvality dat** – klíčová aktivita pro publikaci kvalitních a použitelných datových sad.
  - Přesnost – data by měla být ověřena a validována, aby se zaručila přesnost.
  - Úplnost – data by měla být úplná, chybějící data mohou vést k nízké kvalitě data k jejich nepoužitelnosti.
  - Konzistence – data by měla být konzistentní, logická a nesmí být v rozporu s jinými daty.
  - Aktuálnost – data by měla být aktuální a relevantní pro potřeby, pro které jsou data používána.
  - Spolehlivost – data by měla být získána ze spolehlivých a ověřených zdrojů.
  - Formát dat – data by měla být strukturovaná a organizovaná.
- **Úprava dat** (konsolidace, čištění, agregace, anonymizace, ...)



- Čištění dat – data musí být vyčištěna, bez chyb a nesprávných hodnot. To zahrnuje ověření a odstranění chybějících nebo nesprávných hodnot, duplicit a jiných nepotřebných informací.
- Identifikace chybějících hodnot – mohou být například nahrazeny průměrnou hodnotou nebo vymazány.
- Odstranění duplicit – výskyt opakovaných záznamů nebo chyb v procesu sběru dat. Tyto záznamy musí být identifikovány a odstraněny.
- Korekce chybných hodnot – jako jsou nesprávné čísla nebo záznamy s nesprávným formátem.
- Kontrola kvality dat – zaměřena na odhalení případných zbylých (neidentifikovaných) chyb. Kontrola kvality může být provedena pomocí různých metod, jako jsou statistické analýzy, kontrola srovnání s referenčními daty nebo kontroly provedené lidmi.
- Anonymizace – transformace nebo odstranění osobních údajů takovým způsobem, aby nebylo možné identifikovat jednotlivce, ke kterému se údaje vztahují. Anonymizovaná data již nelze zpětně převést na původní osobní údaje, a tedy je nelze spojit s konkrétní osobou.
- Pseudoanonymizace – osobní údaje jsou transformovány takovým způsobem, aby byly identifikovatelné pouze s pomocí dodatečných informací. Pseudoanonymizovaná data lze stále spojit s konkrétní osobou, ale pouze za předpokladu, že jsou k dispozici další informace, které umožní identifikaci.
- **Zajištění vazeb mezi položkami datové sady a hodnotami číselníků.**
  - Pokud jsou v datové sadě položky kódované číselníky, je důležité zajistit, aby byly tyto číselníky poskytnuty společně s datovou sadou. Tím bude zaručeno, že uživatelé budou moci správně interpretovat jednotlivé položky v datové sadě.
  - Vazby mezi položkami datové sady a hodnotami číselníků lze v souboru zaznamenat například pomocí unikátního identifikátoru, který je společný pro oba soubory. Tento identifikátor může být například číselný kód, který je přiřazen k jednotlivým položkám v datové sadě a k odpovídajícím položkám v číselníku.
  - Pro publikaci agendových údajů, jejichž možné hodnoty jsou vymezeny číselníky platí, že tyto číselníky musí být vedeny v RPP ([§ 51 odst. 6, písm. k\) zákona č. 111/2009 Sb.](#)).
  - Strukturu a formát číselníků upravuje OFN „[Číselníky](#)“.
- **Transformace připravených dat do požadovaných formátů a struktur jednotlivých distribucí.**
  - Transformace dat – do formátu distribuce (např. hierarchické stromové struktury), což může zahrnovat úpravu a přebudování datových položek, vytvoření nových vztahů, přejmenování, filtrování, seskupování nebo agregaci dat.
- **Kontrola připravených distribucí.**
  - Porovnání shody s datovým schématem – ověření, zda jsou data v souladu s datovým schématem. Porovnání může být provedeno ručně nebo pomocí nástrojů pro automatizované testování (evaluaci). [Příklad možností uvedených na POD.](#)
  - Validace datové sady – ověření, zda data jsou správná, úplná a platná. Validace může být provedena pomocí automatizovaných nástrojů nebo manuálně prostřednictvím lidského ověření.

## 5.1.2 Příprava distribucí s nestrukturovanými daty

### 5.1.2.1 Distribuce s obrázky

Příprava distribucí otevřených dat, které obsahují obrázky, vyžaduje specifický přístup a formáty souborů, aby byla zaručena srozumitelnost a interoperabilita dat pro uživatele. Při jejich přípravě je nutné dodržovat [Standardy pro datové sady v podobě kolekce obrázků](#) vydaný garantem otevřených dat veřejné správy.



### 5.1.2.2 Distribuce s dokumenty

Příprava distribucí datových sad, které mají charakter textových dokumentů, musí splňovat [Standardy pro datové sady v podobě kolekce textových dokumentů](#) vydané garantem otevřených dat veřejné správy.

### 5.1.3 Příprava distribucí k publikaci do VDF

V případě distribuce datových sad publikovaných agendových údajů do VDF se jedná výhradně o publikaci strukturovaných údajů a jejich příprava je téměř totožná s postupy uvedenými u kapitoly „[Příprava distribucí se strukturovanými daty](#)“.

Rekapitulace a upřesnění kroků uvedených u strukturovaných dat:

- **identifikace zdrojů** – zdroje dat jsou pevně dané, neboť se jedná o agendové informační systémy.
- **získání dat** – v závislosti na provedení AIS připadají v úvahu pouze:
  - export dat z informačního systému,
  - přímý přístup k databázi a využití dotazovacího jazyka (například SQL),
  - využití API rozhraní,
  - dohoda s externími subjekty (dodavateli veřejné správy), případně poskytnutí data již v přímo požadovaném formátu,
- **posouzení kvality dat** – s velkou pravděpodobností kvalita nebude problém, neboť se jedná o aktuální údaje provozovaných agend zodpovědnými OVM,
- **úprava dat** – může se v minimálním počtu případů vyskytnout, ale v zásadě se dá očekávat, že to nebude problém,
- **zajištění vazeb mezi položkami datové sady a hodnotami číselníků** – zcela určitě bude nutné zajistit ve velkém počtu případů, neboť možné hodnoty údajů lze vymezit pouze veřejnými číselníky, a to buď číselníky ČSÚ ([§ 50 odst. 2 zákona č. 111/2009 Sb.](#)), nebo číselníky ohlašovatele agendy ([§ 54 odst. 1, písm. a\) zákona č. 111/2009 Sb.](#)).
- **transformace připravených dat do požadovaných formátů a struktur** jednotlivých distribucí. – prvotní publikace a související nastavení systému (AIS) je shodné, rutinní transformaci přebere AIS spolu s jejími novými komponentami.
- **kontrola připravených distribucí** – v prvotním nastavení shodné, v rutinním režimu přebírá AIS spolu s novými komponentami.

## 5.2 Stanovení podmínek užití

Detailní výklad problematiky včetně postupů pro jednotlivé způsoby ochrany jsou uvedeny na stránkách Portálu otevřených dat ([odkaz POD](#)).

S publikací otevřených dat je spojeno možné porušení autorských a databázových práv, která se k danému obsahu mohou vázat. Možnou překážkou je i situace, kdy jsou součástí distribuce datové sady osobní údaje. Před publikací je nezbytné vypořádat se s těmito překážkami poskytnutím licencí tam, kde je to nezbytné a možné.

Je nutné také uvést, že poskytovatel dat nesmí zavádět omezující podmínky pro další užití dat, jako je například omezení pouze na nekomerční užití. Toto pravidlo vyplývá přímo z definice otevřených dat uvedené v zákoně č. 106/1999 Sb. Pokud poskytovatel dat takový požadavek v podmínkách užití na příjemce dat klade, nejedná se již o otevřená data. Nad rámec toho je pak třeba uvést, že pouhé umístění takového požadavku do metadatového záznamu poskytovaných dat v naprosté většině případů nebude právně vymahatelné.

V případě, že data, ani žádná jejich součást, nejsou jakkoli chráněna, je vhodné uvést tuto informaci, za účelem zvýšení právní jistoty budoucích uživatelů dat.



### Proč jsou podmínky užití důležité?

- Aby uživatelé mohli publikovaná data využívat a mít jistotu, že je nepoužívají nesprávným způsobem.
- Aby publikace dat nevedla k porušení autorských a databázových práv, která se k jejich obsahu mohou vázat.
- Aby byly ošetřeny situace, kdy obsahem distribuce datové sady jsou i osobní údaje.

#### Předmět ochrany:

autorskoprávně chráněné dílo, dle § 2 odst. 1 Autorského zákona (121/2000 Sb.),  
autorskoprávní ochrana databáze jako díla, dle § 2 odst. 2 a 5 Autorského zákona,  
ochrana zvláštních práv pořizovatele databáze, dle § 88 a

Před publikací otevřených dat je nezbytné k nim připojit informaci o tom, zda a jak jsou chráněna z hlediska práv duševního vlastnictví.

#### Doporučené možnosti nastavení podmínek užití.

- **Veřejná licence CC BY 4.0**, kterou vykonavatel 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:
  - Data jsou chráněna jako originální databáze dle § 2 odst. 2 Autorského zákona (tedy, je chráněna jejich struktura) a zároveň nemohou být úředním dílem dle § 3 písm. a) Aut. Z.
  - Data obsahují autorskoprávně chráněná díla a zároveň můžete tato díla dále licencovat. Pokud je licencovat nemůžete, nelze je ani poskytnout jako otevřená data.
- **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 zužitkovat obsah databáze – což je výsledek, který je zamýšlen otevřenými daty dosáhnout. Použije se častěji než licence CC BY 4.0, zejména proto, že se na zvláštní práva pořizovatele databáze vztahuje výjimka úředního díla až v případě databází pořízených po 1. 1. 2017. Použije se tehdy, když:
  - Při tvorbě databáze byl učiněn pořizovatelem databáze (orgánem veřejné správy) pořízením obsahu (zanesením dat do databáze), jeho ověřením nebo předvedením (zveřejněním) kvantitativně nebo kvalitativně podstatný vklad. Ten může být vyjádřen penězi, tak dalšími způsoby. Pokud je databáze úředním dílem, licencovat není potřeba.
- V případě že distribuce datové sady obsahuje osobní údaje, je třeba na tento fakt upozornit a informovat příjemce dat, že se stává správcem osobních údajů a vzhledem k tomu musí plnit zákonné povinnosti, které mu z účinné právní úpravy ochrany osobních údajů vyplývají.
- Pokud poskytovaná data nejsou jakkoli chráněna, je vhodné tuto informaci výslovně sdělit, aby se tak zvýšila právní jistota příjemce dat.

Obecně platí, že licence je potřebná jen tam, kde je poskytován obsah chráněný některým právem duševního vlastnictví. Zároveň může být udělena jen tehdy, když je k tomu poskytovatel otevřených dat oprávněn (např. má možnost udělovat podlicence). V případech, kdy je poskytován obsah, který není nijak zatížen právy duševního vlastnictví, se licence nesmí nikde u dat objevit (“ani jen tak pro jistotu”).



Licence je v českém kontextu specifický typ smlouvy, která se používá právě při udělení oprávnění užít obsah chráněný právy duševního vlastnictví. Její umístění v jiných případech je právně nesmyslné a jen povede ke zmatení příjemců dat.

Pro vytvoření správného katalogizačního záznamu včetně udělení licencí lze využít interaktivní formulář [průvodce registrace datové sady](#), který ve svém [druhém kroku](#) obsahuje předpřipravené doporučené volby licencování nebo upozornění na fakt, že daný obsah poskytovaná sada neobsahuje.

#### Ukázka z formuláře:

Podmínky užití distribuce [?](#)

Autorské dílo  
 Neobsahuje autorská díla [?](#) Jména autorů díla [?](#)

Originální databáze  
 Není autorskoprávně chráněnou databází [?](#) Jména autorů originální databáze [?](#)

Zvláštní právo polizovatele databáze  
 Licencováno CC0 [?](#)

Osobní údaje  
 Neobsahuje osobní údaje [?](#)

#### Datová sada a její distribuce.

Podmínky užití se dle evropského standardu DCAT-AP [připojují k distribuci datové sady](#). Jedna datová sada může obsahovat **více distribucí, kde každá z nich může mít rozdílné podmínky užití**.

#### Přehledová tabulka.

Následující tabulka představuje obecný přehled sumarizující různé varianty možných ochranných režimů a jejich obecně doporučené řešení.

Kategorie	Licenční podmínky	URL v katalogizačním záznamu, pokud distribuce obsahuje data chráněná v této kategorii
<b>Autorské dílo jako součást distribuce datové sady (zcela výjimečné)</b>	Pokud mohu, udělím licenci dle podlicenčního oprávnění	Pokud distribuce obsahuje více děl, licenční záznam musí být přítomný u každého díla. „Dílo <Jméno díla> autora <jméno autora>, které je součástí distribuce <název distribuce datové sady> datové sady <název datové sady> je zveřejněno za podmínek licence <uvést licenci - dle oprávnění>.“ URL v takovém případě bude <a href="#">Distribuce datové sady obsahuje více děl</a> . Pokud je distribuce samostatným dílem, je možné licenci upravit přímo v katalogizačním záznamu. Katalogizační záznam v takovém případě bude Licence CC BY 4.0 ( <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a> ).
<b>Originální databáze</b>	Pokud není úředním dílem (většinou je), CC BY 4.0	„Licence CC BY 4.0 ( <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a> )“
<b>Ochrana zvláštními</b>	Pokud se neaplikuje	„Licence CC0



právy pořizovatele databáze	výjimka úředního díla, CC0	( <a href="https://creativecommons.org/publicdomain/zero/1.0/">https://creativecommons.org/publicdomain/zero/1.0/</a> ) “
Osobní údaje	Pokud distribuce obsahuje osobní údaje, musím to uvést	<a href="#">Distribuce obsahuje osobní údaje</a>

## 5.3 Publikace otevřených dat

### 5.3.1 Zveřejnění distribucí datové sady a souborů s definicemi datového schématu každé distribuce

Před publikací je vhodné provést kontrolu kompletnosti materiálů:

- informace o zvoleném stupni otevřenosti datové sady,
- strojově čitelný soubor s definicí datového schématu datové sady (pokud standardy vyžadují jeho vytvoření),
- strojově čitelný datový soubor každé distribuce s obsahem datové sady
- katalogizační záznam s vyplněnými všemi aktuálně známými atributy.

Možnosti uložení distribucí, datového schématu a případně podmínek užití:

- Zveřejnění na webových stránkách organizace (poskytovatele) nebo vládní instituce, kde jsou snadno dostupná pro uživatele ke stažení.
- Zveřejnění na datových portálech, kde organizace nebo vládní instituce mohou publikovat svá data, aby byla snadno dostupná pro uživatele.
- Distribuce na cloudových úložištích jako je například Google Drive, Dropbox nebo Microsoft OneDrive, kde jsou snadno dostupná pro uživatele ke stažení.
- Distribuce prostřednictvím API (rozhraní pro programování aplikací).

Je také možné zveřejňovat datové sady v různých dalších kanálech, pokud to pomůže zvýšit viditelnost a dostupnost dat pro uživatele. Například pro datové sady katalogizované v NKOD, může být užitečné její distribuce zveřejňovat také i na specializovaných platformách pro otevřená data, na webových stránkách relevantních institucí a organizací, nebo na sociálních sítích.

Vždy je však důležité zajistit, aby všechny zveřejněné verze datových sad byly aktualizovány a obsahovaly stejnou verzi dat. Pokud jsou data aktualizována nebo upravována, musí být tato změna zaznamenána ve všech kanálech, kde jsou data publikována, aby uživatelé měli přístup k nejnovějším informacím. K zajištění se nabízí různé nástroje, které umožňují zveřejňování datových sad v různých kanálech s minimálním úsilím a bez nutnosti ručně aktualizovat každou verzi dat.

Nedostupnost distribucí na uváděných odkazech v katalogizačních záznamech v NKOD bývá bohužel poměrně častým nedostatkem publikace. Povinností poskytovatele je zajistit dostupnost publikovaných distribucí datových sad a platnost uvedených odkazů v záznamu katalogizace v NKOD. Nedostupnost distribucí datových sad je NKOD pravidelně monitorována a výsledky měření jsou publikovány na POD ([Datová kvalita \(nejen\) v](#)



[oblasti otevřených dat](#)).

Rovněž by bylo vhodné informace o publikaci datové sady a místech uložení jejích distribucí systematicky spravovat v [katalogu témat otevřených dat](#).

### 5.3.2 Finalizace katalogizačních záznamů

Katalogizační záznam datové sady připravený dle postupů uvedených v kapitole „[Příprava katalogizačního záznamu](#)“ je nutné doplnit o informace vyplývající až z publikace jednotlivých distribucí.

Jedná se zejména o doplnění chybějících [metadat](#) (v závislosti na formátu distribuce):

- Odkaz na stažení souboru
- Přístupové URL
- Formát souboru ke stažení

Informace o umístění distribucí datové sady a jejich datových schémata (odkazy ke stažení) je nutné doplnit do katalogizačního záznamu v části zaměřené na distribuce.

- Media type souboru ke stažení
- Odkaz na strojově čitelné schéma souboru ke stažení
- Media type kompresního formátu
- Media type balíčkovacího formátu

Doplnění metadat do již připraveného katalogizačního záznamu v kapitole „[Příprava katalogizačního záznamu](#)“ lze provést opět interaktivním způsobem s využitím [průvodce registrace datové sady](#) v POD dle postupu uvedeném v kapitole „[Příprava katalogizačního záznamu datové sady](#)“.

Postup finalizace katalogizačního záznamu:

- Nahrání rozpracovaného katalogizačního záznamu z kapitoly „Příprava katalogizačního záznamu datové sady“ do [průvodce registrace datové sady](#).
- Postupné vyplnění všech metadat jednotlivých distribucí (každá distribuce má svá vlastní metadata).
- Kontrola kompletního katalogizačního záznamu průvodcem.
- Stažení kompletního katalogizačního záznamu pro odeslání do NKOD.
- Odeslání výsledného dokumentu datovou schránkou (například přes spisovou službu Poskytovatele) jako přílohu datové zprávy s předmětem „NKOD“ a adresovat ji na datovou schránku Národní katalog otevřených dat (NKOD) s ID m3hp53v.

### 5.3.3 Katalogizace v LKOD

Pokud poskytovatel provozuje lokální katalog otevřených dat splňující požadavky Otevřené formální normy [Rozhraní katalogů otevřených dat](#) a má katalog zaregistrován v NKOD, správce katalogu otevřených dat postupuje s vytvořením nebo aktualizací záznamu o datové sadě v jeho lokálním katalogu dle zvoleného řešení a dle záznamu o datové sadě od Kurátora dat.

Tento způsob katalogizace šetří agendu spojenou s posláním 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.

Nutnou podmínkou pro využití LKOD ke katalogizaci je ale úplnost všech vyžadovaných metadat pro standardizovanou katalogizaci v NKOD.

## 5.4 Publikace do VDF

Publikace a katalogizace datové sady do VDF je poměrně specifická, neboť velkou roli v ní sehrává celá infrastruktura VDF, zejména nové komponenty agendových informačních systémů (AIS) a náročnější nastavení vazeb všech komponent VDF pro zajištění automatické publikace.

### Základní specifika:

- nutné úpravy stávajících AIS a doplnění potřebných nových komponent,
  - Systém pro zpřístupnění údajů (AIS),
  - Úložiště datových sad (AIS),
  - LKOD (Lokální katalog otevřených dat AIS),
- příprava prvotní publikace bývá náročnější, a to zejména z důvodu velké vzájemné provázanosti jednotlivých komponent VDF a nastavení jejich vzájemné komunikace,
- údaje kódované číselníky je nutné publikovat s vazbami na číselníky,
- číselníky se publikují do „úložiště číselníků a evidence RPP“
- agendové údaje se publikují do úložiště AIS,
- po ukončení publikace a katalogizace publikované datové sady v NKOD je nutné doplnit u každého registrovaného publikovaného údaje v evidenci v RPP odkaz na datovou sadu v NKOD, ve které každý jednotlivý údaj je publikován ve formátu otevřených dat.

Činnosti v této oblasti jsou upraveny „[Metodikou poskytování dat ve veřejném datovém fondu](#)“, a je nutné se touto metodikou řídit.

Více informací k publikaci do VDF je uvedeno v kapitole „[Stručný popis VDF](#)“.

## 5.5 Katalogizace v NKOD

Správce katalogu otevřených dat po zveřejnění datové sady IT specialistou musí zajistit, že 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 záznam o datové sadě v lokálním katalogu (viz [Katalogizace v LKOD](#)), 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 záznamu datové sady.

Správa záznamu datové sady v NKOD je prováděna zasíláním registračních dokumentů ve formě samostatné přílohy datové zprávy s předmětem „NKOD“ do datové schránky s identifikátorem m3hp53v „Národní katalog otevřených dat (NKOD)“. Tyto dokumenty lze získat pomocí [formulářů](#) dostupných v NKOD přes odkaz „Registrovat novou datovou sadu“ v patičce NKOD, či pomocí symbolu editace či mazání na detailu datové sady. Dokumenty jsou ve formátu JSON-LD (s příponou.jsonld.txt kvůli omezení na typy příloh datových zpráv) a řídí se otevřenou formální normou [Rozhraní katalogů otevřených dat: DCAT-AP-CZ](#). Zpracování došlých zpráv na straně NKOD probíhá jednou denně, z pravidla v noci. **Každý dokument je nutné zaslat jako jedinou přílohu samostatné datové zprávy, jinak bude zpráva ignorována.**



## Katalogizace nové datové sady

- V [Národním katalogu otevřených dat](#) kliknout v patičce na „Registrovat novou datovou sadu“.
- V následujících krocích pro datovou sadu a distribuce vyplnit záznam o datové sadě.
- V posledním kroku formuláře je třeba výsledný dokument poslat datovou schránkou (například přes spisovou službu Poskytovatele) jako přílohu datové zprávy s předmětem „NKOD“ a adresovat ji na datovou schránku Národní katalog otevřených dat (NKOD) s ID m3hp53v.

## Úprava záznamu o datové sadě

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

- V [Národním katalogu otevřených dat](#) je nutné nalézt záznam pro upravovanou datovou sadu a v jejím detailu kliknout na symbol editace vedle jejího názvu.
- V následujících krocích pro datovou sadu a distribuce upravit záznam o datové sadě.
- V posledním kroku formuláře je třeba výsledný dokument poslat datovou schránkou (například přes spisovou službu poskytovatele) jako přílohu datové zprávy s předmětem „NKOD“ a adresovat ji na datovou schránku Národní katalog otevřených dat (NKOD) s ID m3hp53v.

## Zrušení záznamu o datové sadě

V případě rušení datové sady je postup následující:

- V [Národním katalogu otevřených dat](#) je nutné nalézt datovou sadu ke zrušení a v jejím detailu kliknout na symbol smazání vedle jejího názvu
- V posledním kroku formuláře je třeba výsledný dokument poslat datovou schránkou (například přes spisovou službu poskytovatele) jako přílohu datové zprávy s předmětem „NKOD“ a adresovat ji na datovou schránku Národní katalog otevřených dat (NKOD) s ID m3hp53v.

# 6 Specifika oblastí veřejné správy

Specifika publikace otevřených dat vyplývají z různých pravomocí a odpovědností státní správy a samosprávy.

- Státní správa:
  - má za úkol zajišťovat plnění úkolů a povinností stanovených zákonem na celostátní úrovni,
  - využití otevřených dat se zaměřuje na zajištění transparentnosti a efektivity procesů a rozhodování na celostátní úrovni,
  - má zpravidla větší objem a rozmanitost dat, která jsou důležitá pro celostátní politiku a řízení,
  - provozuje národní registry a databáze, a tak může být získání a publikování dat náročné na koordinaci a sjednocení formátů dat.
- Samospráva:
  - je odpovědná za řízení a správu určitého regionu nebo obce a má tedy menší rozsah působnosti než státní správa,
  - zaměřuje se na zajištění kvality života občanů na místní úrovni a na plnění místních potřeb a priorit,
  - vyžití otevřených dat se soustředí na zlepšení kvality života občanů v konkrétním regionu a na poskytování lepších místních služeb,
  - může mít menší objem dat, ale tyto data mohou být důležitá pro konkrétní region, obec nebo město,
  - má také obvykle větší pravomoci při rozhodování o publikaci dat a může tak být schopna rychleji reagovat na konkrétní potřeby obyvatel a zveřejňovat potřebná data.



## 6.1 Státní správa a centrální orgány

### Základní oblasti a specifika publikace otevřených dat:

- Publikace registrovaných veřejných údajů do VDF.
- Publikace ostatních (neregistrovaných) údajů ve formátu otevřených dat.
- Publikace High Value Datasets (HVDS).

#### 6.1.1 Publikace registrovaných veřejných údajů do VDF

Sdílení údajů veřejné správy je základním předpokladem digitalizace výkonu veřejné správy a je zajišťováno údaji agendových informačních systémů a informacemi spravovanými centrálním místem „[Registrem práv a povinností](#)“ (RPP). Jeho úkolem je evidovat a spravovat informace o subjektech a objektech práva, jejich údajích (atributech), datových strukturách a souvisejících legislativních ustanoveních. RPP na základě evidovaných informací o údajích následně umožňuje a zprostředkovává [sdílení a přístup k údajům](#) napříč všemi agendami při jejich výkonu, a to prostřednictvím:

- „Propojeného datového fondu“ (PPDF) – zde je uveden [popis způsobu použití informací](#) z RPP,
- „Veřejného datového fondu“ (VDF) na bázi otevřených dat – zde je uveden [popis způsobu použití informací](#) z RPP.

Úplný seznam všech povinně evidovaných informací v RPP vymezuje [§ 51](#) v aktuálním znění „Zákona č.111/2009 Sb. o základních registrech“, platném od 1.2.2022.

Propojení údajů evidovaných v RPP a údajů evidovaných v „Národním katalogu otevřených dat“ (NKOD) vytváří „organizační“ základ fungování VDF.

- RPP obsahuje informace o agendách a jejich údajích.
- NKOD obsahuje informace o datových sadách, ve kterých jsou veřejné údaje publikovány.
- Výsledný komplex informací tvoří „mapu údajů veřejné správy“ nutnou pro zajištění funkčnosti VDF (detailnější informace lze získat v kapitole „4. Komponenty VDF a jejich vzájemné vazby“ v „[Metodice poskytování dat ve veřejném datovém fondu](#)“).

VDF zajišťuje prostřednictvím formátu otevřených dat přístup k:

- registrovaným veřejným údajům (veřejným údajům agend),
- číselníkům veřejné správy.

Publikace veřejných údajů agend a souvisejících číselníků ve formátu otevřených dat vyplývá z legislativy a týká se výhradně organizací státní správy. Zpřístupněné veřejné údaje prostřednictvím VDF jsou automaticky také zpřístupněny veřejnosti jako běžná otevřená data.

### Základní povinnosti státní správy a centrálních orgánů vyžadovaných při publikaci do VDF:

- vést správnou a úplnou evidenci informací o agendách a jejich údajích v RPP – pokrývá „[Metodika definice údajů vedených v agendě](#)“.
- Modelovat agendy a vytvářet odpovídající sémantické slovníky – pokrývá „[Metodika tvorby a údržby sémantického slovníku pojmů veřejné správy](#)“.
- Publikovat agendové údaje a související číselníky do VDF dle požadavků vyplývajících z legislativy:
  - Přípravu datových sad k publikaci pokrývá tato metodika (Metodika identifikace, publikace, katalogizace a využívání otevřených dat pro samosprávu a pro státní správu).
  - Zveřejnění datových sad ve VDF upravuje a specifikuje „[Metodika poskytování dat ve veřejném datovém fondu](#)“.



## 6.1.2 Publikace ostatních údajů ve formátu otevřených dat

Publikace všech ostatních veřejných údajů se provádí dle postupů uvedených v této metodice.

Mezi tyto údaje spadají i ty agendové údaje, které nejsou registrované v RPP. Rozlišení, které údaje agendy mají být dle [§ 51 odst. 6, písm. k\)](#) zákona č. 111/2009 Sb. o základních registrech evidovány v RPP, a které ne, je plně v kompetenci ohlašovatele agendy.

## 6.1.3 Publikace High Value Datasets (HVDS)

Publikace „Datových souborů s vysokou hodnotou“ spadá v převážné míře do povinností organizací státní správy a centrálních orgánů.

### 6.1.3.1 Základní informace o HVDS.

„Směrnice Evropského parlamentu a Rady (EU) [2019/1024](#) ze dne 20. června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru“ v Článku 2 definuje „Datové soubory s vysokou hodnotou“ (High Value Datasets, HVDS) jako (jedná se o přesné citace uvedené ve směrnici):

- Dokumenty, jejichž opakované použití je spojeno s významnými přínosy pro společnost, životní prostředí a hospodářství, zejména kvůli jejich vhodnosti pro tvorbu služeb s přidanou hodnotou, aplikací a nových, kvalitních a důstojných pracovních míst, jakož i počtu potenciálních příjemců služeb a aplikací s přidanou hodnotou založených na těchto datových souborech.
- „Dokumenty“ se rozumí:
  - jakýkoli obsah na jakémkoli nosiči (v listinné či elektronické podobě nebo jako zvuková, vizuální nebo audiovizuální nahrávka) nebo
  - jakákoli část takového obsahu
- V kapitole V., směrnice uvádí, že HVDS musí být:
  - dostupné bezplatně,
  - strojově čitelné,
  - přístupné prostřednictvím rozhraní pro programování aplikací a
  - tam, kde je to relevantní, poskytnuty s možností stažení kompletní datové sady.
- Příloha I. Směrnice (EU) 2019/1024 stanovuje šest tematických kategorií údajů spadajících do HVDS:
  - geoprostorové,
  - pozorování Země a životní prostředí,
  - meteorologické,
  - statistika,
  - společnosti a vlastnictví společností,
  - mobilita.

Prováděcí nařízení komise (EU) [2023/138](#) ze dne 21. prosince 2022, stanovuje seznam konkrétních datových souborů s vysokou hodnotou ve všech tematických kategoriích a uvádí v článku 3 a 4 opatření pro jejich zveřejnění a opakované použití.

Pro komplexnější pochopení problematiky publikace HVDS jsou dále uvedena vybraná opatření ze směrnice.

### Výběr některých opatření:

- Subjekty veřejného sektoru, které mají v držení datové soubory s vysokou hodnotou uvedené v příloze, zajistí, aby datové soubory, které jsou popsány v příloze nebo na které se v ní odkazuje, byly zpřístupněny ve strojově čitelných formátech prostřednictvím rozhraní pro programování aplikací odpovídajících přiměřeným potřebám dalších uživatelů. Pokud je to uvedeno v příloze, jsou datové



soubory k dispozici též s možností stažení kompletní datové sady.

- Podmínky používání rozhraní pro programování aplikací musí být doplněny dokumentací k rozhraní pro programování aplikací ve formátu čitelném lidským okem i ve strojově čitelném formátu, které jsou otevřené a jsou uznávány na unijní nebo na mezinárodní úrovni.
- Subjekty veřejného sektoru uvedené v odstavci 1 určí kontaktní místo pro dotazy a záležitosti týkající se rozhraní pro programování aplikací s cílem zajistit dostupnost a údržbu rozhraní pro programování aplikací a v konečném důsledku bezproblémové a účinné zveřejňování datových souborů s vysokou hodnotou.
- Subjekty veřejného sektoru, které mají v držení datové soubory s vysokou hodnotou uvedené v příloze, zajistí, aby byly tyto datové soubory v popisu pomocí metadat označeny jako datové soubory s vysokou hodnotou.
- Aby se usnadnila dostupnost datových souborů pro opakované použití za delší časová období, vztahují se povinnosti uložené tímto nařízením rovněž na stávající datové soubory s vysokou hodnotou ve strojově čitelném formátu, jež byly vytvořeny před datem použitelnosti tohoto nařízení.
- Do dvou let od vstupu tohoto prováděcího nařízení v platnost předloží členské státy Komisi zprávu o opatřeních realizovaných za účelem provedení tohoto prováděcího nařízení. Informace podle odstavce 3 lze případně poskytnout prostřednictvím odkazů na příslušná metadata.
- Každý členský stát poskytne aktualizovanou verzi této zprávy na žádost Komise, která by měla žádost vznést každé dva roky.

Nařízení Komise (EU) 2023/138 dále stanovuje, že HVDS musí být zveřejňovány v souladu se:

- společnými sémantickými standardy a ontologiemi, které jsou přijímány v celé EU. Tyto standardy a ontologie mají zajistit, aby data byla srozumitelná a interpretovatelná pro různé organizace a systémy.
- společnými jazykovými standardy a používat společné pojmy a kódování. To zahrnuje například standardy pro identifikaci geografických území, standardy pro popis ekonomických a finančních dat nebo standardy pro kódování jazyků a znakových sad.

### 6.1.3.2 Specifika HVDS vůči uvedeným postupům v metodice.

Publikace HVDS se řídí postupy uvedenými v metodice, pouze v několika krocích je nutné důsledně respektovat opatření vyplývající z „Nařízení Komise (EU) 2023/138“.

- **Oblast identifikace HVDS** – identifikovaná témata je nutné ověřit v seznamu konkrétních datových souborů s vysokou hodnotou v jednotlivých tematických kategoriích, který je přílohou „Prováděcího nařízení komise (EU) [2023/138](#).
- **Oblast přípravy datové sady** – zde je kladen důraz na sémantickou a syntaktickou interoperabilitu, ze kterého vyplývá povinnost tvorby konceptuálních (sémantických) modelů primárně na základě slovníků a ontologií EU.
  - Používání EU Vocabularies, což jsou standardizované slovníky pro označování a popisování dat v evropském kontextu. Jsou navrženy tak, aby umožnily interoperabilitu a sémantickou shodu mezi datovými sadami publikovanými v rámci evropských institucí a členských států. Tyto sémantické slovníky zahrnují terminologii a definice v mnoha oblastech, jako jsou například veřejné zakázky, finance, veřejné služby, doprava a další. V současné době jsou k dispozici všechny základní ontologie a slovníky pro oblasti uvedené v prováděcím nařízení EU 2023/138.
  - EU Vocabularies jsou přístupné na internetové stránce <https://op.europa.eu/en/web/eu-vocabularies/home>.
  - Povinné používání EuroVoc při přípravě datové sady (např. pro vyhledávání pojmů v evropských slovnících).
  - Povinné mapování pojmů vytvořených konceptuálních modelů na odpovídající pojmy v SSP.



- **Oblast publikace datových sad:**
  - povinné zpřístupnění HVDS ve strojově čitelných formátech prostřednictvím rozhraní pro programování aplikací,
  - volitelné zpřístupnění HVDS prostřednictvím stažení kompletní datové sady jako datové soubory.
- **Katalogizace datové sady:**
  - povinné používání EuroVoc a vazeb do slovníků (sémantika) v metadatech,
  - označování datových sad v metadatech jako HVDS,
  - důraz na doplňující dokumentaci publikovaných datových sad.

Další upřesňující specifika pro konkrétní datové soubory HVDS v konkrétních tematických kategoriích jsou uvedena v příloze „Prováděcího nařízení komise (EU) [2023/138](#) ze dne 21. prosince 2022“.

## 6.2 Samospráva

Publikace otevřených dat organizacemi samosprávy ve všech dále uvedených oblastech se plně řídí postupy uvedenými v této metodice (Metodika identifikace, publikace, katalogizace a využívání otevřených dat pro samosprávu a pro státní správu).

V kontextu samosprávy se jedná o data, která jsou poskytována kraji, městy, obcemi a dalšími veřejnými institucemi.

### Základní oblasti a specifika publikace otevřených dat:

- Publikace otevřených dat zaměřených na zlepšení kvality života občanů.
- SMART koncepty.
- Publikace otevřených dat vyplývající z legislativy.

### 6.2.1 Publikace otevřených dat zaměřených na zlepšení kvality života občanů

Samospráva se zaměřuje na zajištění kvality života občanů na místní úrovni a na plnění místních potřeb a priorit. Tomu odpovídá publikace i užití otevřených dat, které se soustředí na zlepšení kvality života občanů v konkrétním regionu a na poskytování lepších místních služeb.

Význam a vliv otevřených dat v oblasti samosprávy:

- mohou pomoci zlepšit kvalitu a efektivitu poskytovaných veřejných služeb (například data o využívání městské dopravy mohou pomoci přizpůsobit trasu a frekvenci autobusů a tramvají potřebám obyvatel a tím zlepšit kvalitu městské dopravy),
- pomáhají zlepšit transparentnost a odpovědnost v samosprávě, (obyvatelé získají přístup k informacím o tom, jak se samospráva řídí a jak využívá své zdroje),
- mohou být využita jako zdroj informací pro rozhodování v samosprávě, (například data o využívání sportovních areálů mohou pomoci samosprávě rozhodnout o investicích do sportovní infrastruktury),
- mohou posloužit jako zdroj pro vznik nových aplikací a služeb, které mohou zlepšit život obyvatel v konkrétním regionu,
- mohou být využita pro vytváření nových produktů a služeb, což může podpořit ekonomický rozvoj regionů,
- umožňují lepší plánování a využívání veřejných zdrojů a služeb, což může vést ke zvýšení efektivity a efektivity řízení samospráv.

#### 6.2.1.1 Výhody standardizace

Typickým rysem organizací samosprávy je skutečnost, že všechny řídí a spravují regiony nebo obce, které vykazují



podobné, nebo téměř stejné charakteristiky, a tedy i postupy uplatňované při jejich správě jsou podobné. Proto se nabízí publikovat otevřená data o stejných „věcech“ stejným způsobem, což nejen, že usnadní samotnou publikaci otevřených dat, ale zejména rozšíří možnosti jejich užití.

K tomuto účelu jsou určeny především otevřené formální normy (OFN), které umožní organizacím samosprávy publikovat otevřená data efektivnějším a snadnějším způsobem (viz metodika). Standardizace publikace vede ke standardizaci datového fondu samosprávy a ke standardizovaným aplikacím nad otevřenými daty, následně k možnosti nechat si vyvíjet aplikace společně pro více organizací samosprávy a reprezentovat tímto způsobem větší trh pro vývojáře aplikací a uplatnit tak výhody z rozsahu – levnější aplikace.

Výhody a přínosy standardizace využitím otevřených formálních norem:

- poskytovatelé dat (kraje, ORP, seskupení ORP, obce, ...)
  - větší a zajímavější trh pro vývojáře,
  - uplatnit výhody z rozsahu – levnější aplikace,
  - nabídka větší pestrosti (standardizovaných) aplikací,
  - efektivnější využití financí:
    - neplatit opakovaně za stejné věci,
    - lepší specifikace požadavků na mobilní aplikace,
    - sdílené náklady, ...
  - příležitosti pro pořádání Hackathonů i na úrovni obcí,
- vývojáři aplikací
  - snadnější tvorba aplikací,
  - samospráva – zajímavější trh,
  - větší vývojářská základna, ...
- občan
  - jedna aplikace pro stejné informace v různých lokalitách,
    - turistika,
    - informace o úřadech,
    - telefonní čísla,
    - podpora mobility,
    - podpora a usnadnění turistiky, cestování, ...
  - větší výběr aplikací dle osobních preferencí.

#### 6.2.1.2 Způsoby využití OD:

- Publikovaná otevřená data jsou zdrojem pro nejrůznější aplikace (např. „Moje město“) a webové portály.
- Jednou publikovaná data jsou tak opakovaně používána a prezentována na více „místech“, což přináší celou řadu výhod:
  - Konzistenci, aktuálnost, sdílení, a hlavně důvěryhodnost publikovaných údajů.
  - Dostupnost více komunikačních kanálů směrem k veřejnosti pro předávání informací.
  - Přenechání samotné (často i nákladnější) vlastní prezentace informací třetím stranám – úspora nákladů poskytovatele dat (např. Mapy.cz, Google, weby regionálních společností, ...).
- Spojování publikovaných datových sad samosprávou
  - vytváření nových kontextů a informací propojením místních datových sad, datových sad různých obcí, regionů a datových zdrojů vytvořených v rámci projektů podporovaných z Evropského fondu pro regionální rozvoj. Cílem nových kontextů je podpora:



- turistiky a pořádaných kulturních akcí.
- shromažďování informací o pamětihodnostech,
- zmapování historie.
- Příkladem aplikací využívající tento typ dat mohou být: můj kraj, můj okres, můj region, ...
- Spojování místních datových sad se sadami publikovanými státní správou
  - státní správa publikuje data z celé řady oblastí, např. z oblasti životního prostředí, nakládání s odpady, výsledky sčítání lidu, statistiky ČSÚ, statistiky MPSV, kriminalitu, ...,
  - spojováním s místními otevřenými daty mohou vznikat zajímavé kontexty a aplikace, jako např. „Poznej svoji obec, okres, kraj“, „Místo, kde žiji“, ...
- Vzájemné spojování datových sad publikovaných státní správou (obsahující data samosprávy).
- tvorba nových kontextů napříč celou samosprávou na základě dat publikovaných státní správou. Názorným příkladem jsou aplikace vzniklé na Hackathonu veřejné správy ver. 2.0 ve dnech 14–15/9/2018 v sídle NKÚ. Např:
  - Míra zadluženosti obcí vs. vybrané výdaje.
  - Šobec (=Srovnávač Obcí), pomocník zvidavého komunálního voliče.
- Publikace datových sad zaměřených na infrastrukturu obce/ města, v tomto případě otevřená data zastávají dvě role:
  - jsou přímým zdrojem dat,
  - současně mohou působit jako samotná integrující platforma jednodušších konceptů SMART Cities.

## 6.2.2 SMART koncepty

### 6.2.2.1 Role Otevřených dat v oblasti SMART CITIES

Koncept SC zahrnuje potřebu spolupráce různých zainteresovaných stran (veřejná správa, občané, poskytovatelé služeb, akademický sektor, výrobní sektor, dodavatelé, ...), což vyžaduje jasně definovaný a otevřený přístup k datům, které tyto spolupráce nebo produkty vzniklé ze spolupráce generují. Tento přístup lze definovat jako:

- požadavek na vytváření otevřených řešení a jejich sdílení,
- používání otevřených standardů,
- publikaci všech dat jako otevřených dat pro zajištění jejich všeobecného sdílení a opakovaného použití,
- zajištění vysoké kvality a dostupnosti poskytovaných dat,
- používání sdílené infrastruktury.

Pro zajištění vnitřně datově provázaného (synergického) kontextu celého města, sestávajícího z heterogenních systémů v rozmanitém organizačním prostředí, je požadavek na zajištění datové propojitelnosti (interoperability) klíčový. Datová propojitelnost musí vykazovat následující charakteristiky:

- být otevřená a standardizovaná,
- umožňovat vzájemné sdílení dat,
- být transparentní,
- musí zaručovat kvalitu všech předávaných dat.

Technická rozmanitost datových zdrojů a jejich vlastníků vede k následujícím problémům, které ovlivňují úspěšnost celého konceptu SC.

- Problém 1 - obtížná shoda nad standardy pro danou oblast / doménu s chybějící vizí dalšího využití otevřených dat pro občany i pro samotné tvůrce datových sad.
- Problém 2 - neochota jednotlivých subjektů (ze soukromého i veřejného sektoru) zpřístupňovat a





sdílet data k všeobecnému použití (často i v důsledku obtížné technické realizovatelnosti).

- Problém 3 - nedostatečná kvalita a spolehlivost nejen sensorických dat.
- Problém 4 - neexistence (případně velká složitost nebo „dodavatelská uzavřenost“) jednotného integračního rozhraní na úrovni města pro napojení různorodých zdrojů dat a řešení jednotlivých domén konceptu SC.
- Problém 5 - neefektivní vynakládání finančních prostředků na řešení „opakovatelného konceptu SC“ nebo jeho částí u 14 krajů a 6 258 obcí ČR.

### 6.2.2.2 Zásady/doporučení pro začlenění otevřených dat do konceptu SMART CITIES

Otevřená data slouží jako:

- integrující platforma konceptu SC,
- základ standardizace dat v konceptu SC,
- základ pro výměnu a sdílení dat mezi všemi zainteresovanými stranami,
- základ tvorby otevřených řešení při dalším zhodnocování dat,
- prostředek vedoucí k úsporám při realizaci SC.

Při naplňování konceptu SC je vhodné až životně nezbytné se v oblasti otevřených dat řídit následujícími doporučeními.

#### Oblast integrace dat a tvorby integrační platformy.

- V případě technicky jednoduchých řešení, nenáročných na specifické formáty výměny dat, je vhodné jako základ pro integraci a výměnu dat použít přímo formát otevřených dat.
- Technicky sofistikované systémy (IoT, výsledky měření ze senzorů, ...) se specifickými formáty dat je nutné řešit specifickým, na dané platformě závislým způsobem. Řešení, ale musí také poskytovat všechna získaná i odvozená data ve formátu otevřených dat. Je nutné zajistit, aby data nezůstala uzamčena v systému. Je nutné mít možnost všechna data kdykoliv ze systému získat ve strojově čitelné podobě a v otevřeném formátu a pracovat s nimi libovolným způsobem i mimo tento samotný systém.

#### Oblast standardizace

- Nezbytnou součástí kvalitní a snadné integrace a opakovaného použití dat je standardizace datových formátů a datových struktur všech používaných dat a harmonizace jejich sémantiky. K tomu slouží otevřené formální normy. Je nutné dodržovat existující normy a aktivně se podílet na jejich tvorbě a rozvoji.
- DIA koordinuje rozvoj otevřených formálních norem ve spolupráci s poskytovateli i uživateli dat. Pro každého poskytovatele je důležité se do rozvoje zapojit.
- V rámci všech zainteresovaných subjektů do konceptu SC je také nutné dosáhnout shody nad dalšími procesními a legislativními pravidly otevřených dat. Jejich aktuální podoba je zveřejněna v podobě Standardů otevřených dat (garantuje DIA).

#### Oblast komunikace

- Pro naplnění základního cíle SC je nutné zajistit, aby veškerá data byla přístupná komukoliv ve formátu otevřených dat. Data nejsou využívána pouze pro konkrétní řešení v rámci konceptu SC, ale jsou také zveřejněna jako otevřená data pro využití v dalších různých službách (mobilní aplikace, webové služby, novinové analytické články, studentské práce atd.) a také proto, aby nové služby mohly vznikat.
- Základní otázka při zveřejňování dat zní "Existuje nějaký důvod data nezveřejňovat?", nikoliv otázka "Existuje nějaký důvod data zveřejňovat?". Druhá otázka má totiž vždy odpověď "ANO". Jestliže data sbíráme, mají pro nás nějaký význam a smysl. A určitě nejsme jediní, kteří v datech tento smysl vidí.

#### Oblast vytváření otevřených řešení



- Doporučení uvedená v částech „Oblast standardizace“ a „Oblast komunikace“ přímo podporují vznik nových otevřených řešení i mimo plánovaný koncept SC. Pro tvůrce aplikací to znamená snadnější tvorbu aplikací a zajímavý trh (6 258 obcí a 14 krajů).
- Pro občany se nabízí větší a bohatší výběr aplikací (dle osobních preferencí) doplňujících záměr města při budování konceptu SC.

### **Oblast financí a hledání úspor při budování konceptu SC**

Doporučení uvedená v částech „Oblast standardizace“ a „Oblast komunikace“ mají přímý dopad do efektivity budování konceptu SC. Sjednocení a standardizace problematiky SC vede k tomu, že:

- témata měst, obcí, regionů a jejich správy se stanou větším a zajímavějším trhem pro vývojáře,
- města, obce a regiony budou moci uplatnit výhody z rozsahu a získat levnější aplikace a řešení,
- bude možné nabídnout občanům (rezidentům, pracovníkům, turistům, ...) větší pestrost aplikací,
- nebude nutné platit opakovaně za stejné produkty a tím efektivněji využívat finance,
- bude možné lépe specifikovat společné požadavky na aplikace a řešení,
- bude možné sdílet náklady.

### **6.2.3 Publikace otevřených dat vyplývajících z legislativy**

Publikace otevřených dat samosprávou na základě povinností vyplývajících přímo z legislativy ČR jsou spíše výjimkou. Zatím se jednalo pouze o publikaci dat z úředních desek na základě [§ 5a odst. 4 zákona č. 106/1999 Sb.](#) o svobodném přístupu k informacím.

V případě další povinné publikace otevřených dat vyplývajících z legislativy, která se bude týkat i organizací samosprávy, bude vždy s velkou pravděpodobností pro takové případy připravena odpovídající otevřená formální norma (OFN), stejně, jako to proběhlo v případě publikace dat z úředních desek.



# Část 2. Komunikace a propagace

## 7 Presentace a propagace dat

### 7.1.1 Komunikační strategie

Komunikaci o otevřených datech z pozice garanta otevřených dat upřesňuje vypracovaná „[Komunikační strategie pro cílové skupiny otevřených dat](#)“, která je dostupná na stránkách POD včetně její [přílohy](#).

Strategie obsahuje analýzu prostředí otevřených dat v ČR, podrobný popis jednotlivých cílových skupin, přehled doporučené komunikace s cílovými skupinami otevřených dat, formáty sdělení, doporučené komunikační kanály a četnosti komunikace.

#### 7.1.1.1 Doplnění komunikační strategie pro komunikaci poskytovatelů otevřených dat.

Základní obecné rady a doporučení:

- Jednoduchý jazyk, který je srozumitelný pro cílovou skupinu.
- Vizualizace pro přehlednější a srozumitelnější předávání informací.
- Příběh (relevantní pro cílovou skupinu) o tom, jak otevřená data mohou pomoci lidem a společností.
- Ukázky, jak lze otevřená data využít v praxi. Vhodné příklady úspěšných aplikací a projektů, včetně sdílení dobré praxe mezi jednotlivými subjekty veřejné správy.
- Podpora a spolupráce s ostatními organizacemi a lidmi, kteří se zajímají o otevřená data. Vytváření fór pro diskusi a spolupráci.
- Dostupnost a srozumitelnost otevřených dat pro širokou veřejnost. Vhodné (datové) portály, webové stránky, aplikace a jiné nástroje k prezentaci dat.
- Transparentní komunikace o tom, jaká data jsou k dispozici a jaká jsou omezení, včetně uvedení zdroje dat a způsobu, jakými byla data získána.

### 7.1.2 Propagace otevřených dat

Propagace otevřených dat se zaměřuje na šíření informací o tom, co jsou otevřená data, jaké jsou jejich výhody a jak se k nim mohou různé skupiny lidí připojit a využít je.

Propagace může být prováděna prostřednictvím různých kanálů, jako jsou webové stránky, sociální sítě, konference, oceněním úspěšných projektů odbornou veřejností a podobně. Cílem propagace je zvýšit povědomí o otevřených datech a podpořit proces otevírání dat subjekty veřejné správy a jejich využití pro širokou veřejnost.

#### Doporučení a aktivity pro cílové skupiny.

- Propagace pro vývojáře by měla být orientovaná na praktické využití dat a měla by poskytovat vývojářům užitečné nástroje a informace, které potřebují k vytváření nových aplikací a služeb.
  - Vytvoření webových stránek, které budou zaměřeny na vývojáře a budou obsahovat informace o dostupných otevřených datech, API, SDK, nástrojích pro zpracování dat a dalších užitečných informací pro vývojáře.
  - Pořádání prezentací a workshopů pro vývojáře s poskytnutím praktických návodů a ukázek.
  - Pořádání hackathonů a soutěží, které budou zaměřené na využití otevřených dat.
  - Aktivně se zúčastňovat akcí pořádaných jinými subjekty s ukázkami způsobů využití dat.
  - Poskytnutí detailní dokumentace API pro otevřená data.
- Propagace otevřených dat pro novináře.



- Příběhy, které ukazují, jak otevřená data mohou pomoci s řešením společenských problémů nebo poskytnutím nového pohledu na nějakou událost.
- Webináře, ve kterých je vysvětleno, jak jim mohou otevřená data pomoci. Ukázky, jak je najít, stáhnout, analyzovat, aby bylo možné vytvářet nové příběhy.
- Nabídka připravených zpravodajských balíčků, které obsahují relevantní otevřená data pro určité události nebo témata.
- Prezentace případových studií, ve kterých je ukázáno, jak novináři již využili otevřená data k vytváření nových reportáží.
- Propagace otevřených dat pro občany.
  - Příběhy o tom, jak otevřená data mohou pomoci řešit problémy ve společnosti nebo pomoci lidem získat nový pohled na určité témata. Tyto příběhy mohou být zveřejněny na sociálních sítích, veřejných akcích nebo v místních novinách.
  - Pořádání veřejných akcí a seminářů, které ukazují, jak se otevřená data používají a jaký mají vliv na společnost. Akce mohou být zpřístupněny na místní úrovni a mohou zahrnovat prezentace a workshopy.
  - Upozornění občanů na otevřená data, která mohou používat pro své vlastní účely. Mnoho měst, krajů a obcí má své otevřené datové portály, které umožňují občanům snadný přístup k datům o svém městě, obci, nebo regionu. Tyto portály mohou obsahovat informace o místním dopravním provozu, úrovni kriminality, zdravotní péči a dalším.
  - Podpora projektů, které využívají otevřená data. Tyto projekty mohou být zaměřeny na řešení konkrétních problémů v oblasti dopravy, životního prostředí, dostupnosti škol a dalších.
  - Spolupráce s místními skupinami a organizacemi, které se zajímají o otevřená data.
- Propagace otevřených dat pro školy.
  - Podpora přípravy vzdělávacích materiálů pro učitele a studenty, které ukazují, jak se otevřená data používají a jaký mají vliv na společnost.
  - Nabídka konkrétních otevřených dat jako zdroj informací pro školní projekty a prezentace.
  - Vytvoření interaktivních her, které pomáhají studentům naučit se používat otevřená data.
  - Pořádání soutěží, které povzbuzují studenty k využívání otevřených dat.
- Propagace otevřených dat pro veřejné společnosti a instituce.
  - Pořádání vzdělávacích akcí, seminářů a workshopů pro zaměstnance veřejných společností, aby se naučili, jak využívat otevřená data pro zlepšení své práce. Tyto akce mohou být zaměřené na konkrétní oblasti, jako je plánování městské infrastruktury, ochrana životního prostředí nebo sociální politiky.
  - Vytváření platforem pro otevřená data, jako jsou webové stránky nebo portály, které umožňují veřejnosti snadný přístup k otevřeným datům.
  - Podpora využití otevřených dat ke zlepšení služeb pro občany. Například lze vytvořit mobilní aplikaci, která umožní občanům snadno najít informace o městské dopravě, stavu silnic, znečištění ovzduší a dalších důležitých faktorech.
- Propagace otevřených dat pro samosprávu.
  - Vytvoření webové stránky s informacemi o otevřených datech, kde by byly uvedeny příklady, jak mohou být data využita, a jakým způsobem jsou dostupná.
  - Systematické zveřejňování dat, která jsou relevantní pro občany a mohou pomoci s efektivním řízením města. Například informace o územním plánování, zdravotnictví, vzdělávání nebo dopravě.
  - Organizování vzdělávacích akcí pro pracovníky samosprávy, aby se naučili, jak mohou otevřená data využít ke zlepšení své práce. Tyto akce by mohly být zaměřené na konkrétní oblasti, jako je



urbanismus, ekologie nebo sociální politika.

- Spolupráce s místními organizacemi a občanskými sdruženími, aby se zvýšila povědomost o výhodách otevřených dat. Tyto organizace by mohly být zapojeny do tvorby aplikací a služeb využívajících otevřená data pro občany.

### 7.1.3 Prezentace otevřených dat

Prezentace otevřených dat se zaměřuje na ukázání, jak jsou data dostupná a jak mohou být použita. Prezentace může být prováděna prostřednictvím různých nástrojů, jako jsou vizualizace, tabulky, grafy, mapy a podobně. Cílem prezentace je umožnit lidem prozkoumat otevřená data a vidět, jak mohou být data použita v praxi.

#### Možné způsoby prezentace otevřených dat.

- Vizualizace dat může být velmi efektivním způsobem, jak představit velké množství dat v srozumitelné a atraktivní formě. Lze použít nástroje, jako jsou grafy, tabulky, mapy, infografiky a další, vhodné pro představení dat v přístupné formě veřejnosti.
- Vytvoření demo aplikace, která využívá otevřená data, je velice působivým způsobem prezentace dat. Aplikace může ukázat, jaká data jsou dostupná, jak je lze použít a jaké jsou jejich výhody.
- Prezentace na konferencích je dalším způsobem, jak prezentovat otevřená data.
- Psaní blogů a článků o otevřených datech. Lze psát o tom, co jsou otevřená data, jaká jsou jejich využití a jak se k nim mohou lidé připojit.
- Vytvoření webových stránek věnovaných otevřeným datům. Stránky mohou obsahovat informace o dostupných datech, návody, jak data používat a ukázky aplikací, které využívají otevřená data.
- Organizování workshopů a školení na téma otevřených dat. Lze vytvořit workshop, kde lidé mohou sami pracovat s daty a naučit se je používat.

#### Prezentace otevřených dat pro veřejnou správu a poskytovatele.

- Vysvětlení, co jsou otevřená data a jak mohou být využita pro zlepšení kvality veřejných služeb a podporu rozhodování.
- Představení dostupných datových sad a ukázka, jak lze k nim přistupovat. To může zahrnovat popis datových zdrojů, webových stránek a případně i dalších zdrojů.
- Ukázat, jak mohou být data použita prostřednictvím vhodných aplikací pro lepší porozumění potřebám obyvatel, pro vylepšení plánování veřejných služeb a zvyšování efektivity procesů.
- Představení nástrojů a technologií, které mohou být použity pro práci s daty. To může zahrnovat různé programovací jazyky, nástroje pro vizualizaci dat a další.
- Ukázka příkladů projektů a aplikací, které využívají otevřená data a které mohou být použity pro zlepšení veřejných služeb a procesů.
- Poskytnutí praktických tipů, návodů a příkladů dobré praxe, jak pracovat s daty a jak je využít pro potřeby veřejné správy a poskytovatelů služeb. To může zahrnovat rady pro získání a čištění dat, používání nástrojů pro vizualizaci dat a další.
- Organizace workshopů a tréninků, kde zaměstnanci veřejné správy a poskytovatelé si mohou prakticky vyzkoušet práci s daty a získat zkušenosti s jejich použitím.

## 8 Využití otevřených dat

Využitím otevřených dat dochází k jejich zhodnocování, a to v oblasti **sociální**, nebo **ekonomické**. Otevřená data vystupují při jejich zhodnocování ve dvou rolích:

- jsou zdrojem pro tvorbu ekonomických hodnot,
- stávají se nástrojem pro utváření společenských hodnot.



**Zhodnocovateli otevřených dat jsou jak veřejná správa tak i soukromý sektor.**

## 8.1 Oblasti možného využití otevřených dat.

- Zlepšení transparentnosti a odpovědnosti vládních institucí a podniků tím poskytováním přístupu veřejnosti k důležitým informacím o jejich činnosti.
- Zlepšení efektivity a produktivity v různých oblastech, jako je například veřejná správa, průmysl nebo zemědělství.
- Zlepšení komunikace a informovanosti mezi lidmi a institucemi poskytováním přesných a aktuálních informací o různých tématech a událostech.
- Zlepšení spolupráce mezi lidmi a institucemi a umožnění sdílení informací, dat a znalostí.
- Vylepšení rozhodovacích procesů v různých oblastech, jako je například zdravotnictví, doprava, životní prostředí nebo bezpečnost.
- Podpora sociální spravedlnosti a zlepšení života lidí v různých komunitách poskytováním informací o nerovnostech a diskriminaci.
- Provádění výzkumu a analýz v různých oblastech, jako jsou sociální vědy, zdravotnictví, vzdělání, doprava, finance a mnoho dalších.
- Podpora environmentální udržitelnosti a zlepšení ochrany přírodních zdrojů poskytováním informací o emisích skleníkových plynů, kvalitě ovzduší, nebo využívání energie.
- Zlepšení vzdělávání a výuky poskytováním studentům a vyučujícím přístup k různým datům a informacím pro výzkum a analýzu.
- Monitorování a predikce trendů v různých oblastech, jako je například ekonomika, doprava, zdravotnictví nebo klimatické změny.
- Podpora inovace v různých oblastech a pomoc při vytváření nových produktů, služeb a řešení.
- Rozvoj aplikací, které mohou poskytovat lepší služby a zlepšit interakce mezi lidmi a institucemi.
- Vytváření vizualizací, které mohou pomoci lidem lépe porozumět datům a informacím.
- Vytváření nových obchodních modelů a vytváření nových produktů a služeb.
- Zlepšení výkonu podniků poskytováním informací o trhu a spotřebitelských preferencích.
- Zlepšení kvality a bezpečnosti výrobků a služeb poskytováním informací o zdravotních rizicích, environmentálních dopadech, nebo dalších bezpečnostních problémech.

## 8.2 Způsoby zhodnocování otevřených dat

Příklady aplikací, portálů a zpravodajských webů využívajících otevřená data je možné nalézt ve [výročních zprávách o stavu otevřených dat v České republice](#) publikovaných na Portálu otevřených dat.

V plánu budoucího rozvoje NKOD je rozšíření jeho funkcionality o správu vytvořených a zaregistrovaných produktů vytvořených s využitím otevřených dat.

- Analýza dat – základní způsob zhodnocení otevřených dat. Analýzou lze identifikovat vzorce a tendence v datech, které mohou být použity k získání nových poznatků a informací. Typy analýz:
  - Identifikace trendů a vzorů.
  - Získání znalostí.
  - Predikce.
  - Odhalení vztahů mezi různými faktory.



- Vizualizace dat – zobrazení dat v grafické podobě, což umožňuje lépe porozumět datům a identifikovat vzorce a tendence.
- Dashboardy – interaktivní vizualizace dat, které umožňují uživatelům data vizualizovat a prezentovat v reálném čase pomocí interaktivních grafů a tabulek.
- Webové mapy – v případě prostorových dat lze využít webových mapových nástrojů, které umožňují vizualizaci geografických dat a umožňují interakci s těmito daty.
- Infografiky – vizuálně atraktivní prezentace dat, které lze snadno sdílet, a kterým je snadné porozumět.

### 8.3 Zhodnocování dat ve veřejné správě

Seznam aplikací, portálů a zpravodajských webů využívajících otevřená data je možné nalézt ve výročních zprávách o stavu otevřených dat v České republice publikovaných na Portálu otevřených dat.

- Monitorování výkonu – otevřená data mohou být použita pro sledování výkonu různých státních institucí a jejich porovnávání s výkony v jiných regionech nebo srovnatelných organizacích. Tyto údaje mohou pomoci identifikovat oblasti, které by mohly být zlepšeny.
- Plánování a rozhodování – otevřená data mohou být také použita pro plánování budoucích akcí a rozhodování. Například lze využít údaje o demografii a geografii pro posouzení potřeb obyvatelstva v určité oblasti a rozhodnutí o tom, jaké služby poskytnout, nebo zajistit.
- Zvyšování transparentnosti – otevřená data mohou být použita pro zvýšení transparentnosti veřejné správy. Například otevřené údaje o rozpočtu a výdajích mohou pomoci občanům porozumět tomu, kam se jejich daně dostávají a jak s nimi subjekty veřejné správy hospodaří.
- Identifikace nezbytnosti změn – otevřená data mohou pomoci identifikovat oblasti, ve kterých jsou nutné změny nebo vylepšení. Například mohou ukázat, kde jsou největší potřeby a jaké služby jsou nejvíce požadovány v daném území nebo regionu.
- Zlepšení služeb pro občany – otevřená data mohou být použita k vylepšení služeb poskytovaných občanům. Například lze využít údaje o provozu ve městě pro lepší plánování dopravních tras a zlepšení mobility.
- Podpora výzkumu – otevřená data mohou být použita k podpoře výzkumu a vývoje. Například lze využít údaje o zdravotním stavu populace pro vývoj nových zdravotnických technologií a léčebných postupů.

### 8.4 Postup vizualizace dat

Zpracování a vizualizaci dat se věnuje záznam školení „[Zpracování a vizualizace otevřených dat](#)“, které je zaměřeno na seznámení se způsoby čištění dat, jejich zpracováním, vyhledáváním a způsoby transformace do konkrétních výstupů.

Zpracování a vizualizaci dat se věnuje e-learningové školení „Zpracování a vizualizace otevřených dat“, které je zaměřeno na seznámení se způsoby čištění dat, jejich zpracováním, vyhledáváním a způsoby transformace do konkrétních výstupů.

- Úvodní podmínky:
  - Nutnost existence konkrétní myšlenky nebo stanoveného účelu pro vizualizaci otevřených dat.
    - Ideální je mít jasně definované cíle a otázky, na které se hledají odpovědi, nebo vzniká



- potřeba vyřešit určitý problém.
- Dobrou metodou je příprava příběhu (Storytelling), který bude „odvyprávěn“ pomocí otevřených dat.
- Bez konkrétního účelu bývá obtížné určit, která data jsou pro daný projekt relevantní, jakým způsobem je zpracovat a jakým způsobem je prezentovat.
  - Představa správné interpretace dat a výběr vhodného nástroje pro vizualizaci.
- Získání přístupu k datům – nezastupitelné místo k nalezení požadovaných dat sehrává NKOD, ale existuje i mnoho dalších zdrojů vhodných dat, jako jsou webové stránky a datové portály, které poskytují otevřená data ve formě datových souborů, API nebo vizualizací.
- Často je vhodné vizualizaci otevřených dat doplnit i zdroji dat, které nejsou publikovány formou otevřených dat.
- Výběr vhodných nástrojů pro vizualizaci na základě potřeb a úrovně znalostí zpracovatele. Mezi nejpoblárnější nástroje například patří:
  - Tableau, Power BI, Google Data Studio, R Studio,
  - webové mapové nástroje: Google Maps nebo OpenStreetMap,
  - pro tvorbu infografik: Canva nebo Piktochart,
  - pro tvorbu Dashboardů: Google Data Studio.
- Zpracování dat – před vizualizací může být nutné data nejprve připravit. To může zahrnovat například úpravu formátu dat, čištění dat, sjednocení dat a další úpravy. Zpracování dat může být složitý proces, který ale může výrazně zlepšit kvalitu výsledné vizualizace.
- Návrh vizualizace – rozhodnutí o způsobu nevhodnější vizualizace dle zamýšleného účelu a obsahu předávané zprávy.
- Vlastní vytvoření vizualizace – předpokladem jsou technické způsobilosti a um zpracovatele v závislosti na použitých nástrojích.

## 8.5 Příprava infografiky a související klíčové faktory

- Kvalita dat – bez kvalitních a relevantních dat by infografika neměla žádnou hodnotu. Proto je důležité získat data z důvěryhodných zdrojů s kvalitní dokumentací a ověřit jejich správnost.
- Kreativní návrh – infografika musí být vizuálně atraktivní a přehledná. Měla by být navržena tak, aby okamžitě upoutala pozornost diváka a byla snadno čitelná. Využití barev, kontrastu, typografie a grafických prvků může pomoci přitáhnout pozornost k důležitým informacím.
- Jasná a jednoduchá prezentace dat – infografika by měla být navržena tak, aby zprostředkovala důležité informace co nejjasněji a nejúčinněji. Je důležité použít jednoduchý jazyk a grafické prvky k vysvětlení složitých konceptů.
- Využití vhodného formátu – infografiky lze vytvořit v mnoha různých formátech, jako jsou statické obrázky, interaktivní grafy a tabulky, animované infografiky apod. Volba vhodného formátu pro prezentaci dat je velmi důležitá.
- Identifikace cílové skupiny – různé cílové skupiny mohou mít různé preference vzhledu, stylu a formátu.
- Testování a optimalizace – infografiku je důležité otestovat a optimalizovat, aby byla co nejefektivnější. Vhodným způsobem je požádat několik lidí o její ohodnocení z hlediska čitelnosti, přehlednosti a srozumitelnosti.

## 8.6 Příprava a vytvoření příběhu (Storytelling), jako základu analýzy a vizualizace otevřených dat.

Příběh je velmi účinným nástrojem pro analýzu otevřených dat. Při vytváření příběhu je vhodné se zaměřit na





určitou problematiku nebo téma, které je relevantní pro analyzovaná data.

Storytelling může pomoci zvýšit povědomí o určitém problému nebo tématu, zlepšit porozumění datům a motivovat lidi k akci.

Příkladem může být například využití otevřených dat k vytvoření příběhu o změnách klimatu v určitém regionu. Pomocí vizualizací a analýzy dat lze ukázat, jak se teploty v regionu mění, jak to ovlivňuje přírodu a životní podmínky lidí v konkrétní oblasti a jaké jsou dopady na životní prostředí. Takové údaje lze prezentovat pomocí interaktivních map a grafů, které ukazují změny v čase a prostoru.

#### **Postup vytvoření příběhu:**

- Proces vytváření příběhu začíná zamyšlením nad otázkou, kterou je možné zodpovědět pomocí dat.
- Po formulaci otázky je možné začít hledat data, která pomohou na otázku odpovědět.
- Po nalezení dat, lze začít vytvářet příběh, který využije data k vysvětlení situace a zodpovězení otázky. Příběh lze prezentovat v různých formách, jako jsou infografiky, vizualizace dat nebo i textový příběh.
- Výhodou příběhů jako nástroje pro analýzu otevřených dat je, že umožňují prezentovat data a informace v poutavější a srozumitelnější formě pro širší publikum. Může to být velmi užitečné při prezentování výsledků analýzy dat organizacím nebo veřejnosti.

## **8.7 Práce s informacemi v NKOD**

Vyhledání vhodných zdrojů otevřených dat je podstatné pro zhodnocování a využití publikovaných dat. K nalezení vhodných datových sad publikovaných veřejnou správou ČR je určen Národní katalog otevřených dat. Možnosti a způsoby práce s informacemi v NKOD vysvětluje e-learningový kurz „Katalogizace otevřených dat“ ve 3. modulu [„Katalogizace OD a Národní katalog otevřených dat“](#).



# Část 3. Znalostní

## 9 Specifické/znalostní oblasti a související skupiny činností

### 9.1 Rekapitulace informací o otevřených datech

#### 9.1.1 Základní charakteristiky otevřených dat.

Otevřená data jsou data zveřejněná na internetu, která jsou:

- úplná,
- snadno dostupná,
- strojově čitelná,
- používající standardy s volně dostupnou specifikací,
- zpřístupněna za jasně definovaných podmínek užití dat s minimem omezení,
- dostupná uživatelům při vynaložení minima možných nákladů a
- jsou evidovány v Národním katalogu otevřených dat.

#### Požadavky na podmínky užití otevřených dat.

- Neomezují jejich uživatele ve způsobu použití dat.
- Opravňují uživatele k jejich dalšímu šíření.
- Při dalším šíření musí mít všichni uživatelé stejná oprávnění s daty nakládat.
- Cílem je umožnit šíření dat pro nekomerční i komerční účely.

#### Definice otevřených dat.

Definice dle zákona 106/1999 Sb. (§3), o svobodném přístupu k informacím:

*„Otevřenými daty se rozumí informace zveřejňované způsobem umožňujícím dálkový přístup v otevřeném a strojově čitelném formátu, jejichž způsob ani účel následného využití není omezen a které jsou evidovány v národním katalogu otevřených dat.“*

Definice uvedená v „Mezinárodní chartě otevřených dat“:

*„Otevřená data jsou digitální data, která jsou k dispozici s technickými a právními charakteristikami nezbytnými k tomu, aby byla volně používána, opětovně používána a distribuována kýmkoli, kdykoli a kdekoli.“*

#### 9.1.2 Zastřešující koncept / filozofie otevřených dat – otevřenost.

##### Právní otevřenost

Zaměřena na:

- legalizaci získávání dat, ustavení možností jejich sdílení, šíření a možností jejich následného využití,
- odstranění legislativních bariér zpřístupnění informací a ošetření těch informací, které souvisí s legislativní ochranou.

##### Technická otevřenost.

Zaměřena na:

- odstranění technických překážek spojených s publikací a používáním otevřených dat, což vyžaduje:



- standardizaci samotné publikace dat,
- standardizaci používaných otevřených formátů,
- standardizaci doprovodné dokumentace,
- zajištění kvality, dostupnosti a způsobu dohledání publikovaných dat.

Zákon č.106/1999 Sb. o svobodném přístupu k informacím v [§ 3a odst. 3](#) řešení této oblasti usnadňuje definicí pojmu „Otevřená formální norma“:

*„Otevřenou formální normou se ... rozumí pravidlo, které bylo vydáno písemně a obsahuje specifikace požadavků na zajištění schopnosti různých programových vybavení vzájemně si poskytovat služby a efektivně spolupracovat.“,*

Současně v [§ 4b zákona č. 106/1999 Sb.](#) uvádí, že jsou pro poskytovatele otevřených dat závazné.

Prakticky se jedná o legislativní nástroj pro stanovení technických doporučení (standardů) s cílem zajištění, že stejná data publikovaná různými poskytovateli budou technicky otevřená a interoperabilní.

### 9.1.3 Interoperabilita otevřených dat.

Interoperabilita je základním předpokladem elektronické komunikace a výměny informací mezi orgány veřejné správy, mezirezortně a přeshraničně. Je tedy i základním předpokladem pro dosažení jednotného digitálního trhu. Programy interoperability v EU se v průběhu času vyvíjejí. Zpočátku se týkaly dosažení interoperability v určitých prioritních oblastech, dnes se již jedná o zavádění celé společné infrastruktury.

Interoperabilita u otevřených dat zajišťuje, aby byla data snadno použitelná a sdílená napříč různými systémy a platformami. K tomu dochází prostřednictvím standardů a specifikací, které popisují formát a strukturu dat.

Existuje mnoho standardů a specifikací pro otevřená data, jako například formáty datových souborů, protokoly pro přenos dat, ontologie pro popis datových sad a další. Tyto standardy a specifikace umožňují různým systémům a aplikacím pracovat s daty v konzistentním formátu a předcházet tak problémům s kompatibilitou.

Kromě toho jsou otevřená data publikována s použitím otevřených licencí, které umožňují další šíření, využití a modifikaci dat. To také podporuje interoperabilitu, protože umožňuje různým uživatelům a systémům pracovat s daty bez omezení.

European Interoperability Framework (EIF) stanovuje požadavky na zajištění interoperability na několika úrovních

- Právní interoperabilita
  - překonání rozdílností mezi právními předpisy v jednotlivých členských státech,
  - zajištění, aby při výměně informací přes hranice byla zachována právní platnost těchto informací,
  - dodržení právních předpisů o ochraně údajů v poskytující i přijímající zemi.
- Sémantická interoperabilita
  - Zajišťuje, že při výměně informací mezi stranami bude správně chápán a zachován smysl vyměňovaných informací (smysl datových prvků a vztahů mezi nimi).
  - Zahrnuje vytvoření slovníku pojmů pro popis vyměňovaných údajů a zajišťuje, aby všechny datové prvky byly všemi komunikujícími stranami chápány shodně.
- Syntaktická interoperabilita
  - popis přesného formátu vyměňovaných informací (dat), pokud jde o:
    - gramatiku,
    - formát
    - schémata.



- Technická interoperabilita
  - technické stránky zajištění propojení informačních systémů a sdílených informací,
    - specifikace rozhraní,
    - spojové služby,
    - služby integrace dat,
    - prezentace a výměny údajů.

### 9.1.4 Principy otevřených dat (Mezinárodní charta otevřených dat, 2015).

Principy charty byly vytvořeny v roce 2015 ze strany vlád, občanské společnosti a expertů z celého světa a reprezentují sadu globálně smluvených norem, které aspirují na to být normou pro publikaci dat.

- **Otevřenost jako standard (Open by default),**
  - zveřejňování všech dat automaticky s určitými výjimkami například z důvodu:
    - bezpečnosti,
    - ochrany osobních údajů a
    - ochrany autorských práv,
  - opakem tohoto principu je přístup zveřejňování dat z předem vymezených oblastí s tím, že ke zveřejnění zbytku informací nedochází.

Vládní program digitalizace České republiky 2018+ „**digitální ; Česko**“ v „Informační koncepci ČR“ zařadil tento princip mezi „Principy eGovernmentu EU“.

#### **P13: Otevřená data jako standard (Open Data by default).**

*„Veřejné údaje evidované orgány veřejné správy ve spravovaných ISVS musí být zveřejňovány jako otevřená data. Pro neveřejné údaje musí být jako otevřená data zveřejňována jejich anonymizovaná podoba, souhrn nebo statistika, nebo obdobná forma, pokud může mít význam pro uživatele těchto dat. V případě, že orgány veřejné správy sdílejí veřejné údaje včetně anonymizované podoby neveřejných údajů, souhrnů nebo statistik*

- **Včasnost a ucelenost dat (Timely and Comprehensive),**
  - data by měla být stále relevantní, tj. měla by být poskytnuta včasně,
  - ucelená, přesná a vysoké kvality,
  - původní, nemoifikovaná.
- **Přístupnost a použitelnost (Accessible and Usable),**
  - data by měla být strojově čitelná a snadno dostupná,
  - poskytovaná zdarma na základě otevřené licence se zohledněním zkušenosti jejich uživatelů.
- **Srovnatelnost a interoperabilita (Comparable and Interoperable),**
  - k docílení snadného vzájemného porovnání a interoperability dat je nutné dodržovat odsouhlasené datové standardy.
- **Zlepšení vládnutí a zapojení občanů (For Improved Governance and Citizen Engagement),**
  - otevřená data zvyšují transparentnost a umožňují občanům se zapojit do správy věcí veřejných;
- **Inkluzivní rozvoj a inovace (For Inclusive Development and Innovation),**
  - otevřená data mají potenciál přinášet výhody a přínosy všem uživatelům, veřejnému sektoru, privátnímu sektor a společnosti.



### 9.1.5 Kvalita otevřených dat a kvalita jejich publikace

Cílovou podobu publikace otevřených dat lze charakterizovat ve třech pohledech:

**Kvalita samotných publikovaných dat** se obecně týká toho, jak dobře jsou data připravena k použití a jak vhodná jsou pro zamýšlený účel. Základními charakteristikami jsou:

**Přesnost:** Data by měla být co nejpřesnější a aktualizovaná.

### 9.1.6 FAIR principy

Otevřená data v ČR jsou v souladu s obecnými principy FAIR – ty požadují, aby data byla:

- **(F)** vyhledatelná (Findable), vyhledatelnost otevřených dat je zajištěna jejich registrací do NKOD a vyplněním všech povinných položek metadat.
- **(A)** dostupná (Accessible), dostupnost je zajištěna jejich vystavením na internetu.
- **(I)** interoperabilní (Interoperable), interoperabilita je zajištěna využíváním otevřených formátů, strojovou čitelností, standardy a propojením se sémantickým slovníkem pojmů veřejné správy..
- **(R)** opětovně využitelná (Reusable). Opětovná využitelnost je pak zajištěna vhodnými (neomezujícími) podmínkami užití.

### 9.1.7 Způsoby vzniku hodnoty na základě otevřených dat (zhodnocování).

Zhodnocováním otevřených dat vznikají hodnoty **sociální**, nebo **ekonomické**. Otevřená data přitom vystupují ve dvou rolích:

- zdroj pro tvorbu ekonomických hodnot,
- nástroj pro utváření společenských hodnot.

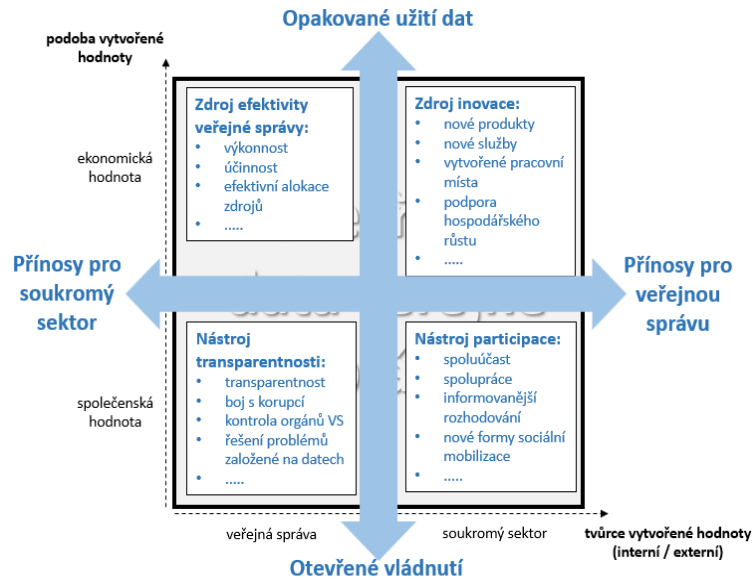
Tvůrcem vytvářených hodnot může být:

- přímo veřejná správa,
- nebo soukromý sektor.

Veřejná správa vytváří hodnotu již správnou a kvalitní publikací otevřených dat, správným výběrem publikovaných dat (např. zaměřením na podporu transparentnosti organizací) a užíváním publikovaných otevřených dat pro výkon veřejné správy.

Soukromý sektor vytváří hodnotu otevřených dat jejich dalším zhodnocením a opakovaným používáním, obvykle tvorbou aplikací, služeb a nových produktů.

Způsoby zhodnocování otevřených dat a jejich význam zachycuje obrázek.



### 9.1.8 Význam a role otevřených dat

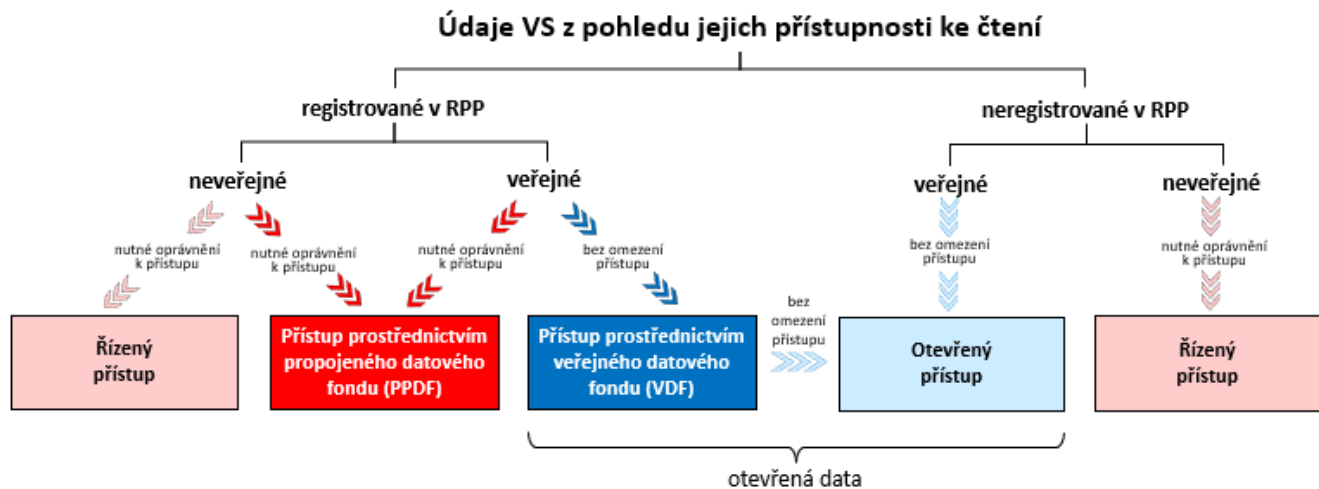
- **Nástroj a zdroj k dosažení cílů Open Governmentu:**
  - Zlepšení vládnutí v působnosti organizace a jejich kompetencí,
    - boj proti korupci,
    - posílení veřejných služeb a alokace zdrojů,
    - zvýšení transparentnosti.
  - Posílení postavení občanů,
    - informovanější rozhodování,
    - nové formy sociální mobilizace,
    - nové způsoby komunikace a přístupu k informacím,
    - posílení postavení občanů, aby převzali kontrolu nad svým životem.
  - Podpora vytváření hospodářských příležitostí,
    - podpora inovací,
    - podpora hospodářského růstu,
    - vytváření pracovních míst.
  - Řešení veřejných problémů,
    - přístup k novým formám hodnocení těchto problémů založených na datech,
    - zapojení založené na datech, které vede k cílenějším intervencím a lepší spolupráci.
- **Základní metoda pro sdílení veřejných informací** mezi veřejnoprávními subjekty navzájem i pro sdílení veřejných údajů mezi veřejnoprávní a soukromoprávní sférou **v ČR (e-Government)**.
- Jeden ze základních pilířů SMART CITIES (REGIONS, ...).
- Základ jednotného digitálního trhu EU.
- Základ společného evropského datového prostoru.

## 9.2 Údaje VS

Údaje veřejné správy jsou údaje evidované jednotlivými OVM v jejich informačních systémech. Z pohledu sdílení údajů je důležité jejich členění.



## 9.2.1 Rozdělení údajů z pohledu jejich sdílení



**Registrované údaje** označují všechny údaje zaevidované do Registru práv a povinností (RPP):

- při ohlášení agend dle [§ 51 odst. 6, písm. k\)](#) zákona č. 111/2009 Sb. o základních registrech,
- číselníky evidované v RPP dle [§ 50 odst. 2](#) a [§ 54 odst. 1, písm. a\)](#) zákona č. 111/2009 Sb. o základních registrech.

**Neregistrované údaje** označují všechny ostatní údaje spravované OVM.

- Neregistrovaným údajem je např. i agregace údaje evidovaného v RPP sloužící ke statistickým účelům, pokud není samotná agregace evidována v RPP.

**Veřejné údaje** jsou údaje, které může kdokoliv číst bez omezení přístupu.

- Veřejné údaje jsou všechny údaje, u kterých neexistuje nějaké zákonné omezení jejich veřejnosti.
- Veřejnost údaje označuje ohlašovatel agendy v RPP tím, že ho označí jako veřejný.
- Číselníky jsou veřejné všechny bez výjimky, [§ 51 odst. 11](#) zákona č. 111/2009 Sb. o základních registrech.

**Neveřejné údaje** jsou údaje, pro jejichž čtení je oprávnění nutné. Pro zápis údajů (neveřejných i veřejných) je nutné oprávnění vždy.

- Neveřejnost údaje označuje ohlašovatel agendy tím, že ho označí jako neveřejný, pokud existuje legislativní ustanovení, ze kterého neveřejnost údaje jednoznačně vyplývá.
- Omezení veřejnosti údajů vychází primárně ze [zákona č. 106/1999 Sb.](#) o svobodném přístupu k informacím, a to zejména z ochrany utajovaných informací, ochrany obchodního tajemství, ochrany důvěrnosti majetkových poměrů, ochrany osobních údajů a dalších omezení práva na informace dle tohoto zákona, sekundárně veřejnost údajů upravují další speciální právní předpisy.

## 9.2.2 Přístupy k údajům veřejné správy

- **Řízený přístup.**
  - Řízený přístup je určen k přístupu k neveřejným údajům (registrovaným i neregistrovaným) kýmkoliv (tj. jednotlivými OVM, SPUÚ mimo výkon agend i veřejností) formou čtení na základě definovaných oprávnění.
- **Přístup prostřednictvím propojeného datového fondu (PPDF).**
  - PPDF je určen ke sdílení registrovaných údajů (veřejných i neveřejných) mezi jednotlivými OVM a ke zpřístupnění těchto údajů jednotlivým SPUÚ pro účely výkonu jejich agend, a to na základě



oprávnění pro čtení nebo pro zápis evidovaných v RPP.

- **Přístup prostřednictvím veřejného datového fondu (VDF).**

- VDF je určen ke sdílení veřejných registrovaných údajů mezi jednotlivými OVM a ke zpřístupnění těchto údajů jednotlivým SPUÚ. Ve všech případech se jedná o čtení údajů v podobě otevřených dat bez omezení přístupu, a tudíž není nutné pro OVM a SPUÚ získat oprávnění pro jejich čtení.
- VDF je ustanoven a definován v Informační koncepci ČR (IKČR) v dílčím cíli 5.10 jako součást eGovernmentu VS ČR:
  - “Veřejný datový fond tvořený publikovanými veřejnými údaji veřejné správy je základní metodou pro sdílení veřejných informací mezi veřejnoprávními subjekty navzájem i pro sdílení veřejných údajů mezi veřejnoprávní a soukromoprávní sférou v ČR. Veřejný datový fond se od pouhé publikace automatizovaně čitelných otevřených dat posune též k publikaci právně závazných, platných a pravidelně aktualizovaných datových sad s jasně definovanou zodpovědností OVS za takové sady.”
- Obecným výchozím principem, který je základem VDF, je princip P13 eGovernmentu “Otevřená data jako standard” (Open Data by Default):
  - “Veřejné údaje evidované orgány veřejné správy ve spravovaných ISVS musí být zveřejňovány jako otevřená data. Pro neveřejné údaje musí být jako otevřená data zveřejňována jejich anonymizovaná podoba, souhrn nebo statistika. V případě, že orgány veřejné správy sdílejí veřejné údaje (včetně anonymizované podoby neveřejných údajů, souhrnů nebo statistik) musí je sdílet jako otevřená data.”
- Detailní popis VDF je k dispozici na stránkách <https://archi.gov.cz> a „[Metodice poskytování dat ve veřejném datovém fondu](#)“.

- **Otevřený přístup.**

- Otevřený přístup je určen k přístupu k veřejným údajům (registrovaným i neregistrovaným)

**Přístup k údajům VS prostřednictvím veřejného datového fondu (VDF) i prostřednictvím otevřeného přístupu je založen na publikovaných údajích VS ve formátu otevřených dat.**

kýmkoliv formou čtení, a to bez omezení přístupu, v podobě otevřených dat dle [§ 3a odst. 5 zákona č. 106/1999 Sb.](#) o svobodném přístupu k informacím.

### 9.2.3 Registrované veřejné údaje

Podle [§ 51 odst. 6, písm. k\) zákona č.111/2009 Sb. o základních registrech.](#) je pro každý údaj vedený nebo vytvářený v rámci agendy v RPP vedena jeho přístupnost veřejnosti a v případě, že údaj není přístupný veřejnosti, je povinností uvést číslo a název právního předpisu a označení jeho ustanovení, na jehož základě není údaj přístupný veřejnosti (resp. seznam takových referencí na ustanovení právních předpisů).

Pokud tedy neexistují legislativní překážky ke zveřejnění údaje, jedná se o údaj přístupný veřejnosti (dále jen veřejný údaj). Veřejný údaj je speciálním případem údaje.

Způsob poskytnutí údaje zveřejněním upravuje [§ 4b zákona č. 106/1999 Sb. o svobodném přístupu k informacím](#), dle kterého se veřejný údaj poskytuje jako informace ve všech formátech a jazycích, ve kterých byla informace





vytvořena; při zveřejnění takové informace v elektronické podobě musí být jeden z těchto formátů otevřený a, je-li to možné, též strojově čitelný. Je-li to možné a vhodné, zveřejní povinný subjekt spolu s informací též metadata, která se k ní vztahují. Formát i metadata musí co nejvíce splňovat otevřené formální normy.

Z důvodu dosažení interoperability mezi informačními systémy veřejné správy a také z důvodu dosažení přeshraniční interoperability je nutné při zveřejňování údajů postupovat podle otevřených formálních norem vydávaných MV ČR a při zveřejňování metadat je nutné postupovat dle otevřené formální normy rozhraní katalogů otevřených dat.

Dále je dle [§ 51 odst. 6, písm. k\) zákona č.111/2009 Sb.](#) pro údaj vedený nebo vytvářený v rámci agendy, jehož možné hodnoty jsou vymezeny číselníkem nutné, aby tento číselník byl evidován v RPP.

Z [§ 54 odst. 1, písm. a\) zákona č.111/2009 Sb.](#) vyplývá, že primárně musí být volen číselník, který již je v RPP veden. Pouze pokud žádný vhodný číselník v RPP veden není, uvádí ohlašovatel agendy svůj vlastní číselník, tj. nový číselník ohlašovatele agendy.

Dle [§ 69 odst. 4 zákona č.111/2009 Sb.](#) MV ČR zveřejnilo otevřenou formální normu pro vytváření číselníků a podle [§ 51 odst. 11 zákona č.111/2009 Sb.](#) jsou číselníky veřejně přístupné způsobem umožňujícím dálkový přístup.

Ze spojení s [§ 5a odst. 2 zákona č. 106/1999 Sb.](#) pak vyplývá, že číselník musí být zveřejněn jako otevřená data podle otevřené formální normy vyplývající z [§ 69 odst. 4 zákona č.111/2009 Sb.](#)

Z pohledu zákona o základních registrech lze rozlišit dva možné typy veřejných číselníků.

- Dle [§ 50 odst. 2 zákona č.111/2009 Sb.](#) číselníky ČSÚ.
- Dle [§ 54 odst. 1, písm. a\) zákona č.111/2009 Sb.](#) číselníky ohlašovatelů agend.

Možné hodnoty údaje lze tedy vymežit buď číselníkem ČSÚ nebo číselníkem ohlašovatele agendy.

## 9.3 Příklad ošetření navržených témat otevřených dat

### 9.3.1 Určení přínosů datových sad

Jak již bylo uvedeno v kapitole „[Ošetření navržených témat otevřených dat](#)“, v případě „povinného otevírání dat“ na základě povinností stanovených v legislativě a uplatňování principu P13 „Open data by Default“ platí, že veřejné údaje se publikují bez ohledu na identifikované přínosy. Z toho pohledu je analýza přínosů nerelevantní, neboť základní přínos vyplývá:

**z veřejného přístupu k veřejným údajům veřejné správy a  
podporou efektivity výkonu veřejné správy prostřednictvím nové  
metody vzájemného sdílení údajů (V/D/E) organizacemi veřejné správy**

Pro ostatní otevřená data (dobrovolně otevíraná data) je důležité, aby jejich existence přinášela nějaký užitek. U každé navrhované datové sady je proto důležité analyzovat, jaké přínosy její otevření přinese.

([odkaz POD](#))

- Analýzu přínosů provádí Koordinátor otevírání dat ve spolupráci s Kurátory dat a je doporučen následující postup:
- Pro datové sady uvedené ve vzorovém publikačním plánu jsou přínosy identifikovány a lze je snadno převzít.
- Pokud jsou do seznamu přidány vlastní datové sady, ohodnocení je vhodné provést pásmově:
  - 0 - žádný přínos,
  - 1 - nízký přínos,



- 2 - střední přínos,
- 3 - významný přínos.
- **Seznam kritérií pro posouzení přínosů:**
  - Posílení transparentnosti,
    - lepší dostupnost a přístup k informacím o fungování orgánů veřejné správy, například přístup ke smlouvám a veřejným zakázkám.
  - Podpora ekonomického růstu,
    - zveřejnění informací může vést k hospodářskému a ekonomickému růstu,
    - zveřejnění informací veřejného sektoru může posloužit jako zdroj pro nové inovace, služby, produkty a podnikatelské příležitosti.
  - 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 orgánem veřejné správy.
  - Podpora opětovného použití dat,
    - strojová čitelnost, dostupnost metadat a právní otevřenost dat zajišťují větší znovu využitelnost dat veřejného sektoru,
    - jednodušší vývoj aplikací a služeb postavených na otevřených datech.
  - Zlepšení vnímání veřejné správy veřejností,
    - instituce státní správy a samosprávy se stanou otevřenější a transparentnější.
  - Zlepšení procesů a dat veřejné správy, zlepšení komunikace a spolupráce veřejné správy,
    - může se zlepšit komunikace mezi jednotlivými orgány veřejné správy,
    - publikace otevřených dat může nahradit někdy zbytečně složité předávání dat mezi jednotlivými institucemi.
  - Zamezení chyb vzniklých při práci s daty,
    - dostupnost dat ve strojově čitelném formátu může minimalizovat potřebu manuálního přepisování dat z dokumentů do jiných systémů a celkově se může snížit chybovost.
  - Snížení počtu dotazů dle zák. č. 106/1999 Sb.,
    - může se výrazně snížit počet žádostí dle zák. č. 106/1999 Sb.
  - Pořádek ve vlastních datech,
    - analýza dat, katalogizace a publikace umožní instituci uspořádat svá vlastní data a optimalizovat tak i vlastní interní procesy.
  - Zvýšení hodnoty dat,
    - strojová čitelnost a otevřenost dat spolu s metadaty přidávají datům novou hodnotu, která může být zhodnocena v různých vyvinutých aplikacích a zejména rozvinuta v kontextu Otevřených propojitelných dat.

### 9.3.2 Určení rizik 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.

[\(odkaz POD\)](#)

- Analýzu rizik provádějí Koordinátor otevírání dat s Kurátory dat a je doporučen následující postup:



- pro datové sady uvedené ve vzorovém publikačním plánu jsou rizika identifikována a lze je snadno převzít,
- pokud jsou do plánu přidány vlastní datové sady, ohodnocení se provede pásmově:
  - 0 - žádné riziko,
  - 1 - nízké riziko,
  - 2 - střední riziko,
  - 3 - významné riziko.
- Celkové zhodnocení rizikovosti datové sady musí být následováno návrhem opatření pro eliminaci či omezení pravděpodobnosti výskytu nebo negativního dopadu rizik.
- **Seznam kritérií pro posouzení rizik:**
  - Zveřejnění dat v rozporu se zákonem,
    - ověření, zda publikaci datové sady nebrání související legislativa.
  - Porušení ochrany obchodního tajemství,
    - zvážení, zda se publikací neporušuje obchodní tajemství, specifikované např. ve smlouvě s dodavatelem.
  - Porušení ochrany osobních údajů,
    - ověření, zda součástí otevřených dat nejsou chráněné osobní údaje.
  - Zveřejnění nevhodných dat či informací,
    - zvážení, zda publikovaná data nebudou mít negativní dopady, např. z hlediska kvality publikovaných dat.
  - Dezinterpretace dat,
    - zvážení, zda otevřená data nebude možno dezinterpretovat, např. nevysvětlením používaných pojmů.
  - Absence konzumentů dat,
    - ověření, zda je o publikovaná data zájem ze strany budoucích uživatelů dat.
  - Překrývání dat,
    - ověření, 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.

### 9.3.3 Ošetření rizik

Celkové zhodnocení rizikovosti 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,
    - pokud nelze poskytnout data z důvodu požadavků na jejich ochranu, je vhodné zvážit, zda lze publikovat alespoň anonymizovaná data či statistiky,
    - u datové sady, která obsahuje osobní údaje je třeba dbát zvýšené opatrnosti.
    - Obecně až na úplné výjimky platí, že pokud neexistuje zákonná povinnost taková data zveřejnit, tak se zveřejnit nemohou.
    - V případech, že zákonná povinnost existuje, je možné osobní údaje zveřejnit vzhledem k



nutnosti splnění takové povinnosti, ale do popisu datové sady je doporučeno vložit následující sdělení určené příjemcům dat:

- „Poskytovaná datová sada obsahuje osobní údaje ve smyslu zákona č. 101/2000 Sb., o ochraně osobních údajů a o změně některých zákonů. Příjemce této datové sady se stává správcem osobních údajů. Vzhledem k tomu musí plnit zákonné povinnosti, které mu z účinné právní úpravy ochrany osobních údajů vyplývají.“
- Návrh vhodné komunikační strategie.
  - Stanovit, jak bude o datové sadě informováno, identifikovat potenciální negativní reakce a připravit odpovědi,
  - vhodné oslovení cílových skupin uživatelů dat před otevřením dat (analýza zainteresovaných stran).
- Vytvoření vhodných metadat a doplňujících informací,
  - 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,
  - vymezení datové sady s ohledem na minimalizaci překryvů s jinými datovými sadami,
  - využití již publikovaných datových sad,
  - propojování na související datové zdroje.

## 9.4 Datové schéma

### 9.4.1 Výklad problematiky vymezení věcného obsahu datové sady

Každé navržené téma je zpravidla definováno svým názvem, odpovědným kurátorem a stručným popisem. Toto ještě ale nemusí znamenat přesné vymezení rozsahu publikovaného kontextu tématu, a tedy ani rozsahu publikovaných údajů. Je nutné si uvědomit, že při použití zejména hierarchických datových modelů model datové sady je obvykle určován jednou kořenovou entitou, a ta předurčuje účel využití příslušné datové sady. Při komplexnějších tématech (zastřešujících například tematickou oblast) to může odpovídající datovou sadu omezit při jejím využití. Pro komplexnější témata je jednoznačně výhodnější používat pro jejich publikaci grafové modely, které obsahují kompletní provázaný kontext celé tematické oblasti.

#### Vysvětlující jednoduchý příklad.

- Navržené téma otevřených dat: Prezenční listina zasedání.
- Popis: Prezenční listina ze zasedání shrnující přítomnost na konkrétním zasedání a případné omluvy.

Tímto způsobem navržené téma je velmi jednoduché, ale také současně málo využitelné. V podstatě se jedná pouze o publikaci jednoduché tabulky.

ID pracovníka	jméno	kontakt	funkce	přítomen	omluven
---------------	-------	---------	--------	----------	---------

Cílem v tomto jednoduchém případě je tedy publikovat informace zachycené v jedné konkrétní prezenční listině, přičemž informace o tom, o jaké zasedání se jednalo a v kterém konkrétním termínu proběhlo, budou obsaženy



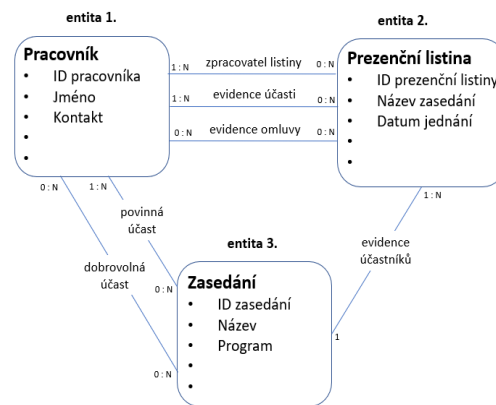
v metadatach publikované sady.

### Rozšíření uvedeného příkladu.

Širším cílem publikace otevřených dat je ale publikace datových sad určených k opakovanému a mnohostrannému využití s možností jejich propojování s dalšími datovými sadami do širších kontextů.

- Navržené téma otevřených dat: „Mini-agenda“ zasedání.
- Popis: Publikace informací o zasedáních.

Zjednodušená představa tématu (konceptuální model) zahrnuje existenci tří provázaných entit.



(Entita je základním pojmem při popisu reálného světa a práci s informacemi o něm. Vlastnosti entit se popisují atributy. Mezi jednotlivými entitami existují vzájemné vztahy – relace.)

Zobrazený konceptuální model nyní nabízí následující možnosti jeho publikace.

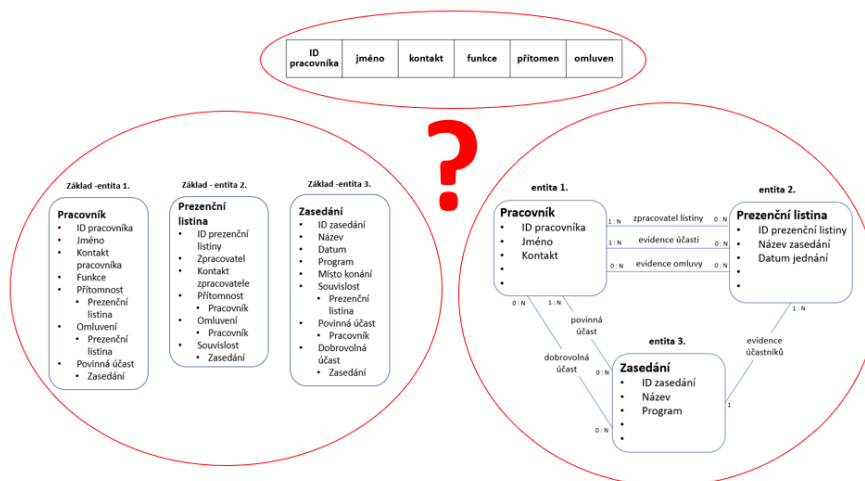
- Publikace datové sady na základě pouze jedné entity bez publikace vazeb na navázané entity (tabulková data).
  - Základ entita 1 – publikace seznamu pracovníků.
  - Základ entita 2 „Prezenční listina“ – publikace účasti na jednom konkrétním zasedání, stejný případ, jako byl uveden u jednoduchého příkladu.
  - Základ entita 3 – publikace přehledu všech zasedání.
- Publikace datové sady na základě jedné kořenové entity a s odvozením dalších atributů prostřednictvím navázaných entit (hierarchicky strukturovaná data).
  - Základní entita 1 a entita 2 – datová sada bude obsahovat informace o pracovnících, jejich účasti na zasedáních, včetně termínů a zda byli případně omluveni za neúčast. Bude reprezentovat nasazení a disciplínu každého pracovníka.
  - Základní entita 1 a entity 2 a 3 – datová sada přináší navíc informace ohledně vyžadované povinné účasti pracovníků na zasedáních a jakých témat zasedání se měli zúčastnit. Bude reprezentovat plnění pracovních povinností každého pracovníka.
  - Základní entitou je entita 2 – datová sada bude obsahovat informace o prezenčních listinách,



kterých zasedání se týkaly, v jakých termínech, jaká byla na nich účast, kdo se jich zúčastnil a kdo byl omluven.

- Základní entitou je entita 3 - datová sada bude obsahovat informace o zasedáních a jejich programech, o povinných účastnících a informace o skutečné účasti.
- Publikace datové sady se všemi entitami a vztahy mezi nimi (propojená data).
  - Způsob použití a možnosti vytěžení informací z datové sady je omezen pouze kvalitou namodelování reality – kvalitou konceptuálního schématu.

Rozhodnutí o rozsahu publikovaného kontextu tématu je zásadním předpokladem pro zajištění široké využitelnosti publikovaných datových sad.



### Další zvažovaná hlediska při vymezení rozsahu obsahu datových sad:

- počet entit (subjektů a objektů práva v případě agendových údajů) s atributy v údajích a jejich vzájemné vazby – složité datové sady vedou ke komplikovaným datovým schémátům a konceptuálním modelům,
- vymezení tematických oblastí, např. v případě agend zákon ustavující příslušnou agendu může přesně vymezovat (oddělovat) vedené rejstříky, seznamy a evidence,
- smysluplný kontext – i samostatná datová sada by měla být cíleně použitelná a měla by dávat smysl,
- cíl publikace a použití publikované datové sady – vymezuje kořenovou entitu a její atributy, předurčuje jednoduchost / složitost její struktury,
- místo evidence a správy publikovaných údajů (entit, jejich atributů a vzájemných vazeb), v případě rozproštění informací mezi více zdrojů (informačních systémů) je nutné mít na paměti, že automatizovaná publikace datových sad by byla komplikovaná,
- místa evidence a správy informací o subjektech a objektech práva agendy (jednotlivé ISVS), v případě rozproštění informací mezi více ISVS je nutné mít na paměti, že publikace datových sad (zpřístupnění a katalogizace) je vždy zajišťována tím ISVS, který informace spravuje (publikace společné datové sady s informacemi z více ISVS by byla komplikovaná),



- množství publikovaných atributů – hodně atributů vzniká v důsledku vazeb na jiné objekty (entity), v takových případech je vhodnější údaje rozdělit do několika datových sad (například podle jednotlivých entit) a tím zmenšit datové struktury, a přitom dodržet publikaci všech atributů,
- celkové množství publikovaných dat – obtížnější práce s velkými datovými soubory.

## 9.4.2 Stanovení stupně otevřenosti

Detailní výklad problematiky stanovení stupně otevřenosti je uveden na stránkách Portálu otevřených dat ([odkaz POD](#)).

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

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.

Míru otevřenosti dat je možné vyjádřit pomocí 5 stupňů otevřenosti:

- [stupeň 1](#) - datová sada je dostupná v síti WWW s vhodnými podmínkami užití otevřených dat,
- [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.

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. Neznamená to však, že otevřená datová sada nesmí být publikována na stupni otevřenosti 2.

**Je vždy nutné zajistit publikaci kompletního obsahu datové sady 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 strojově čitelná data, ale



právě aplikace, které jsou nad nimi vytvořené). Existence nebo neexistence takových aplikací však nijak neovlivňuje otevřenost datové sady.

## 9.5 Datové specifikace

Pokud poskytovatel dat provozuje pouze aplikaci (např. HTML formulář pro vyhledávání záznamů a jejich následné zobrazení) nad otevřenými daty a neposkytuje současně kompletní obsah datové sady na stupni otevřenosti minimálně 3, **neposkytuje otevřená data**.

Interoperabilita publikovaných otevřených dat je zajištěna standardizací jejich publikace na základě Otevřených formálních norem (OFN) vydaných garantem otevřených dat a Datových specifikací vytvářených jednotlivými poskytovateli dat.

Datová specifikace je z pohledu obsahu a struktury všech vytvořených artefaktů v podstatě shodná s [OFN](#), se dvěma zásadními rozdíly:

- není oficiálně vydaná a autorizovaná garantem OD veřejné správy ČR dle § 3a odst. 3 zákona č. 106/1999 Sb.
- součástí datové specifikace nejsou předpřipravená metadata v podobě vzorových metadatových záznamů pro registraci do NKOD.

OFN je v principu specializací Datové specifikace, přičemž její specializace je dána především její přípravou, způsobem publikace a autorizací garantem OD veřejné správy ČR.

Datová specifikace popisuje strukturu a sémantiku konkrétní datové sady nebo datových sad. Obsahuje dokumentaci významu obsahu datové sady v podobě jejího konceptuálního modelu, jeho popisu a datová schémata určující datové struktury, ve kterých je obsah distribuován v různých formátech.

**Datovou specifikaci na rozdíl od OFN vydává konkrétní poskytovatel a specifikuje pomocí ní konkrétní podobu datové sady, datových sad, nebo často se opakujících částí jeho konkrétních publikovaných datových sad (obdobu sdílených specifikací v případě OFN). ve kterých**

### Základní struktura datové specifikace:

- Přehledové schéma (konceptuální model, datový model).
- Specifikace datových tříd, jejich vlastností a vazeb na jiné datové třídy nebo sdílené specifikace. Pro každou vlastnost je uveden její identifikátor, který je pro její reprezentaci použit ve všech datových formátech, její název, datový typ, popis a příklad. Všechny třídy, které nejsou specifikovány jinou otevřenou formální normou nebo číselníkem, dědí vlastnosti a vazby třídy „Věc“.
  - Příklad specifikace vlastnosti „Tituly před jménem“ datové třídy „Člověk“:
    - Vlastnost: jméno
    - Vlastnost: tituly\_před
    - Typ: Text
    - Jméno: Tituly před jménem
    - Popis: Tituly před jménem.





- Povinnost: Nepovinná
- Příklad: RNDr.
- Uvedené příklady reprezentace v různých úrovních detailu, ve formátu JSON a JSON-LD.

**Informace o použité „Datové specifikaci“, nebo o použité OFN při návrhu publikované datové sady je uvedena v jejím katalogizačním záznamu v položce „Specifikace“.**

Více informací k přípravě „Datové specifikace“ lze nalézt v „[Návodu k použití OFN poskytovateli dat](#)“ a také v již připravených a publikovaných [OFN](#).

## 9.5.1 OFN

### 9.5.1.1 Stručné seznámení s OFN

Otevřené formální normy (OFN) jsou technická doporučení zaměřená na vybrané datové sady. Zajišťují, že stejná data publikovaná různými poskytovateli budou interoperabilní a bude je možné jednodušeji využívat nezávisle na tom, kdo data poskytuje. Pro poskytovatele otevřených dat, kteří jsou povinnými subjekty dle zákona č. 106/1999 Sb. o svobodném přístupu k informacím, jsou doporučení vyplývající z OFN závazná dle § 4b odst. 1.

OFN jsou v principu specializací Datové specifikace, přičemž její specializace je dána především její přípravou, způsobem publikace a autorizací garantem OD veřejné správy ČR

Dále uvedené postupy v této kapitole jsou pro názornost demonstrovány formou „ručních operací“, jejichž výsledkem je pouze zajištění syntaktické interoperability. Pro zajištění i plné sémantické interoperability jsou nezbytné připravené konceptuální modely s vazbami do slovníku sémantických pojmů, pro jejichž přípravu lze výhodně využít tzv. „výrobní linku pro modelování“. Z vytvořených konceptuálních modelů lze totiž následně pomocí specializovaného nástroje (dataspecer) přímo generovat nejen datová schémata (datové specifikace), ale i nové otevřené formální normy, a to vše včetně využití již vydaných stávajících OFN a sdílených specifikací. Více v kapitole „[Podpora nástrojů](#)“.

#### **Obecný princip konstrukce OFN bez zřetele na konkrétní témata.**

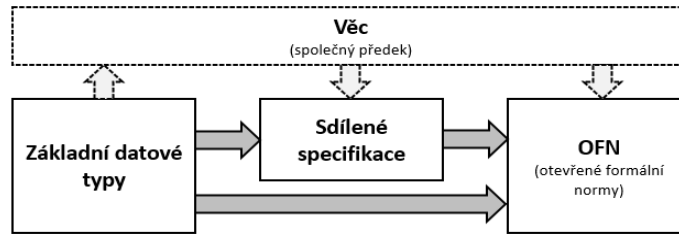
Konstrukce OFN je založena na výstavbě datových specifikací (datových modelů) pomocí standardizovaných „stavebních prvků“. Jednotlivé stavební prvky jsou vytvářeny postupnou konstrukcí a kombinacemi od těch nejjednodušších prvků (základní datové typy) až po ty nejsložitější (OFN pro úplná datová témata). Všechny vytvářené prvky dědí vlastnosti od svého společného předka „Věc“. Ve strojovém vyjádření jsou konstrukce vytvářených „vyšších“ stavebních prvků tvořeny pouze pomocí vzájemných referenčních odkazů, což znamená, že „nižší“ stavební prvky jsou vzájemně pouze sdíleny.

Základním stavebním prvkem OFN jsou základní datové typy, pomocí kterých je jednak vytvořen společný předek „Věc“ pro všechny další prvky vytvářených OFN, a také jsou základem vytvořených sdílených specifikací. Výsledná OFN je následně tvořena kombinací všech vytvořených prvků (základní datové typy a sdílené specifikace) a zděděnými vlastnostmi společného předka „Věc“.

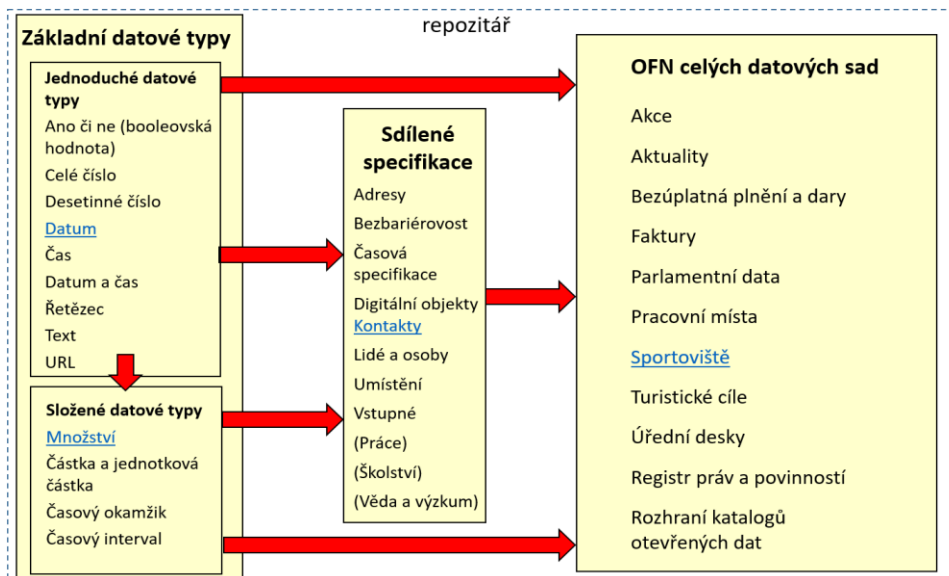
Při tvorbě OFN pro konkrétní datové sady sehrává klíčovou roli téma OFN a jeho významový kontext vyjádřený konceptuálním modelem..



## Obsah a struktura repozitáře OFN:



- **Základní datové typy**
  - Ano či ne (booleovská hodnota), Celé číslo, Desetinné číslo, Datum, Čas, Datum a čas, Řetězec, URL, Cena, Množství
- **Věc**
  - Společný předek sdílených specifikací, charakterizovaný základními vlastnostmi: identifikátor, název, popis, čas zneplatnění apod.
- **Sdílené specifikace**
  - Společné části vyskytující se ve více datových sadách, na které se odkazují ostatní otevřené formální normy a specifikace.
  - V přehledových obrázcích v OFN jsou tyto typy entit označovány rámečkem s šedě podbarveným nadpisem.
- **Specifické OFN**
  - Zachycují ucelený kontext konkrétního publikovaného tématu, nebo řešeného problému.
  - Dělí se na:
    - OFN pro konkrétní datové sady: specifikují, jak publikovat data o konkrétních věcech, např. sportovištích, akcích apod.
    - Obecnější OFN pro různé typy dat: nejsou spojeny s žádným konkrétním tématem, ale spíše s typy, nebo způsobem publikace datových sad.

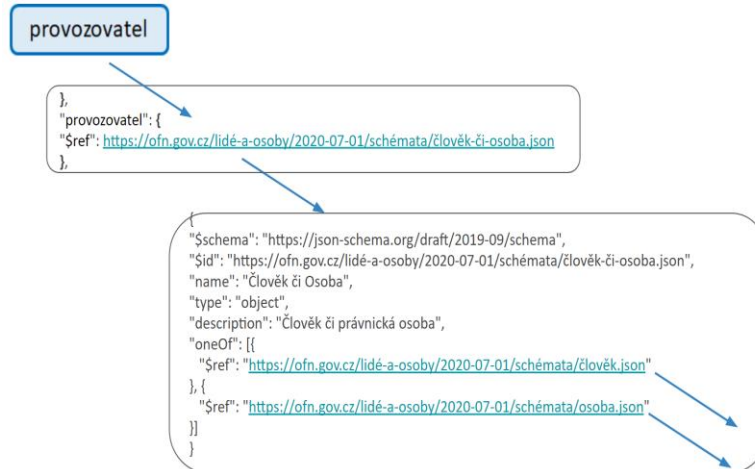


[Obsah repozitáře OFN](#) je dostupný na stránkách Portálu otevřených dat.

Repozitář OFN si lze také představit jako množinu hierarchických stromových modelů dat (reprezentující datová



schémata a vyjádřených ve formátu JSON), které navzájem sdílí společné prvky repozitáře od těch nejjednodušších (základních datových typů) až po ty nejsložitější (specifické OFN). Toto sdílení výrazně zjednodušuje vytváření OFN i datových schémat datových sad. Při jejich tvorbě stačí pro každou položku datového schématu vždy jenom odkazovat na vrchol hierarchického datového modelu odpovídajícího ošetřované položce datového schématu (odpovídající základnímu datovému typu, nebo sdílené specifikaci).



### Návrh datového schématu

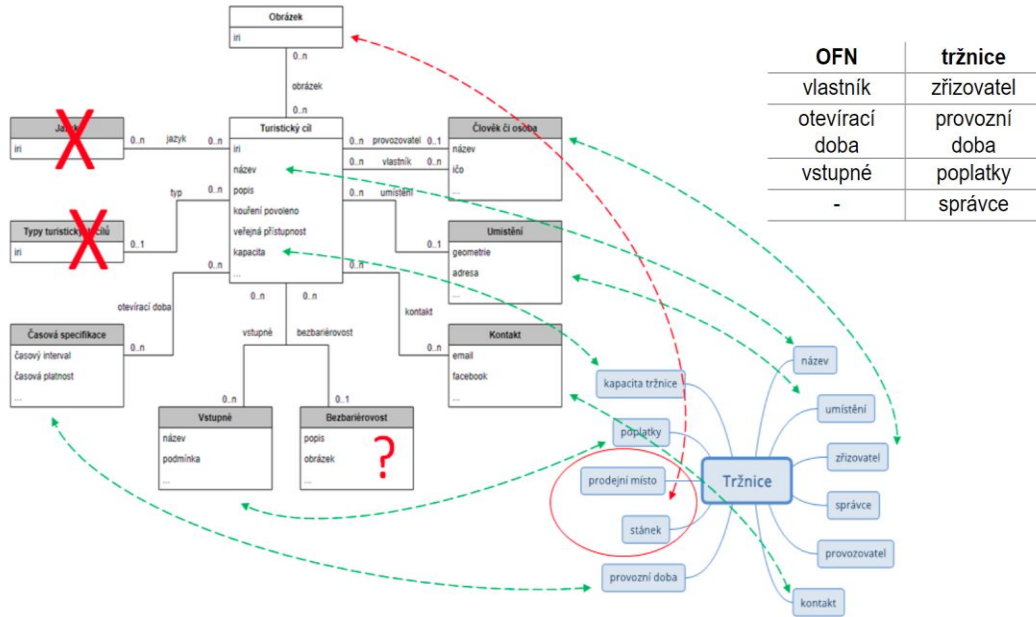
Návrh na základě prvků repozitáře OFN se soustředí do tří kroků:

- správné odvození datového modelu z konceptuálního modelu tématu datové sady,
- ošetření každého prvku datového modelu referenčním odkazem na vrchol datového modelu odpovídajícího prvku repozitáře OFN,
- sestavení strojově čitelného datového schématu pomocí základních datových typů a referenčních odkazů na prvky repozitáře OFN ve stejné hierarchické struktuře, jako má datový model datové sady.

```
"správce": {
  "$ref": "https://ofn.gov.cz/lidé-a-osoby/2020-07-01/schéματα/člověk-či-osoba.json"
},
"umístění": {
  "$ref": "https://ofn.gov.cz/umístění/2020-07-01/schéματα/umístění.json"
},
"kontakt": {
  "type": "array",
  "items": {
    "$ref": "https://ofn.gov.cz/kontakty/2020-07-01/schéματα/kontakt.json"
  }
},
},
```

#### 9.5.1.2 Příklad přípravy datového schématu na základě podobné OFN

Jedná se pouze o vizualizaci principu jednoduchého využití připravených OFN. Prakticky není nutné s modely tímto způsobem pracovat, vše lze provést pouze na základě hrubého konceptuálního modelu / konceptu datové sady a změny zavést pouze ve strojovém vyjádření datového schématu (včetně odkazů do SSP v případě zajištění sémantiky datové sady).



Výsledkem porovnání je zjištění, že 3 vlastnosti OFN nejsou zapotřebí.

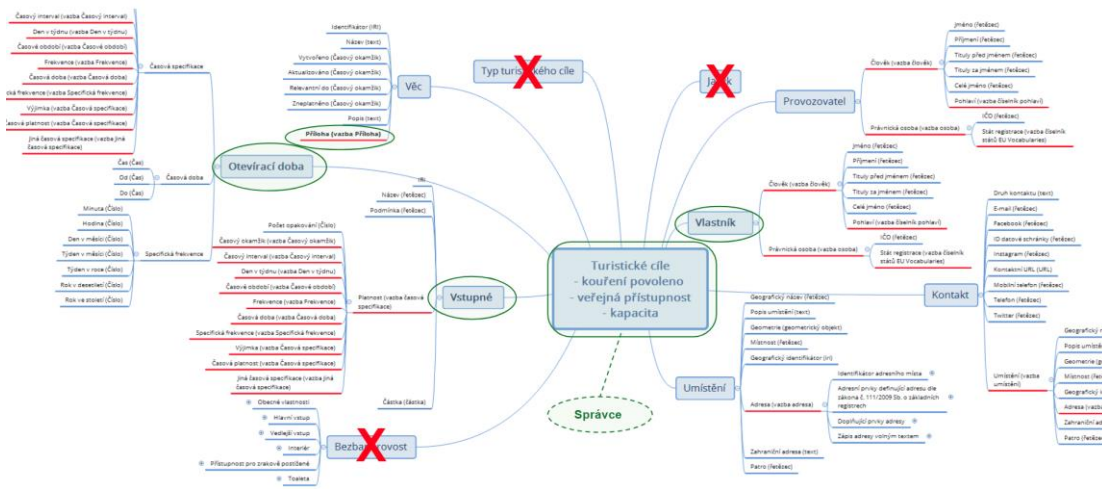
- Dvě vlastnosti OFN (jazyk a typy turistických cílů) jsou nerelevantní pro téma tržnice,
- bezbariérovost je k diskusi (i u tržnice to může být důležitá informace), ale pro jednoduchost ji není nutné řešit.

A naopak schází dvě zamýšlené položky k publikaci – prodejní místo a stánek.

Tady se nabízí několik možností řešení.

- Obě položky je možné považovat za složené základní typy (např. množství) a pouze uvádět počet prodejních míst a počet prodejních stánků.
- Obě položky řešit jako samostatné entity se svými vlastními vlastnostmi a mající vztah ke kořenové entitě tržnice (podobně jako sdílené specifikace).
- Informace o obou položkách uvést v dokumentu (nákrese, plánu, ...) a tento dokument navázat jako přílohu kořenové entity. Vybereme tuto variantu jako nejjednodušší řešení.

**Vizualizace prováděných úprav v datovém modelu:**





Červeným křížkem jsou označeny položky (případně celé sdílené specifikace), které u cílové datové sady Tržnice nebudou využity:

- Typ turistického cíle
- Jazyk
- Bezbariérovost

Zelenou elipsou jsou zvýrazněny části, které budou upraveny, případně doplněny

- Otevírací doba
- Vstupné
- Vlastník
- Atributy kořenové entity
- A nová položka – správce.

### Výsledný datový model:



- Nevužité položky jsou odstraněny, některé přejmenovány, aby významově více odpovídaly tématu Tržnice.
- Je přidána nová vazba na sdílenou specifikaci „Člověk či osoba“ s názvem „Správce“.
- Změny názvů:
  - Otevírací doba – Provozní doba
  - Vstupné – Poplatky
  - Vlastník – Zřizovatel
  - Nová položka – Správce
- U sdílené specifikace Věc je přidána jako příloha plánek tržnice, na kterém budou informace o stáncích, jejich umístění a jednotlivých prodejních místech (příklad zjednodušení uvedení chybějících atributů v původní OFN).
- Úprava strojově čitelného datového schéma.
  - Podobným způsobem s využitím stejných úprav jako u datové struktury by bylo upraveno i strojově čitelné datové schéma, a to následujícím postupem:



- Stažení datového schématu z umístění publikované OFN Turistické cíle na portálu otevřených dat.
- Provedení úprav schématu
- Odstranění nepoužitých sdílených specifikací
  - Typ turistického cíle
  - Jazyk
  - Bezbariérovost
- Přidání atributu „Správce“ s odkazem na sdílenou specifikaci „Člověk či osoba“
- Změna názvů:
  - Otevírací doba – Provozní doba
  - Vstupné – Poplatky
  - Vlastník – Zřizovatel
- Přidání odkazu na přílohu s plánkem tržnice.
- Uložení souboru s upraveným datovým schématem.
- Použití souboru při registraci datové sady.

Tématu, jakým způsobem pracovat s různými formáty a jak konkrétně provádět úpravy, se věnuje e-learningový kurz „[Technické aspekty otevřených dat](#)“.

Přizpůsobení připravených strojově čitelných datových schémat není obtížné, jedná se přejmenování datových položek a doplnění chybějících částí.

V případě publikace datové sady s částečnou vazbou do sémantického slovníku úpravy souboru s publikovaným kontextem schématu (soubory typu JSON-LD kontext v OFN) také nejsou obtížné, za předpokladu, že sémantický slovník již obsahuje pojmy odpovídající významu jednotlivých datových položek datové sady.

Pokud ve slovníku pojmy zavedeny nejsou, a datová sada má být publikována se sémantikou, je nutné provést analýzu datové sady dle „[Metodiky tvorby a údržby konceptuálních datových modelů agend](#)“ a pojmy do slovníku doplnit.

Doplňující informace včetně výkladu jsou k dispozici i v e-learningovém kurzu „[Úvod do otevřených formálních norem \(OFN\)](#)“ na Portálu otevřených dat v sekci vzdělávání. Problematice podobných OFN se věnuje modul „4. Využití připravených OFN pro tvorbu datových schémat“.

### **9.5.1.3 Příprava datového schématu s využitím sdílených specifikací.**

Příprava datového schématu je založena na:

- datovém modelu (vycházejícím z konceptuálního modelu nebo konceptu) datové sady k publikaci,
- využití prvků repozitáře OFN (sdílených specifikací a základních datových typů) postupným nahrazováním jednotlivých položek datového modelu vhodnými sdílenými specifikacemi a jejich dekompozicí na jednodušší struktury.

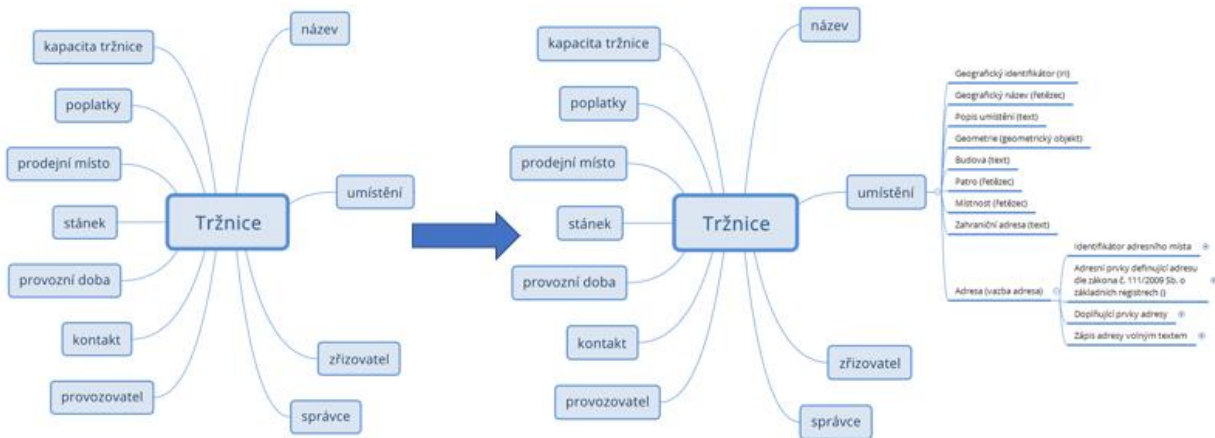
S výhodou je možné využít skutečnosti, že dekompozice sdílených specifikací je v podstatě základem jejich konstrukce, a že tato dekompozice je plně zachycena v repozitáři OFN prostřednictvím vzájemných referencí.

Samotný návrh datového schématu ze strany poskytovatele se tak zaměřuje zejména na návrh datového modelu a náhradu jeho položek prostřednictvím referenčních odkazů na sdílené specifikace a základní datové typy.

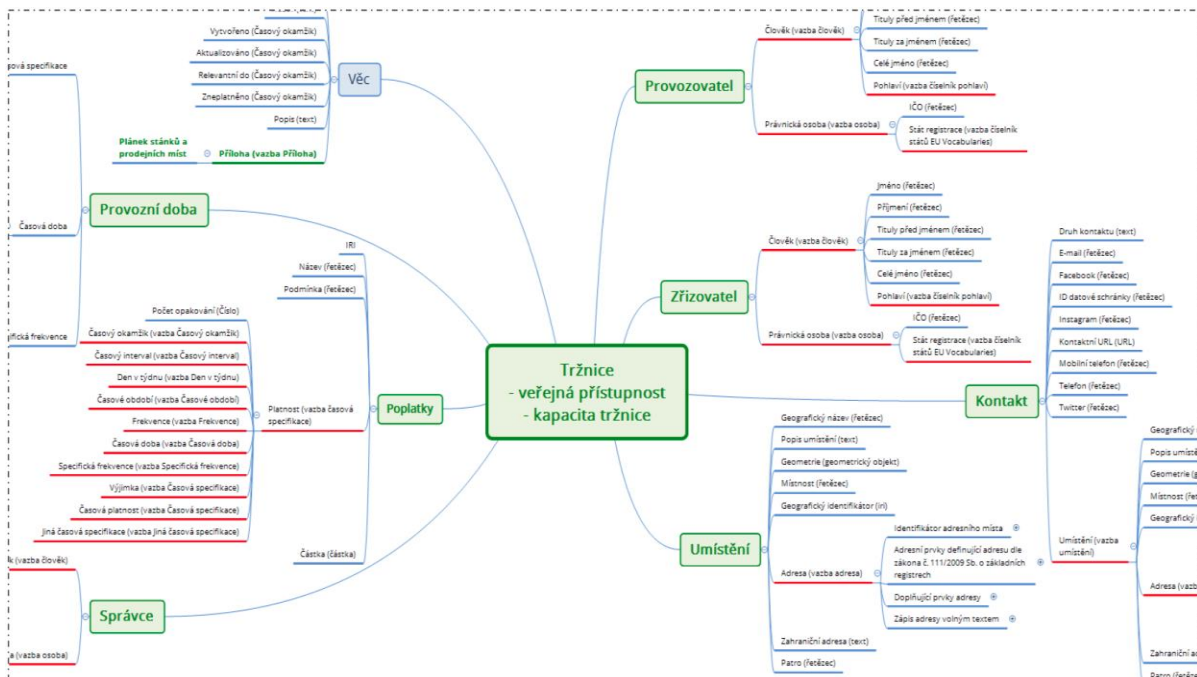
Návrh datového schématu je tak posunut na logicky vyšší úroveň, kdy není nutné řešit technické a jiné detaily a stačí se plně soustředit na zachycení věcné tematiky publikované oblasti.



## Vizualizace postupné náhrady vlastností „Tržnice“ sdílenými specifikacemi:



## Finální výsledek:

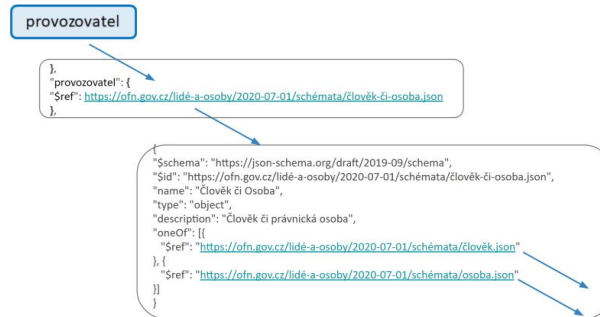


- Ze schématu je patrné, že takové kompletní datové schéma je opravdu kompletní a pokrývá všechny možné varianty datově vyjádřit položky uvedené v konceptu datové sady.
- Jedná se o maximalistickou verzi datové struktury publikované datové sady. Zcela stejný přístup, jako je tomu v případě OFN pro konkrétní tematické oblasti.
- Na rozdíl od připravených OFN v tomto případě není k dispozici připravené datové schéma ve strojově čitelném formátu, jediné, co je k dispozici jsou popisy sdílených specifikací na portálu otevřených dat.
- K vlastní tvorbě datového schématu lze přistoupit individuálním ručním výběrem reprezentací pouze použitých datových položek ve využitých sdílených specifikacích.
- Jednalo by se, ale o pracnou práci, jejíž výsledek by navíc představoval málo pružné řešení a při každé eventuální úpravě struktury publikované datové sady by bylo nutné také opět ručně upravovat i soubor s datovým schématem.
- Mnohem vhodnějším a efektivnějším řešením je plně využít vlastnosti a možnosti repositáře OFN.
- Dekompozice sdílených specifikací je v podstatě základem jejich konstrukce, a tato dekompozice je



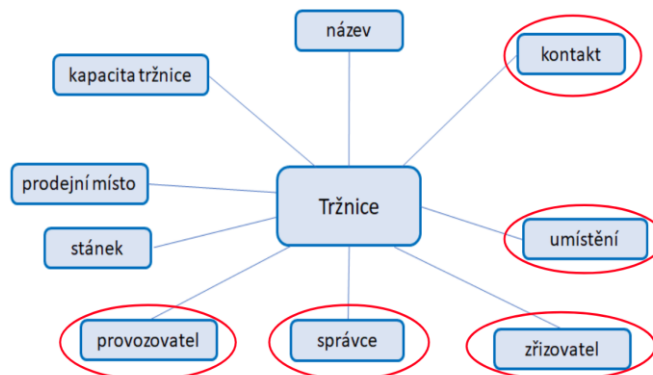
plně zachycena v repozitáři OFN prostřednictvím vzájemných referencí. Repozitář OFN si tak lze představit jako množinu hierarchických stromových modelů dat (reprezentující datová schémata a vyjádřených ve formátu JSON), které navzájem sdílí společné prvky repozitáře od těch nejjednodušších (základních datových typů) až po ty nejsložitější (specifické OFN).

- Při náhradě každé položky datového schématu vždy stačí jenom odkazovat na vrchol hierarchického datového modelu odpovídajícího ošetřované položce datového schématu (odpovídající základnímu datovému typu, nebo sdílené specifikaci).
- Příklad ošetření položky „umístění“ a její dekompozice prostřednictvím vnitřních referenčních vazeb. K využití celé sdílené specifikace potom stačí jediný referenční odkaz na tuto specifikaci



- V tomto případě položce konceptu datové sady „provozovatel“ odpovídá sdílená specifikace „člověk či osoba“. Pro vyřešení této položky stačí vazba na odpovídající sdílenou specifikaci.
- Jakým způsobem je sdílená specifikace navržena a jak je udržována její aktuálnost spadá plně na správce repozitáře. Poskytovatel je od těchto „malicherností“ odstíněn a má jistotu, že používá komponenty plně standardizované a aktuální.
- Repozitář obsahuje další dekompozice sdílených specifikací na detailnější struktury, sdílená specifikace „člověk či osoba“ je dekomponována na další dvě sdílené specifikace „člověk“ a „osoba“.
- Dalším krokem by byla dekompozice například „člověka“ na jeho jednotlivé specifikované atributy a ty by byly opět dále dekomponovány až na úroveň základních datových typů.
- V uvedeném provázání a sdílení prvků repozitáře OFN spočívá základní výhoda používání sdílených specifikací při návrhu datových schémat, vedoucí k celkovému efektivnímu využití připraveného repozitáře OFN při návrhu datového schématu.

### Vyznačené vlastnosti odpovídající sdíleným specifikacím:

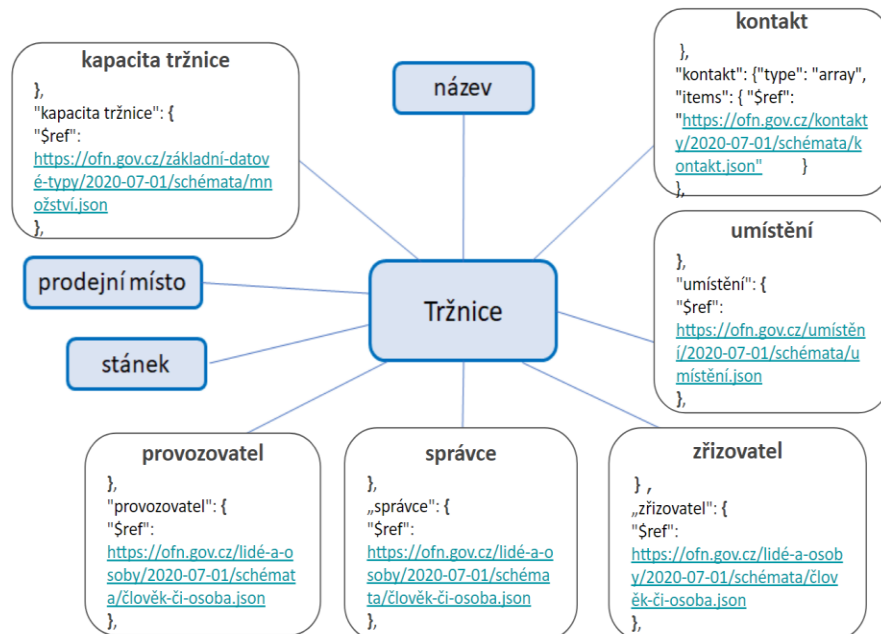


- Zjednodušený postup v uvedeném příkladu pro demonstraci:
  - Označení atributů, které je možné reprezentovat sdílenými specifikacemi.
    - Kontakt
    - Umístění





- Provozovatel
  - Správce
  - Zřizovatel
  - Kapacita tržnice
- Ošetření sdílených specifikací uvedením referenčních odkazů na jejich detailní specifikace.



- Atribut „kapacita tržnice“ je ošetřen odkazem na základní datový typ „množství“.
- Vytvoření strojově čitelného datového schématu,
- sestavení datového schématu pomocí základních datových typů a referenčních odkazů na prvky repozitáře OFN ve stejné hierarchické struktuře, jako má datový model datové sady.
- Ukázka výsledného datového schématu ve strojově čitelném formátu.



```
"kapacita_tržnice": {
  "$ref": "https://ofn.gov.cz/základní-datové-typy/2020-07-01/schéματα/množství.json"
},
"zřizovatel": {
  "type": "array",
  "items": {
    "$ref": "https://ofn.gov.cz/lidé-a-osoby/2020-07-01/schéματα/člověk-či-osoba.json"
  }
},
"provozovatel": {
  "$ref": "https://ofn.gov.cz/lidé-a-osoby/2020-07-01/schéματα/člověk-či-osoba.json"
},
"správce": {
  "$ref": "https://ofn.gov.cz/lidé-a-osoby/2020-07-01/schéματα/člověk-či-osoba.json"
},
"umístění": {
  "$ref": "https://ofn.gov.cz/umístění/2020-07-01/schéματα/umístění.json"
},
"kontakt": {
  "type": "array",
  "items": {
    "$ref": "https://ofn.gov.cz/kontakty/2020-07-01/schéματα/kontakt.json"
  }
},
}
```

Doplňující informace včetně výkladu jsou k dispozici i v e-learningovém kurzu „[Úvod do otevřených formálních norem \(OFN\)](#)“ na Portálu otevřených dat v sekci vzdělávání. Problematice využití sdílených specifikací se věnuje modul „5. Tvorba datových schémat s využitím sdílených specifikací“.

## 9.6 Propojená data

Propojená data jsou ideální způsob publikace dat na webu. Jsou inspirována způsobem fungování dnešního webu složeného z webových stránek a jsou také podepřena řadou doporučení konsorcia W3C, které k této formě vydává webové de facto standardy. Cílem publikace dat jako propojených dat je odstraňovat překážky mezi publikací a užitím dat. Je toho dosahováno několika základními způsoby:

- Globální identifikace a lokalizace datových entit
  - Žádný jiný způsob publikace dat na webu neumožňuje globální a jednoznačnou identifikaci a lokalizaci jednotlivých datových položek.
- Sémantický popis dat
  - Žádný jiný způsob publikace dat nedosahuje takové úrovně popisu významu dat jako propojená data.
- Jednotný formát
  - Propojená data používají jednotný formát reprezentace dat, což velmi usnadňuje práci s nimi.
- Integrace
  - Propojená data jsou na web vystavována již integrována s jinými datovými zdroji pomocí linků. Toto přesouvá integrační zátěž od konzumenta dat k poskytovateli, který má pro integraci lepší znalosti a zdroje.

Aby bylo možno na tyto výhody dosáhnout, je třeba dodržovat základní sadu konvencí pro propojená data. Ty jsou popsány v Otevřené formální normě „[Propojená data](#)“, její součástí jsou také vyčerpávající informace k problematice propojených dat.

[Podobný dokument](#) pro svá propojená data vydala například [standardizační organizace GS1](#).



Doplňující informace včetně výkladu jsou k dispozici i v e-learningovém kurzu „[Technické aspekty otevřených dat](#)“ na Portálu otevřených dat v sekci vzdělávání. Problematice propojených dat se věnuje modul „4. Formáty pro otevřená data – Propojená data a model RDF“.

### 9.6.1 Vytvoření RDF modelu

Pro reprezentaci propojených dat se používá datový model RDF – Resource Description Framework [[rdf11-concepts](#)]. Aktuální verze 1.1 byla vydána konsorciem W3C v roce 2014.

Jedná se o grafový datový model, tj. data jsou reprezentována jako uzly a hrany v grafu, kde uzly reprezentují entity a datové hodnoty, a hrany reprezentují jejich propojení.

Takový graf se dá popsat pomocí množiny trojic <uzel 1, hrana, uzel 2>, které říkají, že existuje entita uzel 1, existuje entita uzel 2 a jsou propojeny hranou hrana.

V RDF se jednotlivým částem každé trojice říká **subjekt**, **predikát** a **objekt**, a trojice říká, že objekt je hodnotou vlastnosti predikát nějaké entity subjekt.

Objektem může být primitivní hodnota (řetězec, datum, číslo, ...) nazývaná **literál**, nebo jiná entita. **Subjektem je vždy entita**. Entity a predikáty se identifikují pomocí jejich IRI.

Příklad trojice v RDF

- Alice je identifikována jako <https://příklad.cz/lidé/Alice>.
- Bob je identifikován jako <https://sociálnísít.cz/uživatelé/Bob>.
- Predikát říkající, že jeden člověk zná jiného má IRI <http://xmlns.com/foaf/0.1/knows>.

Trojice, která říká, že Alice zná Boba tedy vypadá takto:

- ✓ <<https://příklad.cz/lidé/Alice>>
- ✓ <<http://xmlns.com/foaf/0.1/knows>>
- ✓ <<https://sociálnísít.cz/uživatelé/Bob>>.

Tímto způsobem lze reprezentovat jakákoliv data v RDF.

### 9.6.2 Serializace RDF modelu

Datový model je nutné serializovat, tj. zapsat například do souboru, nebo v nějaké jiné formě přenést prostřednictvím Internetu. Existuje 7 standardních serializací modelu RDF.

#### N-Triples

Serializace N-Triples [[n-triples](#)] je doporučení konsorcia W3C z roku 2014. Jedná se o nejpřímočařejší RDF serializaci. IRI jednotlivých částí trojice se uzavírají do ostrých závorek < a > a trojice se ukončují tečkou. Pokud je objektem literál, uzavře se jeho hodnota do uvozovek ". IRI datového typu literálu se připojuje za znaky ^^

Soubor s RDF serializací N-Triples má příponu .nt. Tato serializace se hodí tam, kde záleží na jednoduchosti a rychlosti zpracování a nezáleží tolik na lidské čitelnosti a na velikosti dat.

#### Turtle

Serializace Turtle [[turtle](#)] je doporučení konsorcia W3C z roku 2014. V serializaci N-Triples se řada IRI nebo jejich částí neustále opakuje, navíc tato serializace není příliš čitelná pro lidi. Serializace Turtle proto přidává optimalizace vedoucí k lepší lidské čitelnosti a úspoře počtu znaků.

- Znak středník ; říká, že následující trojice má stejný subjekt jako aktuální, a tedy stačí specifikovat pouze nový predikát a objekt.



- Znak čárka , říká, že následující trojice má stejný jak subjekt, tak predikát a tedy stačí specifikovat pouze nový objekt.
- Pro případy opakujících se některých části IRI (například ty, které identifikují predikáty z jednoho slovníku, nebo entity patřící do stejného nadřazeného prvku), serializace Turtle zavádí tzv. prefix - krátce pojmenovaná opakující se část IRI, kterou lze použít pro zkrácený zápis IRI v dokumentu.

Soubor s RDF serializací Turtle má příponu .ttl.

### N-Quads

Serializace N-Quads [[n-quads](#)] je doporučení konsorcia W3C z roku 2014 a rozšiřuje serializaci N-Triples o čtvrtou složku, IRI pojmenovaného grafu, do kterého daná trojice patří.

Soubor s RDF serializací N-Quads má příponu .nq.

### TriG

Serializace TriG [[trig](#)] je doporučení konsorcia W3C z roku 2014. Stejně jako serializace Turtle zavádí zkratky do serializace N-Triples, serializace TriG stejným způsobem rozšiřuje serializaci N-Quads. Jedná se tedy zároveň o rozšíření serializace Turtle o podporu pojmenovaných grafů.

Soubor s RDF serializací TriG má příponu .trig.

### JSON-LD

Serializace JSON-LD [[json-ld](#)] je doporučení konsorcia W3C z roku 2014. Jedná se o serializaci RDF v syntaxi JavaScript Object Notation (JSON) [[ECMA-404](#)], která je již známá velkému počtu stávajících vývojářů webových aplikací. Hlavní motivací pro tuto serializaci je možnost poskytovat jedním způsobem propojená data jak vývojářům, kteří znají pouze JSON, tak vývojářům znalým technologií propojených dat. Mapování hodnot z JSON do RDF je specifikováno v klíči @context, který mohou JSON vývojáři ignorovat.

Soubor s RDF serializací JSON-LD má příponu .jsonld.

### RDF/XML

Serializace RDF/XML [[rdf-syntax-grammar](#)] je doporučení konsorcia W3C z roku 2014 a specifikuje, jak zapsat datový model RDF do XML dokumentů. Historicky se jedná o první RDF serializaci, nicméně aktuálně ztrácí na významu, jelikož je lidsky poměrně nečitelná, přidává složitost pravidel pro XML dokumenty. Zpracování softwarem, který rozumí XML ale nerozumí RDF se již nepředpokládá.

Soubor s RDF serializací RDF/XML má příponu .rdf.

### RDFa

Serializace RDFa [[rdfa-core](#)] je doporučení konsorcia W3C z roku 2015 a specifikuje, jak lze datový model RDF zapsat do atributů běžných HTML dokumentů [[html53](#)]. Serializace nalezne uplatnění tam kde je požadováno, aby lidsky čitelný zápis v HTML byl doplněn o strojově čitelné informace ve stejném dokumentu, s minimalizací opakování stejných dat pro obě reprezentace.

## 9.6.3 Pravidla pro výběr slovníků pro vybrané typy dat

Nedílnou součástí propojených dat je také standardizace jejich obsahu pomocí tzv. **slovníků** – popisů významu jednotlivých tříd a predikátů pro daný typ dat.

Obecným pravidlem pro tvorbu reprezentace propojených dat je pokrývání jejich obsahu již existujícími slovníky. Až teprve v případě, že ještě neexistuje vhodný slovník pro popis nějaké části publikovaných dat, je možné si dodefinovat svůj slovník, který je ale nutné řádně publikovat, aby ho mohli použít ostatní poskytovatelé dat. K



vyhledání existujících slovníků je určen registr [Linked Open Vocabularies \(LOV\)](#).

V prostředí veřejné správy ČR je pro tvorbu reprezentace propojených dat a pokrývání jejich obsahu výchozím slovníkem „[Sémantický slovník veřejné správy](#)“ (SSP). Pro popis těch částí publikovaných dat, pro které ještě neexistuje vhodný slovník, je možné si také dodefinovat svůj vlastní slovník, který je ale nutné řádně publikovat, aby ho mohli použít nejen ostatní poskytovatelé dat ale také jejich uživatelé. Důvodem ustavení „[Sémantického slovníku veřejné správy](#)“ jako základu pro pokrývání obsahu datových sad je skutečnost, že slovník je silně navázán na legislativu ČR (vychází z pojmů a jejich definic v legislativě), stejně jako správa a publikace dat VS ve formátu otevřených dat (více v kapitole „[Popis sémantiky veřejné správy](#)“).

Nejvíce používané typy dat a slovníky dostupné na webu pro reprezentaci propojených dat:

- Číselníky - [Simple Knowledge Organization System](#) (SKOS)
  - Číselníkem je typicky plochý seznam položek, kde každá má minimálně kód a název. Číselníky se mohou používat například jako seznam možných hodnot pro různé vlastnosti datových entit, což lze využít pro formuláře pro zadávání dat, aplikace vizualizující data apod. Pro reprezentaci číselníků v propojených datech se používá slovník **SKOS**.
- Statistická data - [The RDF Data Cube Vocabulary](#)
  - The RDF Data Cube Vocabulary (DCV) [[vocab-data-cube](#)] je doporučení W3C z roku 2014 pro reprezentaci datových kostek v datovém modelu RDF. Datový model DCV je kompatibilní s datovým modelem [SDMX](#).
- RDF Schema
  - RDF Schema 1.1 [[rdf-schema](#)] je doporučení W3C z roku 2014 a slouží k jednoduchému popisu tříd a vlastností a také k tvorbě jednoduchých hierarchií dědičnosti tříd a vlastností. Ke každé třídě a vlastnosti je možno specifikovat její název a popis, pro vlastnosti pak navíc definiční obor a obor hodnot.
- OWL
  - Web Ontology Language 2 [[owl2-overview](#)] je doporučení W3C z roku 2012 pro tvorbu ontologií. Umožňuje modelovat i velmi složité sémantické vztahy. **V prostředí propojených dat se nejvíce používá predikát owl:sameAs, který říká, že 2 IRI identifikují stejnou entitu reálného světa.**

#### 9.6.4 Zpřístupnění publikovaných propojených dat

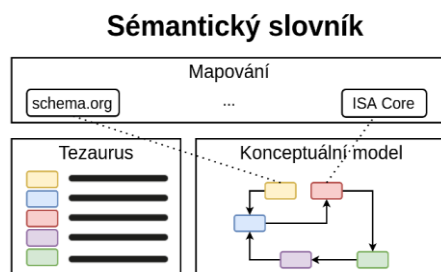
- Dostupnost souborů ke stažení
  - Základním způsobem zveřejňování 5\* otevřených dat je jejich poskytnutí ve formě RDF dumpu, tj. souboru ke stažení v jedné ze standardních RDF serializací, tj. Turtle, TriG či JSON-LD. Volitelně může být použita komprese gzip.
- Dereference IRI – Již ze základních principů propojených dat vychází potřeba zpřístupňovat data o konkrétní entitě při přístupu na IRI této entity, tj. jeho využití jako URL. Data o datové entitě jsou poskytnuta ve formátu RDF. Tato funkcionality může, ale nemusí využívat SPARQL endpoint, který opět může, ale nemusí být zpřístupněn veřejnosti.
- Dostupnost SPARQL endpointu
  - Publikace 5\* otevřených dat v podobě veřejného SPARQL endpointu může být velmi náročná na hardwarové prostředky, a není bezpodmínečně nutná, uživatelé 5\* otevřených dat si mohou stáhnout RDF dump a použít pro dotazování vlastní SPARQL endpoint. Sémantika údajů veřejné správy



## 9.7 Popis sémantiky veřejné správy

Sémantický slovník pojmů veřejné správy je nástrojem pro postupnou harmonizaci významu (sémantiky) jejich dat. Je katalogem pojmosloví používaného v rámci veřejné správy. Zahrnuje pojmy, jejich definice, vazby pojmů na legislativu, vzájemné významové vazby pojmů mezi sebou i významové vazby pojmů na standardní veřejné slovníky používané v zahraničí (především z iniciativy EU, např. ISA Core Vocabularies). Jako takový je prostředkem usnadňujícím zajišťování syntaktické a sémantické interoperability dat veřejné správy.

Sémantický slovník pojmů je tvořen tzv. tezaurem (glosářem), který definuje důležité pojmy v dané oblasti zájmu (např. pojmy zavedené nějakým zákonem nebo pojmy používané v nějaké agendě, datové sadě, ...) s **hierarchickým uspořádáním jednotlivých pojmů od nejabstraktnějších po nejkonkrétnější**, a konceptuálním modelem, který pojmy propojuje pomocí významových (sémantických) souvislostí.



### 9.7.1 Slovník pojmů (Glosář, Tezaurus).

Základním stavebním kamenem glosáře je pojem. Pojem (též koncept, nebo také význam) je slovo nebo sousloví opatřené definicí, synonymy, zdrojem, příp. dalšími charakteristikami, které zpřesňují jeho význam tak, aby se omezila možnost mu špatně porozumět nebo jej špatně použít.

Základní pojmy glosáře jsou tvořeny pojmy provozovaných agend veřejné správy. Ty jsou odvozené z legislativních dokumentů stejně jako přesné definice jejich významů. Ne všechny pojmy agend jsou ale explicitně uváděny v legislativě, proto je nutné k zajištění sémantické interoperability dat veřejné správy glosář pojmů průběžně doplňovat o další identifikované pojmy (a jejich významy), a tedy i během publikace otevřených dat.

Pojmy reprezentují:

- subjekty a objekty v reálném nebo abstraktním světě,
- vlastnosti těchto subjektů a objektů (atributy a vztahy),
- osoby, které jsou subjekty práv nebo povinností v agendě,
- věci, které jsou předmětem (objektem) práv a povinností v agendě,
- role subjektů (role objektů se neuvažují, neboť nemohou konat a „brát na sebe“ roli),
- fáze subjektů nebo objektů, tj. jejich konkrétní specifické stavy,
- události, které jsou pro danou doménu relevantní, a které nějakým způsobem ovlivňují subjekty nebo objekty,
- činnosti (konání, chování, úkony) v rámci domény,

Konzistence pojmů a jejich hierarchické uspořádání od nejabstraktnějších po nejkonkrétnější v glosáři jsou zajištěny:

- klasifikací subjektů a objektů dle různých společných vlastností (zatřídění),
  - generalizací a specializací mezi těmito klasifikačními třídami (zachycení specializační vazbou fakt, že určitý typ subjektu, nebo objektu je speciálnější než jiný typ, je tedy jeho speciálním případem

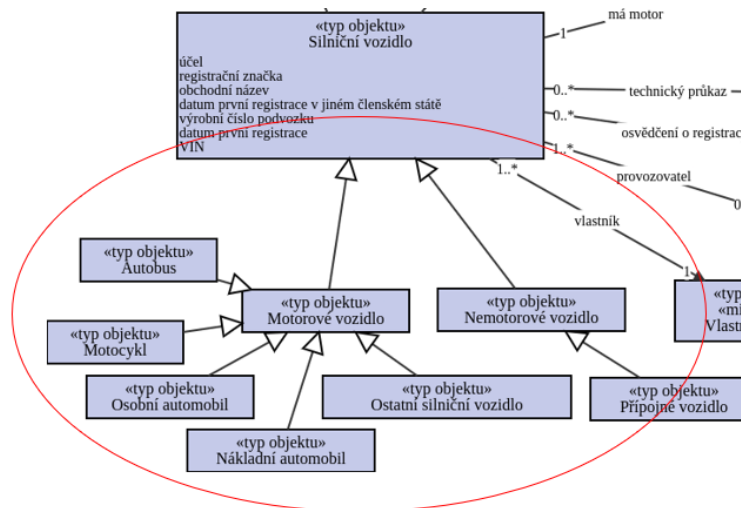


a každý prvek speciálního prvku je také prvkem specializovaného).

### 9.7.2 Konceptuální model.

- Konceptuální model je zcela obecný a základní pojmový model, nezávislý na jakékoliv potenciální implementaci jeho obsahu.
- Konceptuální model je vyjádřen v grafické podobě diagramu.
- Vzniká na základě pojmů definovaných slovníkem a tyto pojmy propojuje významovými vazbami (vztahy), které zachycují klíčové vazby v analyzované doméně.
- Doplnuje význam pojmů zachycených v glosáři o kontexty, ve kterých se modelované pojmy v analyzované doméně vyskytují.
- Věcné (sémantické) souvislosti mezi pojmy doplňuje následujícím způsobem:
  - klasifikuje pojmy glosáře do typů (např. „typ subjektu“, ...),
  - přiřazuje typy vlastností k ostatním typům,
  - přiřazuje typy vztahů k typům, které se vztahu účastní,
  - přiřazuje typy událostí a činností k typům, které se jich týkají,
  - propojuje typy pomocí obecných pojmenovaných vzájemných asociací k zachycení souvislostí, které nejsou takového významu, aby byly modelovány jako typy vztahů,
  - určuje specializační vazby, které modelují, že určitý typ je speciálnější než jiný typ,
  - přiřazuje typy subjektů k typům rolí.

Ukázka specializace (hierarchie) pojmů.



### 9.7.3 Sémantický slovník

Komplet slovníku pojmů (s jejich významy) a souvisejícího konceptuálního modelu (s vazbami mezi pojmy a zachycenými kontexty) představuje **sémantický slovník**.

Není přitom nutné nejprve vytvořit úplný glosář a teprve potom přistoupit k tvorbě konceptuálního modelu. Glosář lze opakovaně rozšiřovat o nové pojmy.

**Sémantický slovník pojmů veřejné správy (SSP)** je tvořen všemi slovníky pojmů a všemi konceptuálními modely veřejné správy. Jeho prohlížení, vytváření a využití je dostupné prostřednictvím odkazu <https://slovník.gov.cz>.

Struktura slovníků vzniká postupně (inkrementálním způsobem) s každou analyzovanou agendou, nebo její částí, věcnou doménou a také s každou publikovanou datovou sadou. Což znamená, že například pro každou agendu



(analyzovanou věcnou doménu, datovou sadu) musí být vytvořen vlastní sémantický slovník agendy, který sestává ze sémantických slovníků jednotlivých AIS, které agendu podporují, a propojuje je do jednoho uceleného slovníku agendy. Sémantické slovníky agend jsou pak propojeny do sémantického slovníku úřadu/resortu. Sémantické slovníky úřadů/resortů společně tvoří propojený sémantický slovník pojmů celé veřejné správy.

#### Struktura sémantického slovníku:

- Základní slovník (Z-SGoV)
  - Základní sémantický slovník pojmů definuje modelovací jazyk, který je používán pro tvorbu konceptuálních modelů a především součástí Sémantického slovníku pojmů.
- Slovník veřejného sektoru (V-SGoV)
  - Sémantický slovník pojmů veřejného sektoru (V-SGoV) zavádí konsensuální pojmosloví používané napříč veřejnou správou zejména v oblasti základních pojmů teorie práva, struktury dat a jejich evidence.
- Doménové slovníky
  - Doménový sémantický slovník může např. definovat pojmy z daného zákona, rozšiřovat sémantický slovník pojmů daného zákona o pojmy používané v dané agendě, nebo zavést technické pojmy používané v kontextu daného AIS nebo datové sady v samostatném slovníku.
  - Při tvorbě doménových sémantických slovníků je nutné postupovat podle Metodiky tvorby a údržby sémantických slovníků veřejné správy.
  - Výchozí sémantické slovníky:
    - **generický sémantický slovník pojmů veřejného sektoru** – zavádí konsensuální pojmosloví používané napříč veřejnou správou zejména v oblasti základních pojmů teorie práva, struktury dat a jejich evidence,
      - např.: subjekt práva, objekt práva, právní vztah, číselník, evidenční systém,
    - **sémantický slovník pojmů základních registrů** – zavádí pojmosloví základních registrů, které není specifické pro jeden konkrétní základní registr,
      - např.: osoba, fyzická osoba,
    - **sémantický slovník pojmů registru osob** – zavádí pojmosloví specifické pro registr osob,
      - např.: osoba v registru osob, právnická osoba, podnikající fyzická osoba,
    - **sémantický slovník pojmů registru obyvatel** – zavádí pojmosloví specifické pro registr obyvatel,
      - např.: subjekt údajů vedený v registru obyvatel,
    - **sémantický slovník pojmů registru práv a povinností** – zavádí pojmosloví specifické pro registr práv a povinností,
      - např.: orgán veřejné moci, soukromoprávní uživatel údajů, agenda, činnost, služba,
    - **sémantický slovník pojmů registru územních identifikátorů, adres a nemovitostí** – zavádí pojmosloví specifické pro registr územních identifikátorů, adres a nemovitostí,
      - např.: územní prvek.

#### Význam a výhody sémantického slovníku:

- zajišťuje podmínky:
  - pro sdílení významu pojmů v rámci veřejné správy ČR,
  - automatické vytváření výkladových slovníků,





- k zajištění sémantické interoperability agendových informačních systémů v rámci propojeného datového fondu (PPDF) veřejné správy a veřejného datového fondu (VDF),
- k zajištění interoperability publikovaných otevřených dat veřejné správy,
- zvyšuje kvalitu popisu otevřených datových sad:
  - přispívá k omezení duplicit v pojmenování stejných prvků různými jmény, nebo různých prvků stejnými jmény,
  - umožňuje propojování souvisejících atributů dvou (dosud nepropojených) datových sad významovou vazbou,
- zjednodušuje přípravu a návrh publikovaných otevřených dat,
  - sjednocuje způsob modelování jednotlivých datových sad,
  - zjednodušuje modelování jednotlivých datových sad prostřednictvím již vytvořených konceptuálních modelů v sémantickém slovníku,
  - podporuje tvorbu datových schémat založenou na používání Otevřených formálních norem,
  - automatické generování datových schémat otevřených dat,
- zajišťuje podmínky pro nové funkcionality NKOD založené na významovém propojení obsahu publikovaných otevřených dat (vyhledávání vzájemných významových souvislostí mezi datovými sadami, schémata a rozhraními),
- zjednodušuje evidenci údajů subjektů a objektů práva při ohlašování agendy v Registru práv a povinností, a to včetně přesné specifikace významu s odkazem na legislativu, a také identifikováním a zachycením významových vazeb na další pojmy.

#### 9.7.4 Analýza agendy a odvození údajů

V rámci analýzy je nutné identifikovat a popsat nejen samotné údaje potřebné v agendě, ale také subjekty práva, které v agendě vystupují v agendových rolích (kontextech), a objekty práva, které jsou předmětem práv nebo povinností těchto subjektů práva. Identifikované subjekty práva ve svých agendových rolích, objekty práva a údaje potřebné v agendě o těchto subjektech a objektech práva pak musí ohlašovatel agendy, s případnou podporou správce agendového informačního systému, ohlásit do registru práv a povinností (RPP).

V rámci analýzy je vytvořen tzv. sémantický slovník pojmů agendy. Pro každou agendu musí být vytvořen vlastní sémantický slovník agendy, který sestává ze sémantických slovníků jednotlivých AIS, které agendu podporují, a propojuje je do jednoho uceleného slovníku agendy.

Jako vstupy pro analýzu slouží legislativní předpisy (tj. zákony, prováděcí vyhlášky, nařízení atd.) definující agendu a AIS, které agendu podporují.

Výstupem analýzy je sémantický slovník pojmů agendy, ve kterém jsou identifikovány subjekty a objekty práva, jejich agendové role a agendové údaje připravené pro evidenci v registru práv a povinností.

Celá problematika analýzy agendy je detailně vysvětlena v [doplňujících zdrojích](#).



### 9.7.5 Doplnění entit, atributů a vazeb mezi entitami dle datové sady

Zařazení (specializace) nového pojmu (entity – subjektu/objektu).

Uvedený postup je uveden pouze pro přiblížení problematiky, reálné postupy se budou záviset na výsledku zavedení modelovacích nástrojů do běžného užívání. Aktuálně jsou k dispozici pouze prototypy nástrojů, která vyžadují absolvování k tomu určených školení.

- Postupné procházení hierarchické struktury SSP až do nalezení významově nejbližšího pojmu.
- Specifikace názvu nového pojmu dle Metodiky.
- Zavedení (specializace) nového pojmu v Termitu.
- Vyplnění všech potřebných souvisejících náležitostí.
- Zobrazení všech zděděných atributů a vazeb zavedeného pojmu (zobrazení kontextu pojmu).
- Doplnění nových atributů a jejich specifikací.
- Doplnění nových vazeb.

### 9.7.6 Správa Sémantického slovníku veřejné správy.

[Správcem Úložiště konceptuálních modelů je DIA](#), která je také správcem a věcným garantem ontologie veřejné správy. Konceptuální modely kontextů vypracovávají ohlašovatelé agend a ukládají je do Úložiště konceptuálních modelů. DIA, Odbor hlavního architekta eGovernmentu pro tvorbu konceptuálních modelů poskytuje volně dostupný nástroj pro konceptuální modelování kontextů a také udržuje Metodiku konceptuálního modelování kontextů. Kontroluje požadovanou kvalitu konceptuálních modelů dodaných do Úložiště konceptuálních modelů. jejich vazeb na evidenci definic údajů v RPP, v případě zjištěných nedostatků vyžaduje jejich nápravu ze strany ohlašovatelů agend.

### 9.7.7 Podpora nástrojů

Sémantické slovníky lze realizovat dvěma základními způsoby:

- Nezávisle na existenci a správě Sémantického slovníku pojmů veřejné správy (SSP), což znamená bez naplnění cíle harmonizace významů pojmů veřejné správy, a bez zajištění sémantické interoperability otevřených dat, a to:
  - stylem „tužka/papír“, tedy bez jakýchkoliv technických prostředků, nebo
  - s pomocí libovolného nástroje pro konceptuální datové modelování (např. volně dostupný PlantUML nebo komerční Enterprise Architect).
  - V případě takto připravených slovníků pojmů je nutné, aby poskytovatel datových sad je veřejně zpřístupnil k zajištění sémantiky publikovaných dat.
  - Pojmy obsažené v těchto slovnících ale nebudou součástí (ani konzistentní) se slovníkem pojmů veřejné správy SSP.
- V plně integrovaném a konzistentním prostředí sémantického slovníku pojmů veřejné správy (SSP). V tomto případě je základní sémantický slovník pojmů SSP postupně rozšiřován a doplňován o nové pojmy s každým novým konceptuálním modelem publikované datové sady nebo analyzované agendy. Stávající pojmy jsou sdíleny, a navíc případně uváděny do nových souvislostí (kontextů) vyplývajících z každé nově modelované oblasti.



- Tento způsob tvorby sémantického slovníku je již ale nutné provádět s vhodnou podporou k tomu určených a volně dostupných nástrojů (zatím ve verzí prototypů), které jsou připraveny agenturou DIA.
- Nástroje podporují postupy uvedené v příslušných metodikách a umožňují vytvářet sémantický slovník pojmů v podobě, která je dále strojově zpracovatelná.
- Připravené nástroje:
  - **Termit** – tvorba slovníků pojmů, nástroj pro správu odborného pojmosloví a jeho propojení na legislativu,
  - **Ontographer** – tvorba konceptuálních modelů, nástroj pro správu konceptuálních modelů využívajících vytvořené pojmosloví,
  - **Dataspecer** – nástroj pro automatizovanou tvorbu datových specifikací a OFN přímo z vytvořených konceptuálních modelů,
  - **ShowIT** – webový prohlížeč a vyhledávač pojmů v sémantickém slovníku pojmů (SSP).
- Aktuálně jsou k dispozici pouze zcela funkční a použitelné prototypy nástrojů, ale pro práci s nimi je vhodné absolvovat příslušná školení a seznámení se s [Metodikou tvorby a údržby sémantického slovníku veřejné správy](#).
- Po zavedení prototypů nástrojů do běžného užívání budou připraveny ze strany DIA návody a doporučení pro všechny připravené nástroje.

### 9.7.8 Doplnující zdroje

- [Koncepte sémantického slovníku pojmů pro potřeby konceptuálního datového modelování agend.](#)
- [Metodika tvorby a údržby sémantického slovníku veřejné správy.](#)
- [Metodika definice údajů vedených v agendě.](#)
- [Definice scénářů využití konceptuálních modelů.](#)
- [Návrh a prototypování výrobní linky pro tvorbu a údržbu konceptuálních modelů agend.](#)
- [Ověření prototypu výrobní linky v podobě vytvoření konceptuálních modelů vybraných agend.](#)
- [Analýza existujících informačních systémů veřejné správy vč. jejich příslušné legislativy a identifikace datových sad pro veřejný datový fond. Příloha.](#)

## 9.8 Katalogizační záznam.

### 9.8.1 Metadata záznamu datové sady a pravidla pro jejich vyplňování.

Příklad katalogizačního záznamu datové sady v NKOD



### Seznam jednotlivých položek katalogizačního záznamu:

- **Název datové sady** (povinný atribut).
  - Krátký název datové sady, ze kterého je zřejmé, co je obsahem datové sady.
  - V názvu je vhodné použít i název publikující organizace.
  - Název by měl být optimálně (max.) v rozsahu jednoho řádku textu.
- **Popis datové sady** (povinný atribut).
  - Delší text, který detailněji popisuje, co je obsahem datové sady.
  - Pokud je pro správné pochopení nutné, je vhodné uvést i o čem data nevyprávějí.
  - Optimální velikost textu max. v rozsahu jednoho odstavce do 10 řádků.
- **Téma datové sady** (povinný atribut).
  - Povinné: Výběr [tématu z evropského číselníku datových témat](#), které datovou sadu nejlépe vystihuje.
  - Volitelně také pojmy ze sémantického slovníku pojmů
- **Periodicita aktualizace** (povinný atribut).
  - Výběr periodicity aktualizace datových zdrojů (publikace nových nebo aktualizace obsahu již publikovaných distribucí). Např. při výběru týdenní periodicity to znamená, že každý týden bude přidán nový datový zdroj nebo aktualizován obsah již publikovaného datového zdroje.
  - Periodicita aktualizace se vyplňuje dle [EU slovníku pro frekvence](#). Příklady možností:
    - Denně
    - Týdně
    - Měsíčně
    - Čtvrtletně
    - Pololetně
    - Ročně
    - Nikdy / jednorázová publikace
    - Nepravidelně
- **Klíčová slova** (povinný atribut).
  - Jedno nebo více libovolných klíčových slov, které charakterizují obsah datové sady.
- **Související geografické území** (povinný atribut).
  - Výběr geografického území, kterého se datová sada týká. Může to být např.:
    - území, které by datová sada pokrývala, pokud by byla zobrazena na mapě,



- správní území orgánu veřejné správy, který datovou sadu publikuje (v případě, že by zobrazení na mapě nedávalo smysl).
- K identifikaci geografického území lze využít identifikátory RÚIAN, např.:
  - Datová sada s veřejnými parkovišti na území obce má jako související geografickou oblast přiřazenou obec (pokud by se parkoviště zobrazily na mapě, rozprostíraly by se na území obce).
  - Datová sada s rozpočtem krajského úřadu má jako související geografickou oblast přiřazeno území vyššího samosprávného celku, tj. kraje (rozpočet nedává smysl zobrazovat na mapě, vhodnější formou je správní území kraje).
- K identifikaci se používá se dvojice hodnot:
  - typ prvku RÚIAN (Území obce, Území státu),
  - kód prvku RÚIAN (Praha (554782), Česká republika (1)).
- *Dotčené časové období* (doporučený atribut).
  - Časové období (datum od, případně do), které je pokryto obsahem datové sady.
  - Vyplňuje se pouze, pokud to dává 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.
- *Jméno Kurátora dat* (doporučený atribut).
  - Jméno konkrétní osoby nebo název útvaru organizace, který je Kurátorem dat pro danou datovou sadu.
- *Email Kurátora dat* (doporučený atribut).
  - Kontaktní email Kurátora dat, může být i jednotný pro celou organizaci.
- *Odkaz na dokumentaci datové sady* (doporučený atribut).
  - Odkaz na dokument, který je určený pro použití lidmi, a který popisuje obsah datové sady a také vysvětluje jeho správné použití/interpretaci.
- *Klasifikace dle EUROVOC* (doporučený atribut).
  - Vybrané koncepty z taxonomie EUROVOC. [Tezaurus EUROVOC](#) obsahuje koncepty, jejichž názvy (pojmy, termíny) jsou udržovány ve všech jazycích Evropské unie, a je proto vhodný pro klasifikaci datových sad.
  - Měl by být vybrán vždy alespoň jeden koncept charakterizující obsah datové sady.
  - [Způsob vyhledání a použití EUROVOC konceptu](#) je uveden na POD.
- *Odkaz na specifikaci*.
  - Odkaz na specifikaci, jíž se datová sada řídí. Takovou specifikací jsou zejména Otevřené formální normy.



## 9.8.2 Metadata záznamu distribuce datové sady

Distribuce datové sady

Annotations in the image:

- Název
- Odkaz přístupový bod služby
- Podmínky užití
- Formát souboru
- Odkaz na soubor ke stažení
- Odkaz na schéma souboru ke stažení
- Media typ

- **Podmínky užití** (povinný atribut).
  - Vyplněné podmínky užití dle Stanovení podmínek užití otevřených dat.
- Odkaz na stažení datového zdroje (povinný atribut).
  - Odkaz na datový soubor ke stažení.
  - V případě, že distribuce datové sady je zpřístupněna datovou službou, odkaz na přístupový bod.
- Formát souboru ke stažení (povinný atribut).
  - Uvedený formát souboru ke stažení dle evropského číselníku typů souborů.
- Media type souboru ke stažení (povinný atribut).
  - Uvedený media typ popisující formát datového zdroje.
- Odkaz na strojově čitelné datové schéma datového zdroje (doporučený atribut).
  - Odkaz na strojově čitelné datové schéma popisující strukturu datového zdroje.
- **Název distribuce** (doporučený atribut).
  - Název datového souboru ke stažení nebo název API.
- Poskytovatel – název organizace, která datovou sadu publikuje.

## 9.8.3 Metadata záznamu datové série

Častou chybou při návrhu katalogizačního záznamu datové sady je chybné dělení distribucí.

Jednotlivé distribuce jedné datové sady **se od sebe totiž mohou lišit pouze formátem**, například JSON-LD a CSV, případně distribuce mohou popisovat různé datové služby, přes které je datová sada přístupná.

**Nelze však jako distribuce jedné datové sady mít soubory za různá časová období**, jako třeba rozpočty obcí, nebo za různé obce, například turistické cíle v okrese. To jsou totiž různé datové sady.

Ty však mohou být sdružené do datových sérií.

- Datové série umožňují věcně související datové sady vidět na jednom místě.



- Pro datové série se použije část záznamu pro datovou sadu, ale datová série samotná nemá žádné distribuce.
- Skládá se pouze z dalších datových sad.
- V detailu datové sady lze potom vidět, že je součástí datové série.
- Pro využití funkcionality datových sérií je třeba používat Lokální katalog otevřených dat (LKOD)

The screenshot shows the ARES - Administrativní registr ekonomických subjektů website. The left sidebar has a section 'Datové sady této datové série' with a red box and arrow pointing to 'ARES - Číselník veřejných rejstříků'. The main content area has a section 'Součást datové série' with a red box and arrow pointing to the same dataset. Below, there is a 'Distribuce datové sady' section with a table of distribution options.

Podmínky užití distribuce	Soubor ke stažení
Neobsahuje Autorské dílo	Stáhnout
Neobsahuje Originální databáze	text/csv
Není chráněna Zvláštní právo	Typ média

## 9.9 Lokální katalog otevřených dat (LKOD)

Pokud Poskytovatel provozuje lokální katalog otevřených dat s rozhraním dle [Otevřené formální normy pro datové katalogy](#), 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 registrační soubor jako přílohu datové zprávy s předmětem „NKOD“.

Správa záznamu lokálního katalogu v NKOD je prováděna zasíláním registračních dokumentů ve formě přílohy datové zprávy s předmětem „NKOD“ do datové schránky s identifikátorem m3hp53v „Národní katalog otevřených dat (NKOD)“ spravované Digitální a informační agenturou. Tyto dokumenty lze získat pomocí formulářů dostupných v NKOD přes odkaz „Registrovat nový lokální katalog“ v patičce NKOD, či pomocí symbolu mazání na detailu datové sady v případě mazání lokálního katalogu. Vždy je nutné zasílat pouze jednu přílohu jednou datovou zprávou, jinak bude zpráva ignorována. Dokumenty jsou ve formátu JSON-LD (s příponou [.jsonld.txt](#) kvůli omezení na typy příloh datových zpráv).

### 9.9.1 Implementace Lokálního katalogu otevřených dat

Pro implementaci Lokálního katalogu otevřených dat nejsou stanoveny žádné doporučené nebo přímo vyžadované sw řešení. Prakticky to znamená, že z pohledu garanta otevřených dat veřejné správy ČR jsou přípustná jakákoliv řešení poskytovatelů, za předpokladu, že budou mít implementované referenční rozhraní pro komunikaci s NKOD dle předepsané Otevřené formální normy „[Rozhraní katalogů otevřených dat](#)“.

Příklad implementace jednoduchého katalogu otevřených dat (LKOD), který plně vyhovuje požadavkům NKOD, lze nalézt prostřednictvím odkazu [Referenční implementace Lokálního katalogu otevřených dat \(LKOD\)](#).



## 9.9.2 Registrace lokálního katalogu

Registrace lokálního katalogu je provedena zasláním registračního dokumentu, který lze získat vyplněním formuláře po kliknutí na odkaz „Registrovat nový lokální katalog“ v patičce NKOD. Postup registrace lokálního katalogu v NKOD je následující:

- Otevření [registračního formuláře](#).
- Zadání všech požadovaných údajů ve formuláři.
  - Název lokálního katalogu otevřených dat (povinný atribut).
    - Tento název rozlišuje mezi jednotlivými katalogy otevřených dat dané instituce. Pokud má jeden, může to být třeba 'Katalog otevřených dat XXX'.
  - Jméno a email správce katalogu (povinné atributy).
    - Jméno a email správce lokálního katalogu otevřených dat, se kterým lze komunikovat o případných problémech s katalogem (např. problém s rozhraním API, jeho přístupností, apod).
  - Typ API lokálního katalogu (povinný atribut).
    - Označuje typ API (rozhraní) registrovaného katalogu. Podrobnosti viz Otevřená formální norma [Rozhraní katalogů otevřených dat](#).
  - URL API lokálního katalogu (povinný atribut).
    - URL API katalogu dle Otevřené formální normy [Rozhraní katalogů otevřených dat](#).
  - Domácí stránka katalogu (volitelný atribut).
    - URL domácí stránky lokálního katalogu, pokud existuje. Lokální katalog může obsahovat větší množství informací a souvislostí, než se přenesou do Národního katalogu otevřených dat. Z tohoto důvodu by bylo vhodné, aby se uživatelé po nalezení datové sady poskytovatele v NKOD mohli podívat i do jeho katalogu.
- Posledním krokem po vyplnění všech údajů ve formuláři je stažení výsledného vyplněného dokumentu.
- Zbývá jeho odeslání prostřednictvím datové schránky na adresu datové schránky Národního katalogu otevřených dat (NKOD) s ID m3hp53v. Dokument je možné odeslat například přes spisovou službu poskytovatele jako přílohu datové zprávy s uvedeným předmětem zprávy „NKOD“.

## 9.9.3 Zrušení registrace lokálního katalogu

Registraci lokálního katalogu otevřených dat je možno zrušit zasláním dokumentu, který lze získat pomocí formuláře pro smazání lokálního katalogu. Postup zrušení je následující:

- V [Seznamu lokálních katalogů v NKOD](#) je nutné nalézt rušený lokální katalog a v jeho detailu kliknout na symbol smazání.
- Následně je nutné stáhnout dokument s vygenerovaným zrušením katalogu a ten poslat datovou schránkou (například přes spisovou službu poskytovatele) jako přílohu datové zprávy s předmětem „NKOD“ a adresovat ji na datovou schránku Národní katalog otevřených dat (NKOD) s ID m3hp53v.

## 9.9.4 Otevřená formální norma "[Rozhraní katalogů otevřených dat: DCAT-AP-CZ](#)"

Specifikace rozhraní lokálních katalogů otevřených dat (LKOD) pro jejich harvestování Národním katalogem otevřených dat (NKOD). Stejná specifikace platí i pro datový model NKOD a data poskytovaná z NKOD v podobě souborů ke stažení či SPARQL endpointu.





## 9.10 Stručný popis VDF

### 9.10.1 Základní charakteristiky VDF

Veřejný datový fond (VDF) je definován v [Informační koncepci ČR](#) (IKČR) jako [dílní cíl 5.10](#), je součástí eGovernmentu VS ČR a jeho poslání a role je definována následujícím způsobem:

*“Veřejný datový fond tvořený publikovanými veřejnými údaji veřejné správy je základní metodou pro sdílení veřejných informací mezi veřejnoprávními subjekty navzájem i pro sdílení veřejných údajů mezi veřejnoprávní a soukromoprávní sférou v ČR. Veřejný datový fond se od pouhé publikace automatizovaně čitelných otevřených dat posune též k publikaci právně závazných, platných a pravidelně aktualizovaných datových sad s jasně definovanou zodpovědností OVS za takové sady.”*

Obecným výchozím principem VDF je princip [P13 eGovernmentu “Otevřená data jako standard”](#) (Open Data by Default):

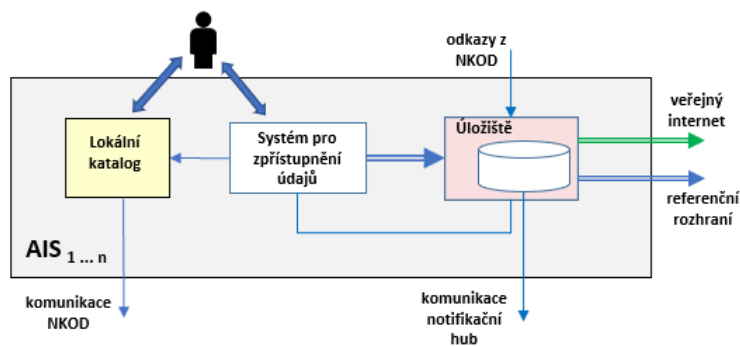
*“Veřejné údaje evidované orgány veřejné správy ve spravovaných ISVS musí být zveřejňovány jako otevřená data. Pro neveřejné údaje musí být jako otevřená data zveřejňována jejich anonymizovaná podoba, souhrn nebo statistika. V případě, že orgány veřejné správy sdílejí veřejné údaje (včetně anonymizované podoby neveřejných údajů, souhrnů nebo statistik) musí je sdílet jako otevřená data.”*

Příprava distribucí datových sad k publikaci do VDF probíhá stejným způsobem jako v případě přípravy distribucí publikovaných mimo VDF (běžná publikace otevřených dat). Zásadním rozdílem je způsob publikace, místa ukládání publikovaných distribucí a důraz na následnou rutinní automatizaci publikace, což vyžaduje doplnění agendových informačních systémů (AIS) o nové funkční oblasti a moduly.

Veřejný datový fond (VDF) je tvořen:

- všemi veřejnými údaji veřejné správy,
- nezbytnou softwarovou podporou zajišťující jejich sdílení a garantovanou dostupnost,
- stanovenými pravidly a definovanými standardy pro zajištění interoperability sdílených údajů,
- poskytovatelem dat definovanými a implementovanými procesy pro zajištění jejich právní závaznosti, platnosti, aktuálnosti a jasně definovanými zodpovědnostmi za sdílené údaje.

### 9.10.2 Základní logické komponenty VDF AIS a jejich význam



#### Systém pro zpřístupnění údajů

- Aktivně v pravidelných intervalech definovaných správcem ISVS exportuje obsah veřejných údajů do podoby datových sad a dávkově je předává do úložiště datových sad.
- Jedná se o logickou součást každého AIS, prakticky se může jednat o samostatný systém nebo to může být modul v rámci existujícího systému.

#### Úložiště datových sad



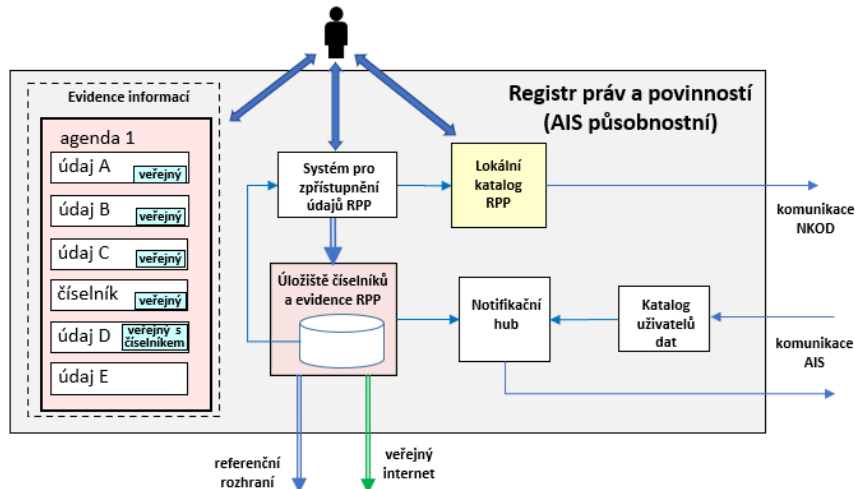
- Obsahuje publikovaný obsah registrovaných veřejných údajů vedených v daném publikujícím AIS (ISVS) v podobě distribucí datových sad.
- Provádí technickou kontrolu datových sad předaných “Systémem pro zpřístupnění údajů” a transformuje vytvořené datové sady do všech dalších formátů (distribucí) definovaných příslušnou datovou specifikací.

### LKOD (Lokální katalog otevřených dat)

- Lokální katalog otevřených dat obsahuje a spravuje katalogizační záznamy datových sad publikovaných veřejných údajů.
- Zpřístupňuje katalogizační záznamy k jejich katalogizaci do NKOD.
- Nabízí se dva možné způsoby jejich realizace:
  - LKOD pro každý konkrétní AIS (komponenta AIS) určený pro své publikované údaje,
  - LKOD konkrétní organizace VS, který katalogizuje všechny údaje publikované organizací.

V případě publikace distribucí číselníků se jedná o podobné publikační mechanismy a rovněž AIS Působnostní (RPP) je k tomuto účely již aktuálně doplněn o potřebné moduly.

### 9.10.3 Společné logické komponenty VDF související s AIS Působnostním (RPP)



#### Úložiště číselníků a evidence RPP – nová komponenta RPP

- Obsahuje a zpřístupňuje datové sady publikovaných veřejných číselníků ve formátu CSV, JSON-LD a pomocí SPARQL endpointu.
- Obsahuje a zpřístupňuje obsah (údaje) RPP jako datové sady ve formátech JSON-LD a prostřednictvím SPARQL endpointu.

#### Lokální katalog RPP – nová komponenta RPP

- Lokální katalog otevřených dat RPP obsahuje katalogizační záznamy pro:
  - datové sady zastřešující verze číselníků,
  - datové sady s verzemi číselníků,
  - datové sady s údaji, které jsou evidovány a spravovány v RPP (obsah RPP).
- Zpřístupňuje katalogizační záznamy k jejich katalogizaci v NKOD.

#### Katalog uživatelů dat – nová komponenta RPP

- Eviduje využívané veřejné údaje v jednotlivých agendách (zajišťuje ohlašovatel agendy).

#### Notifikační hub – rozšíření RPP



- Nástroj zajišťující s pomocí katalogu uživatelů dat notifikační službu, která informuje subjekty registrované v katalogu uživatelů dat VDF o změnách publikovaných veřejných údajů.

#### **Systém pro zpřístupnění údajů RPP**

- Připravuje obsah RPP a veřejné číselníky do podoby datových sad a dále je předává do úložiště datových sad k publikaci.

Nejsou uvedeny logické komponenty zabezpečující čtení publikovaných údajů pro výkon agendy. Kompletní problematika je detailně popsána a vysvětlena v „[Metodice poskytování dat ve veřejném datovém fondu](#)“.

### **9.10.4 Význam metodiky pro publikaci do VDF**

Podstatou publikace do VDF je vytvoření podmínek a nastavení všech modulů AIS pro zajištění automatického provádění všech nutných operací.

Obsah činností nutných k nastavení AIS a jejich nových komponent je v podstatě identický s činnostmi prováděnými v průběhu publikace jakýkoliv otevřených dat, neboť i pro datové sady publikované do VDF (registrované veřejné údaje) platí, že jsou to současně běžná otevřená data přístupná veřejnosti. Je tedy nutné posoudit jejich zpřístupnění z pohledu potenciálních přínosů a rizik, ošetřit podmínkami užití, vybrat vhodné datové modely, způsoby publikace a připravit katalogizační záznamy odpovídající požadovaným standardům.

#### **Oblasti činností popisovaných v metodice a jejich využití při publikaci do VDF:**

- Identifikace dat
  - Návrh témat otevřených dat – omezené využití návrhů v metodice, neboť publikace dat do VDF je především dána legislativními dokumenty a informační strategií veřejné správy ČR. Ty ukládají povinnost publikace registrovaných veřejných údajů v podobě otevřených dat.
  - Ošetření navržených datových sad – nutné plně respektovat doporučení uvedené v metodice (publikovaná data do VDF jsou současně i otevřená data publikované pro veřejné použití).
  - Správa katalogu témat otevřených dat – doporučeno, ale rozhodnutí závisí na každém poskytovateli, soulad s metodikou.
  - Analýza agendy a odvození údajů – povinné a vždy jako první krok přípravy datových sad k publikaci dle „[Metodiky tvorby a údržby konceptuálních datových modelů agend](#)“.
- Příprava datové sady
  - Příprava konceptuálního modelu datové sady – soulad s metodikou.
  - Výběr způsobů publikace – povinný způsob „publikace se sémantikou dat“, soulad s metodikou a „[Metodikou poskytování dat ve veřejném datovém fondu](#)“.
  - Návrh datových modelů a datových schémat – soulad s metodikou.
  - Příprava katalogizačního záznamu – soulad s metodikou.
  - Publikace a katalogizace datové sady – tato oblast je nejvíce specifická, neboť velkou roli v ní sehraávají nové komponenty AIS, příprava prvotní publikace bývá náročnější, a to i z toho důvodu, že je nutné nastavit vzájemnou komunikaci více komponent VDF. Publikace je složitější i z toho důvodu, že číselníky se publikují do úložiště číselníků v RPP a publikace datových sad s údaji kódovanými číselníky (její nastavení pro automatizaci publikace) se tak stává pracnější. Po ukončení publikace a katalogizace publikované datové sady v NKDO je nutné ještě doplnit u každého registrovaného publikovaného údaje v evidenci v RPP odkaz na datovou sadu v NKOD, ve které každý jednotlivý údaj se nachází. Tato oblast činností je upravena „[Metodikou poskytování dat ve veřejném datovém fondu](#)“.

#### **Shrnutí základních specifik publikace do VDF:**

- nutnost úprav stávajících AIS a doplnění potřebných nových komponent (nezbytná podmínka),



- všechny publikované údaje musí být evidovány v RPP,
- evidence musí být provedena dle metodiky a mělo by jí předcházet:
  - vytvoření sémantického slovníku pojmů agendy,
  - odvození údajů subjektů a objektů práva,
  - následně jejich evidence v RPP,
- konceptuální modely by měly být vytvořeny:
  - na základě pojmů slovníku veřejné správy (zajištění sémantiky a sémantické interoperability),
  - nebo z něho přímo extrahovány,
- měla by převládat preference grafových modelů a publikace dat v úplném kontextu (nejlépe propojená data),
- zdrojem dat je vždy datový kmen příslušné agendy,
- datové sady (distribuce) jsou vždy uloženy v úložišti příslušného AIS, v případě číselníků v „úložišti číselníků a evidence RPP“.

### 9.10.5 Publikace distribucí datové sady do datového úložiště agendy

- Systém pro zpřístupnění údajů připraví obsah datové sady v podobě datového souboru. Připravený datový soubor musí dodržovat připravenou Datovou specifikaci nebo příslušnou OFN.
- Systém pro zpřístupnění údajů zašle připravený obsah datové sady do úložiště datových sad.
- Úložiště datových sad provede technickou kontrolu zasláného obsahu:
  - Kontrola správného formátování (např. JSON, CSV nebo XML formátování).
  - Kontrola validity datové struktury vůči datovým schémátům definovaným Datovou specifikací nebo OFN.
  - V případě špatné syntaxe zašle zpět systému pro zpřístupnění údajů chybové hlášení a skončí.
- Úložiště datových sad zpřístupní všechny vytvořené distribuce jako datové soubory dostupné ke stažení prostřednictvím ISGOD a z veřejného internetu prostřednictvím URL pro stažení datového souboru (URL pro stažení datového souboru je stejné pro přístup prostřednictvím ISGOD i veřejného internetu).

### 9.10.6 Katalogizace v LKOD agendy

Katalogizace v LKOD je principiálně shodná s katalogizací mimo VDF (katalogizace běžných otevřených dat), specifiky jsou dána pouze konkrétním řešením v AIS, nebo LKOD publikující organizace.

V rutinním režimu publikace systém pro zpřístupnění údajů konkrétního AIS vytváří kompletní katalogizační záznam o datové sadě, včetně doplnění metadat o distribucích vytvořených úložištěm datových sad a vše katalogizuje v lokálním katalogu automaticky.

### 9.10.7 Informování notifikačního hubu o změně publikovaného veřejného údaje

Informování notifikačního hubu probíhá automaticky, systém pro zpřístupnění údajů mu informaci o změně publikovaného veřejného údaje (obsahu datové sady).

Notifikační hub je nástroj zajišťující s pomocí katalogu uživatelů dat notifikační službu, která informuje subjekty registrované v katalogu uživatelů dat z VDF o změnách veřejných údajů, vedoucích následně i ke změnám v souvisejících publikovaných datových sadách.

Notifikace probíhá automaticky při ohlášení změny veřejného údaje v publikovaných datových sadách ze strany poskytujícího ISVS.



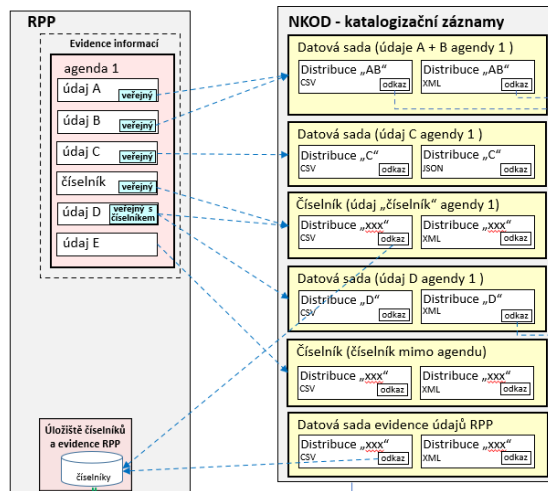
### 9.10.8 Doplnění evidence údajů v RPP

Evidenci údajů v RPP a s ní související způsoby práce s RPP plně řeší "[Metodika evidence údajů v RPP](#)".

K zajištění funkčnosti VDF:

- označení údajů vedených v agendách a registrech příznakem "veřejný údaj / neveřejný údaj", s uvedením odkazu na konkrétní legislativu, ze které vyplývá jejich případná neveřejnost,
- uvedení IRI datových sad v NKOD, jejichž prostřednictvím je veřejný údaj publikován,
- v případě údaje kódovaného číselníkem uvedení také IRI datové sady v NKOD, v níž je číselník zpřístupněn a publikován jako otevřená data.

Ukázka provázání evidence v RPP s katalogizačními záznamy v NKOD:



Informace zde uvedené k problematice publikace údajů do veřejného datového fondu jsou pouze velmi stručným výtahem z dokumentů zaměřených na VDF.

Kompletní problematika je detailně popsána a vysvětlena v „[Metodice poskytování dat ve veřejném datovém fondu](#)“.

Popis globální architektury VDF je dostupný na znalostním webu odboru Hlavního architekta eGovernmentu v části „[Veřejný datový fond](#)“, k dispozici je také záznam školení „[Veřejný datový fond \(VDF\) v architektuře veřejné správy](#)“ dostupný na portále otevřených dat v sekci vzdělávání.