

Koncepce katalogizace otevřených dat VS ČR

(zkrácená verze)

Praha, květen – září 2012

Zpracovali:

Dušan Chlapek Fakulta informatiky a statistiky, Vysoká škola ekonomická v Praze

Jan Kučera Fakulta informatiky a statistiky, Vysoká škola ekonomická v Praze

Martin Nečaský Matematicko-fyzikální fakulta, Univerzita Karlova v Praze

Spolupracovali:

Tomáš Kroupa Ministerstvo vnitra České republiky

Vladimír Weis Ministerstvo vnitra České republiky

Obsah

1	Koncepce katalogizace otevřených dat VS ČR	5
1.1	Úvod od Koncepce	5
1.2	Přínosy katalogu otevřených dat VS ČR	6
1.2.1	Usnadnění přístupu k datům veřejné správy	6
1.2.2	Vytvoření předpokladu pro snazší opětovné použití dat veřejné správy ČR	6
1.2.3	Vytvoření předpokladu pro využívání otevřených propojitelných	6
1.2.4	Vytvoření předpokladu pro dosažení vyšší transparentnosti veřejné správy.....	7
1.3	Katalogizovaná data	7
1.3.1	Vymezení otevřených dat a rozsahu katalogizace	7
1.4	Role a místa zodpovědná za katalogizaci otevřených dat VS ČR	8
1.4.1	Správce Datového katalogu	9
1.4.2	Provozovatel Datového katalogu.....	9
1.4.3	Poskytovatel dat	9
1.4.4	Kurátor dat	10
1.4.5	Redaktor.....	10
1.4.6	Koncový uživatel.....	10
1.5	Postup katalogizace a související doporučení	10
1.5.1	Postup katalogizace.....	10
1.5.2	Pravidla katalogizace.....	11
1.5.3	Struktura katalogizačního záznamu	11
1.5.4	Doporučení ohledně klasifikačních struktur (slovníků a taxonomií)	14
1.6	Správa a provoz katalogu dat VS ČR	14
1.7	Zajištění kvality obsahu datového katalogu.....	14
1.7.1	Kvalitativní atributy obsahu katalogu a jednotlivých záznamů	14
1.7.2	Využití koncových uživatelů (veřejnosti)	15
1.7.3	Využití metod strojového učení	16
1.8	Zajištění bezpečnosti provozu datového katalogu	16
1.8.1	Možná ohrožení Datového katalogu a cíle z hlediska bezpečnosti.....	16
1.8.2	Bezpečnostní požadavky a opatření pro jejich naplnění	17
1.8.3	Další doporučení pro zajištění bezpečnosti Datového katalogu a jeho provozu	18
1.8.4	Přístup uživatelů do Datového katalogu a jejich oprávnění	19
1.9	Cílová softwarová architektura nástroje pro Datový katalog	19
1.10	Požadavky na nástroj pro datový katalog.....	20
1.11	Uživatelé datového katalogu a podmínky jeho použití	21
1.11.1	Zapojení veřejnosti do plnění datového katalogu.....	22
1.12	Vzdělávání v oblasti katalogizace otevřených dat VS ČR	22
1.13	Legislativní změny související s katalogizací otevřených dat.....	22
1.13.1	Podmínky užívání otevřených dat VS ČR.....	22
1.13.2	Pojem otevřených dat	23
1.13.3	Legislativní ukotvení katalogizace otevřených dat VS ČR.....	24
1.14	Vazba na připravovaný portál Publicdata.eu	24

1.14.1	Katalogizační nástroj	24
1.14.2	Taxonomie pro klasifikaci katalogizačních záznamů	25
1.15	Scénáře užití datové katalogu VS ČR	25
2	Plán realizace koncepce	26
2.1	Etapy a hlavní skupiny činností	26
2.2	Popis etap realizace koncepce	27
2.2.1	Zjednodušený Datový katalog	27
2.2.2	Příprava legislativních podmínek pro publikaci otevřených dat	28
2.2.3	Výběr nástroje, jeho dodavatele a provozovatele	28
2.2.4	Implementace vybraného nástroje	28
2.2.5	Ověřovací provoz plnohodnotného nástroje	29
2.2.6	Zahájení provozu	29
2.3	Seznam výstupů	29
2.4	Pracnost katalogizace	30
2.4.1	Pracnost pro redaktory obsahu Datového katalogu	32
2.4.2	Pracnost pro kurátory dat	33
2.5	Odhad potřebných nákladů	33
2.5.1	Pořizovací náklady Datového katalogu	33
2.5.2	Náklady provozu a údržby Datového katalogu	34
2.5.3	Náklady vytváření katalogizačních záznamů a provádění redakce	34
2.6	Způsob provozování katalogizačního nástroje	35
2.7	Doporučení pro výběr dodavatele a provozovatele katalogizačního nástroje	36
2.8	Přehled odhadovaných nákladů jednotlivých etap realizace Koncepce	36
3	Použité zkratky	37
4	Zdroje	38
4.1	Odkazované předpisy	40
4.1.1	České právní předpisy	40
4.1.2	Evropské předpisy	40

1 Konceptce katalogizace otevřených dat VS ČR

1.1 Úvod od Konceptce

V rámci tohoto dokumentu je popsána Konceptce katalogizace otevřených dat veřejné správy České republiky (dále Konceptce). Tato Konceptce vznikla v rámci plnění závazku Akčního plánu Partnerství pro otevřené vládnutí „Zpřístupnění dat a informací“, který cílí na zlepšení přístupu k datům vytvářených veřejnou správou ČR (VS ČR) [2], a který byl schválen usnesením Vlády České republiky ze dne 4. dubna 2012 č. 243 [57].

Konceptce katalogizace otevřených dat VS ČR představuje ucelenou sadu doporučení, která souhrnně odpovídají na otázku, kdo a jak by měl katalogizovat otevřená data VS ČR. Konceptce je navrhována s časovým výhledem realizace v délce 3 až 5 let, navazující část Plán realizace Konceptce potom rozpracovává jednotlivé kroky realizace a harmonogram realizace.

Konceptce je rozdělena do kapitol, které jsou vždy zaměřeny některou z dále uvedených otázek. Tabulka 1 u každé otázky uvádí i příslušné kapitoly, které se zabývají jejím řešením.

Tabulka 1: Otázky řešené v Konceptci a související kapitoly

Otázky řešené v konceptci	Kapitola
Jaké jsou přínosy datového katalogu VS ČR?	kap. 1.2
Jaká data katalogizovat?	kap. 1.3
Kdo by měl data katalogizovat?	kap. 1.4
Jak by měl při katalogizaci postupovat?	kap. 1.5
Kdo bude spravovat a provozovat katalog dat VS ČR?	kap. 1.6
Jak bude zajištěna kvalita obsahu datového katalogu?	kap. 1.7
Jak bude zajištěna bezpečnost provozu datového katalogu?	kap. 1.8
Jaké jsou požadavky na nástroj pro datový katalog?	kap. 1.10
Kdo bude moci datový katalog využívat a za jakých podmínek?	kap. 1.11
Budou moci do datového katalogu přispívat občané ČR z řad veřejnosti?	kap. 1.11
Jak by měl být nástroj pro datový katalog pořízen?	kap. 1.10
Kdo a jak bude zajišťovat vzdělávání v oblasti katalogizace otevřených dat VS ČR?	kap. 1.12
Jaké budou potřebné či vhodné legislativní změny pro fungování datového katalogu?	kap. 1.13
Jak bude zajištěna vazba na připravovaný portál Publicdata.eu?	kap. 1.14
Jaká by měla být sw architektura nástroje pro datový katalog?	kap. 1.9

Řešení jednotlivých otázek vycházejí ze situace, kdy je realizována doporučená (v plné verzi konceptci vybraná) varianta budování Datového katalogu jako centralizovaného a homogenního řešení, kdy vkládání dat provádí více pověřených subjektů a je prováděna úplná redakce.

1.2 Přínosy katalogu otevřených dat VS ČR

Hlavní přínosy Katalogu otevřených dat veřejné správy jsou:

- usnadnění přístupu k datům veřejné správy,
- vytvoření předpokladu pro snazší opětovné použití dat veřejné správy,
- vytvoření předpokladu pro využívání otevřených propojitelných dat
- vytvoření předpokladu pro dosažení vyšší transparentnosti veřejné správy.

I když budou dále přiblíženy hlavní přínosy Datového katalogu, je třeba si uvědomit, že Datový katalog sám o sobě není cílem, ale pouze prostředkem pro efektivní využívání otevřených dat veřejné správy. Podle Sdělení Evropské komise „*Veřejně přístupné údaje, hnací síla inovací, růstu a transparentní správy*“ [19] mají informace shromažďované veřejnoprávními subjekty¹ nevyužitý potenciál v možnosti opakovaného použití v nových produktech a službách. Dle studie [56] je pak odhadováno, že obrát trhu vytvářeného přímo opětovným použitím informací veřejného sektoru na území 27 států Evropské unie (EU27) v roce 2008 činil 28 miliard EUR. V citované studii se dále uvádí, že ekonomické přínosy až do výše 40 miliard EUR ročně z přímého využití informací veřejného sektoru na území EU27 by mohlo přinést další usnadnění přístupu k těmto informacím.

Jednou z možností, jak využít výše naznačeného potenciálu, může být aplikace principů otevřených dat pro data veřejné správy. Důvodem je skutečnost, že otevřená data cílí na usnadnění využívání dat (po technické i právní stránce).

Otevřená data veřejné správy jsou důležitým zdrojem informací o tom, jak je realizována politika a jak je nakládáno s veřejnými rozpočty. Aktivním publikováním otevřených dat se může zvýšit informovanost občanů a dalších zájemců o činnosti veřejné správy a tím může dojít i k posílení transparentnosti veřejné správy.

1.2.1 Usnadnění přístupu k datům veřejné správy

Jak je uvedeno v Akčním plánu České republiky „*Partnerství pro otevřené vládnutí*“: „*Otevřená data jsou smysluplná, pokud jsou dohledatelná a přístupná*“ [2]. Otevřená data, která nemohou jejich potenciální uživatelé snadno najít, nemohou ani efektivně využívat. Centrální Datový katalog bude sloužit jako jedno místo, kde budou moci občané, obchodní společnosti a jiné instituce včetně veřejnoprávních subjektů vyhledávat otevřená data veřejné správy. I po zavedení Datového katalogu budou moci orgány veřejné správy zpřístupňovat otevřená data způsobem, který jim vyhovuje, například na svých webových stránkách nebo portálech. Datový katalog by ale měl potenciálním uživatelům přístup k datům usnadnit, protože díky němu odpadne (někdy pracné a zdlouhavé) prohledávání oddělených webových stránek orgánů veřejné správy. Díky záznamům v Datovém katalogu bude možné nejenom zjistit, že určitá data existují, ale i kde se nacházejí a jak je lze získat.

1.2.2 Vytvoření předpokladu pro snazší opětovné použití dat veřejné správy ČR

Vytvořením Datového katalogu se vytváří také předpoklad pro opětovné využití dat, které orgány veřejné správy zpřístupní jako otevřená data, protože aby data mohli zájemci využívat, musí se nejprve dozvědět o jejich existenci. Datový katalog bude jednotným místem, kde zájemci budou moci zjistit, zda, a případně kde, jsou publikována data z určité oblasti. Jako jeden z kroků, pomocí kterého lze podpořit využití potenciálu ukrytého v opětovném použití dat veřejné správy, je vybudování datového katalogu uvedeno i ve výše uvedeném Sdělení Evropské komise [19].

1.2.3 Vytvoření předpokladu pro využívání otevřených propojitelných

V souvislosti s otevřenými daty se také hovoří o tzv. otevřených propojitelných datech. Jedná se o otevřená data využívající technologie propojitelných dat (Linked Data) [5]. Tyto technologie umožňují vzájemně propojovat související data pocházející z různých zdrojů a navíc vyjádřit, o čem data a propojení mezi nimi vypovídají (tj. vyjádřit sémantiku dat). Publikace a propojování dat je založena na otevřených standardech nezávislých na konkrétním dodavateli či software. Každý může otevřená propojení mezi daty využívat a hledat tak v datech skryté souvislosti. Každý také může svá vlastní data propojit na určitou část otevřených propojitelných dat VS ČR a získat tak vazby i na další související data

¹ Zde je použita terminologie citovaného Sdělení, informacemi shromažďovanými veřejnoprávními subjekty lze nicméně rozumět i data orgánů veřejné správy.

VS ČR. Jinými slovy může snadno umístit svá data do informačního kontextu dat VS ČR a v rámci tohoto kontextu pracovat.

Technologie propojitelných dat mají navíc velký potenciál výrazně snížit náklady procesu publikace a propojování (tj. integrace) dat. Může být totiž postupný - rozložený v čase i mezi jednotlivé poskytovatele dat a koncové uživatele. Poskytovatel zveřejňuje pouze svá primární data. Sekundární data (tj. data spravovaná a zveřejňovaná primárně někým jiným) poskytovatel nezveřejňuje. Místo toho svá primární data na sekundární data propojuje a dále se nestará o jejich správu. Propojení navíc může vytvořit a zveřejnit v rámci stanovených pravidel i třetí subjekt (např. konzument dat, který propojení potřebuje).

Technologie propojitelných dat jsou postaveny na standardech současného webu pocházejících z dílen konsorcia W3C. Jedná se tedy o neproprietární a na konkrétním výrobci nezávislé technologie, s nimiž je možné pracovat prostřednictvím řady běžných softwarových nástrojů. Jejich využití tedy není finančně náročné.

1.2.4 Vytvoření předpokladu pro dosažení vyšší transparentnosti veřejné správy

Výše bylo uvedeno, že otevřená data hrají důležitou roli v naplňování principů otevřeného vládnutí a posilování transparentnosti veřejné správy. V této snaze hraje datový katalog klíčovou úlohu díky tomu, že představuje jedno centrální místo, kde zájemci mohou vyhledávat požadovaná otevřená data o fungování veřejné správy [14]. Datový katalog tak představuje předpoklad či nástroj pro dosažení vyšší transparentnosti veřejné správy.

1.3 Katalogizovaná data

Tato kapitola odpovídá na klíčovou otázku spojenou s katalogizací otevřených dat veřejné správy ČR a to, jaká data by vlastně měla být katalogizována.

1.3.1 Vymezení otevřených dat a rozsahu katalogizace

V cílovém stavu by v Datovém katalogu dat měla být katalogizována otevřená data poskytovaná orgány veřejné správy, u nichž se předpokládá volné a opakované použití ostatními orgány veřejné správy i dalšími právními subjekty a soukromými osobami². Povinnost publikace a katalogizace otevřených dat bude ukládat jednotlivým orgánům veřejné správy příslušná legislativa (viz kapitola 1.13).

V prvním kroku předpokládáme uložení povinnosti katalogizovat otevřená data povinně ve vybraných oblastech (viz závazek ČR v Akčním plánu Partnerství pro otevřené vládnutí [2]) a dobrovolně v ostatních oblastech.

Současně předpokládáme, že v Datovém katalogu mohou být dobrovolně katalogizována i data, která nesplňují některou z podmínek otevřených dat, přičemž musí být jasně indikováno, kterou z uvedených podmínek otevřených dat nesplňuje. Otevřená data veřejné správy jsou data, která jsou³:

1. úplná - data jsou zveřejněna v maximálním možném rozsahu. Rozsah může být definován právním předpisem, usnesením vlády, příp. poskytovatelem dat. Například seznam **všech** nemovitostí s číslem popisným nebo evidenčním v obci XY, nebo seznam **všech** památkově chráněných objektů v obci XY.
2. primární (původní) - data, která jsou zveřejněna původcem dat v podobě, v jaké byla původcem jako primární (původní) vytvořena⁴. Za primární data se považují i
 - a. referenční údaje ze základních registrů,
 - b. data z registrů⁵ a rejstříků VS,
 - c. agregovaná data (např. výsledky voleb) pokud není možné zveřejnit data, z nichž byla provedena agregace,

² Je třeba ještě zmínit, že na základě Směrnice 2007/2/ES je zřízen portál INSPIRE sloužící i jako katalog geodat. Bez změny evropské legislativy tak nelze přesunout katalogizaci těchto dat, která jsou zároveň otevřenými daty, z portálu INSPIRE na zde navrhovaný centrální Datový katalog. V Datovém katalogu by tak měl existovat katalogizační záznam o portálu INSPIRE a případně je vhodné do budoucna zvažovat zajištění vyhledávání geodat katalogizovaných v na portálu INSPIRE z Datového katalogu.

³ Upraveno dle [54]

⁴ Za původce dat je považován orgán veřejné správy, který údaje vytvořil.

⁵ Pokud taková datová data ukládá publikovat jako veřejná data nějaká právní norma ČR.

- d. agregovaná data - (např. statistiky nad jinými otevřenými daty) pokud je uveden způsob agregace a odkaz na zveřejněná primární data, z nichž byla agregace provedena.
3. zveřejněná bez zbytečného odkladu - zveřejnění dat není zdrženo činnostmi, které nesouvisí s jejich přípravou; činnosti nezbytné pro publikaci dat jsou provedeny v čase, který umožní jejich zveřejnění bez nepřiměřeně dlouhé prodlevy od okamžiku vzniku dat,
4. snadno dostupná - data jsou dostupná a dohledatelná běžnými ICT nástroji a prostředky,
5. strojově čitelná - data ve formátu, který je strukturovaný takovým způsobem, že pomocí programové aplikace lze z dat získat žádané (vybrané) údaje
6. neomezuující přístup - data dostupná způsobem, který nediskriminuje jednotlivce nebo skupinu osob,
7. používající standardy s volně dostupnou specifikací (otevřené standardy) - data musí být ve formátu, který je volně (bezplatně) dostupný pro libovolné použití nebo do takového formátu převoditelný volně (bezplatně) dostupnou aplikací,
8. zpřístupněna za jasně definovaných podmínek užití dat (licence) s minimem omezení - podmínky musí být jasně a zřetelně definovány a zveřejněny,
9. stále dostupná - data jsou dostupná on-line po dobu uvedenou jejich poskytovatelem,
10. dostupná uživatelům při vynaložení minima možných nákladů na jejich získání - poskytovatelé jsou v souvislosti s poskytováním dat oprávněni žádat úhradu maximálně ve výši, která nesmí přesáhnout náklady spojené s jejich zpřístupněním uživateli; poskytovatel dat může jednorázově vyžádat i úhradu za mimořádně náročné pořízení dat, pokud si uživatel zpřístupnění těchto dat vyžádá⁶.

Data veřejné správy jsou považována za otevřená, pokud z výše uvedených podmínek splňují alespoň podmínky č. 1, 4, 5, 7, 8 a 10, tj. jsou to data úplná, snadno dostupná, strojově čitelná, používající standardy s volně dostupnou specifikací, zpřístupněná za jasně definovaných podmínek užití s minimem omezení a dostupná uživatelům při vynaložení minima možných nákladů na jejich získání. Pro to, aby data veřejné správy byla považována za otevřená tak není striktně vyžadováno, aby navíc byla primární, zveřejněná bez zbytečného odkladu, neomezuující přístup a stále dostupná, protože zajištění těchto podmínek nemusí být vždy snadné. Protože ale splnění těchto podmínek dále zlepšuje využitelnost dat jejich potenciálními uživateli, je vhodné se o jejich naplnění snažit. Data veřejné správy splňující všech deset výše uvedených podmínek tak lze považovat za dobře publikovaná otevřená data.

K vymezení otevřených dat je třeba ještě podotknout, že otevřená data jsou publikována, aby mohla být dále využívána ostatními orgány veřejné správy i dalšími právními subjekty a soukromými osobami. Veřejná správa disponuje ale i celou řadou dat, která nejsou určena veřejnosti. Takováto chráněná data nemohou být publikována jako otevřená data a to ani v případě, že splňují jednu nebo více výše uvedených vlastností otevřených dat. Ačkoli jsou referenční údaje ze základních registrů považovány za primární data, neznamená to, že by automaticky všechny tyto údaje bylo možné považovat za otevřená data. Zákon č. 111/2009 Sb., o základních registrech, stanoví, jak lze k referenčním údajům ze základních registrů přistupovat a jak s nimi lze nakládat a to individuálně pro jednotlivé základní registry a data z těchto registrů.

Obecně je třeba při publikaci otevřených dat nejprve rozhodnout, zda se jedná o data určená veřejnosti a až následně je třeba se zabývat tím, jak při publikaci těchto dat vyhovět výše uvedeným podmínkám. Pokud data v primární podobě jsou chráněná a nelze je poskytnout veřejnosti, je vhodné se zabývat i tím, zda by nebylo možné veřejnosti zpřístupnit alespoň data odvozená z primárních chráněných dat (např. zpřístupnění souhrnných statistik, anonymizovaných dat apod.). U dat, která jsou určena veřejnosti nebo jsou zveřejňována na základě právního předpisu, se ale jeví vhodné vždy zvážit, zda by je bylo možné zveřejnit jako otevřená data, tj. aby zveřejněná data splňovala výše uvedené vlastnosti.

1.4 Role a místa zodpovědná za katalogizaci otevřených dat VS ČR

V návaznosti na doporučenou variantu budování Datového katalogu (viz část Analýza současného stavu a varianty řešení v plné verzi Koncepce) jsou v této kapitole podrobněji vymezeny role a zodpovědnosti dotčených orgánů VS ČR ve vztahu ke katalogizaci otevřených dat. Kapitola tak odpovídá na otázku, kdo by měl data katalogizovat.

⁶ Upraveno podle § 17 odst. 1 zák. č. 106/1999 Sb. a podle návrhu úpravy směrnice 2003/98/ES [18].

V této koncepci navržený způsob katalogizace dat nebude⁷ probíhat automatizovaně, ale je nutný zásah člověka, který bude provádět záznam do katalogu a následnou redakci záznamu o katalogizovaných datech.

Role v rámci katalogizace otevřených dat VS:

- správce Datového katalogu,
- provozovatel Datového katalogu,
- poskytovatel dat,
- kurátor dat,
- redaktor,
- koncový uživatel⁸.

1.4.1 Správce Datového katalogu

Správce Datového katalogu je subjekt odpovědný za Datový katalog. Subjekt plnící roli správce Datového katalogu může plnit i roli provozovatele Datového katalogu.

Předpokládáme, že roli správce Datového katalogu bude vykonávat MV ČR.

Správce Datového katalogu bude zodpovídat za:

- zajištění výběru nástroje,
- zajištění implementace vybraného nástroje,
- definici parametrů pro provoz Datového katalogu,
- sběr podnětů na zlepšení Datového katalogu a zadávání a zajištění implementace změnových požadavků,
- zajištění provozu Datového katalogu (vlastními silami nebo pověřením jiného subjektu rolí provozovatele Datového katalogu),
- organizaci práce s Datovým katalogem,
- určování redaktorů,
- zajišťování přístupových práv.

1.4.2 Provozovatel Datového katalogu

Provozovatel Datového katalogu provádí činnost spočívající zejména v zajištění běžného chodu a údržbě Datového katalogu a potřebného softwarového a hardwarového vybavení. Provozovatel odpovídá za zajištění bezpečnosti Datového katalogu.

Provozovatel Datového katalogu bude zodpovídat za:

- zajištění realizace a testování úprav parametrů a konfigurace prostředí pro provoz Datového katalogu podle pokynů správce Datového katalogu,
- provoz Datového katalogu v rámci dohodnuté úrovně poskytovaných služeb (dostupnost, doba odezvy, ...),
- zálohování a archivaci Datového katalogu,
- zpracování reportů o provozu Datového katalogu.

1.4.3 Poskytovatel dat

Poskytovatelem dat je pro oblast otevřených dat ve veřejné správě ČR jakýkoliv orgán veřejné správy, který má v rámci svých kompetencí právo či povinnost data zveřejňovat, a který rozhodne o zveřejnění dat ve formě otevřených dat. Pokud data pocházejí z některého z informačních systémů veřejné správy, je poskytovatelem těchto dat správce příslušného ISVS.

Poskytovatel dat bude zodpovídat za:

- zveřejnění dat ve formě otevřených dat,
- identifikaci potenciálních přínosů a rizik plynoucích z poskytování dat formou otevřených dat,

⁷ Automatizovaně budou probíhat specializované procedury, které budou přispívat ke zvýšení kvality záznamů v katalogu dat (např. kontroly existence odkazů na data, příp. značkování dat).

⁸ Také označován jako průzkumník - např. v ISDP.

- určování podmínek, za jakých jsou data zveřejněna ve formě otevřených dat, a to zejména právních (licence)⁹, ekonomických, časových a technických,
- pověření kurátora správou dat.

1.4.4 Kurátor dat

Kurátorem je osoba, která zajišťuje zveřejnění otevřených dat a údržbu záznamů o datech v Datovém katalogu.

Kurátor bude zodpovídat za:

- zveřejnění otevřených dat,
- vytvoření a aktualizaci záznamu o poskytnutých otevřených datech v Datovém katalogu, včetně zařazení záznamů do příslušných klasifikačních struktur a vazby na ISDP a IS o ISVS,
- zneplatnění záznamů v Datovém katalogu.

1.4.5 Redaktor

Redaktor je osoba, která provádí obsahovou kontrolu vložených záznamů o otevřených datech VS ČR.

Redaktor zodpovídá za:

- ověření korektnosti a klasifikace záznamu v Datovém katalogu, tj. za ověření:
 - vyplnění povinných atributů,
 - existenci odkazu na zdroj dat,
 - shodu popisu a obsahu dat,
 - zařazení datové množiny do klasifikačních struktur¹⁰ (např. EUROVOC),
 - provázání se záznamy v systémech ISDP a IS o ISVS,
- komunikaci s poskytovateli dat ohledně kvality záznamu v Datovém katalogu,
- označení záznamu ke zveřejnění.

Obsazování role redaktora bude spadat do působnosti MV ČR.

1.4.6 Koncový uživatel

Koncový uživatel je osoba, která s použitím Datového katalogu:

- vyhledává záznamy o otevřených datech VS ČR,
- posílá správci Datového katalogu podněty na vytvoření nových záznamů, resp. úpravu stávajících záznamů v Datovém katalogu.

V roli koncového uživatele může vystupovat kdokoli bez nutnosti registrace v Datovém katalogu.

1.5 Postup katalogizace a související doporučení

Zatímco předcházející kapitola se věnovala tomu, kdo by měl katalogizovat otevřená data VS ČR, tato kapitola se zabývá otázkou, jak by příslušné orgány VS měly při katalogizaci postupovat. V této kapitole je navržen obecný postup katalogizace otevřených dat VS ČR. Předpokládáme, že jeho konkrétní podoba bude později popsána v podrobné metodice katalogizace otevřených dat VS ČR.

1.5.1 Postup katalogizace

Vložení záznamu do katalogu dat:

- Určení dat ke katalogizaci¹¹
- Založení katalogizačního záznamu¹² - vyplnění povinných atributů
- Zařazení záznamu do klasifikačních struktur (např. EUROVOC)
- Provázání se záznamy v systémech ISDP a IS o ISVS

Redakce vloženého záznamu v katalogu dat

⁹ V mezích platné legislativy.

¹⁰ Klasifikační struktury budou součástí návrhu Metodiky katalogizace otevřených dat VS ČR.

¹¹ Lze předpokládat, že v budoucnu budou k dispozici poloautomatizované nástroje, které umožní vyhledávat otevřená data a upozornit na jejich existenci.

¹² V Datovém katalogu budou uloženy pouze katalogizační záznamy a nikoli vlastní katalogizovaná data.

- Ověření korektnosti a klasifikace záznamu:
 - vyplnění povinných atributů
 - existenci odkazu na zdroj dat
 - shodu popisu a obsahu dat
 - zařazení datové množiny do klasifikačních struktur (EUROVOC)
 - provázání se záznamy v systémech ISDP a IS o ISVS

Oprava povinných atributů záznamu o katalogizaci dat bude prováděna příslušnými kurátory dat na základě podnětu redaktora zasláného poskytovateli dat, jejichž záznam má být opraven.

- Označení záznamu jako záznamu, který prošel redakcí

1.5.2 Pravidla katalogizace

Provádění katalogizace otevřených dat veřejné správy by se mělo řídit těmito základními pravidly:

- Jsou katalogizována pouze existující otevřená data.
- Data jsou katalogizována bez zbytečného odkladu po uveřejnění nebo současně s ním.
- Katalogizační záznam má českou i anglickou jazykovou verzi (pro zajištění využitelnosti katalogizačních záznamů i v připravovaném evropském datovém katalogu).
- I katalogizační záznamy samotné jsou zpřístupňovány ve formě otevřených dat.

1.5.3 Struktura katalogizačního záznamu

Záznam Datového katalogu (katalogizační záznam) by měl mít následující strukturu. Prvky struktury, které jsou povinné, jsou označeny písmenem *P*, nepovinné prvky jsou pak označeny písmenem *N*. Podrobnosti k vyplňování hodnot navržených atributů katalogizačního záznamu budou uvedeny v Metodice katalogizace otevřených dat VS ČR. Položky označené „K:” vyplňuje kurátor při katalogizaci dat. Položky označené „R:” vyplňuje redaktor při redakci záznamu v Datovém katalogu. Ostatní položky jsou generovány automaticky. Cílem je maximálně zjednodušit ruční práci při zakládání katalogizačního záznamu.

Prvky katalogizačního záznamu jsou rozděleny na základní a rozšiřující. Základní prvky je třeba realizovat v Datovém katalogu od samého počátku. Rozšiřující prvky katalogizačního záznamu mají nižší prioritu a je možné je do Datového katalogu přidat až s určitým časovým odstupem. Rozšiřující prvky jsou zpravidla prvky, které umožňují fungování některých pokročilých funkcí Datového katalogu, ale nejsou pro jeho fungování zcela nezbytné.

Základní prvky katalogizačního záznamu jsou:

- **K: Název záznamu/název dat (P)** - výstižné pojmenování katalogizačního záznamu, respektive dat, která katalogizační záznam popisuje
- **URL katalogizačního záznamu (P)** - jednoznačný neměnný identifikátor záznamu používaný v rámci sítě internet, bude vytvořen automaticky
- **K: Jazyk záznamu (P)** - jazyk¹³, ve kterém je uveden popis záznamu a jeho další atributy
- **K: Platnost záznamu (P)** - obecné označení, zda je katalogizační záznam platný či nikoli¹⁴
- **R: Stav redakce záznamu (P)** - označení, zda katalogizační záznam prošel redakcí či nikoli
- **K: Popis dat (P)** - text, který jasně uvádí, jaká data katalogizační záznam popisuje a o čem tato data vypovídají
- **K¹⁵: Poskytovatel dat (P)** - identifikace poskytovatele dat, skládá se z následujících atributů
 - **K: IČO (P)** - identifikační číslo poskytovatele dat
 - **Název poskytovatele dat (P)** - platný úplný název poskytovatele dat¹⁶
 - **Web poskytovatele dat (N)** - odkaz na domovskou stránku webu poskytovatele dat¹⁷

¹³ Katalogizační nástroj musí připouštět možnost vícejazyčného popisu katalogizačního záznamu. Výchozí hodnotou bude „CZ”.

¹⁴ Výchozí hodnota bude nastavena na „Platný”. Ručně nutno změnit na „Neplatný”.

¹⁵ Až bude v plném provozu JIP / KAAS (Jednotný identitní prostor / Katalog autorizačních a autentizačních služeb) bude automatizovaně vkládáno.

¹⁶ Bude automaticky doplněn z registru osob (ROS) na základě zadaného IČO poskytovatele dat (<http://www.szrcr.cz/registr-osob>).

¹⁷ Bude automaticky doplněn z Portálu veřejné správy (<http://portal.gov.cz>).

- **Kontakt (P)** - kontaktní email poskytovatele, na kterém poskytovatel poskytuje informace široké veřejnosti¹⁸
- **K: Podmínky užití dat (P)** - informace o právních podmínkách využití dat uvedených v katalogizačním záznamu, skládá se z následujících atributů
 - **K: Přehled splnění podmínek otevřenosti dat (P)** - přehled podmínek, které musí data splňovat, aby byla považována za otevřená data veřejné správy spolu s přehledem splnění těchto podmínek v případě dat uvedených v katalogizačním záznamu
 - **K: Podmínky užití / Licence (N)** - označení podmínek / licence
 - **K: Odkaz na popis podmínek užití / znění licence (N)** - nejlépe pomocí uvedení odkazu na webovou stránku se zněním podmínek / licenční smlouvy
- **K: Vazba na ISVS (N)** - identifikace informačního systému veřejné správy v systému IS o ISVS, pokud data uvedená v katalogizačním záznamu z takového systému pocházejí, identifikace se skládá z následujících atributů, které jsou povinně vyplněny v případě, že je identifikace ISVS uváděna
 - **K: Název ISVS (N)** - úplný název příslušného ISVS
 - **K: Odkaz na IS o ISVS (N)** - identifikátor ISVS v IS o ISVS
- **K: Vazba na ISDP (N)** - identifikace datových prvků, které jsou použity v datech uvedených v katalogizačním záznamu, identifikace každého datového prvku se skládá z následujících atributů, které jsou povinně vyplněny, pokud je identifikace datového prvku uvedena
 - **K: Název datového prvku (N)** - úplný název datového prvku
 - **K: Odkaz na ISDP (N)** - identifikátor datového prvku v ISDP
- **K: Související katalogizační záznamy (N)** - identifikace katalogizačních záznamů, které souvisejí s daným záznamem, identifikace se skládá z následujících atributů, které jsou povinně uváděny, pokud je identifikace souvisejícího záznamu uváděna
 - **K: Název záznamu (N)** - název odkazovaného katalogizačního záznamu
 - **K: Typ vazby (N)** - označení typu vazby mezi záznamy (např. „doplňuje“, „nahrazuje“, atd. Konkrétní typy vazeb budou definovány v rámci metodiky)
 - **K: URL katalogizačního záznamu (N)** - jednoznačný identifikátor odkazovaného záznamu používaný v rámci sítě internet
- **K: Klasifikace záznamu (P)** - klasifikace záznamů provedená pomocí přiřazení příslušných konceptů klasifikačních struktur a případně i volně tvořených textových značek
 - **K: Koncepty klasifikačních struktur (P)** - výčet konceptů klasifikačních struktur používaných pro klasifikaci dat uvedených v katalogizačním záznamu
 - **K: EUROVOC (P)** - seznam příslušných konceptů z taxonomie EUROVOC, u každého konceptu jsou uvedeny následující atributy
 - **Název konceptu (P)** - úplný název konceptu v jazyce záznamu
 - **Odkaz na koncept (P)** - odkaz na webovou stránku konceptu z webového portálu taxonomie
 - **K: CZ-NACE (N)** - seznam příslušných konceptů z taxonomie CZ-NACE, u každého konceptu jsou povinně uvedeny následující atributy, pokud je koncept použit
 - **Název konceptu (N)** - úplný název konceptu v češtině (v angličtině, pokud je jazyk záznamu angličtina)
 - **Odkaz na koncept (N)** - odkaz na webovou stránku konceptu z webového portálu taxonomie
 - **K: Volně tvořené značky (N)** - seznam použitých volně tvořených značek použitých pro klasifikaci dat v katalogizačním záznamu, značkou se zde myslí volně tvořený textový řetězec
- **K: Související geografické území (N)** - identifikace geografického území, ke kterému se vztahují data uvedená v katalogizačním záznamu, identifikace geografického území se skládá z následujících atributů, kdy je povinně uváděn alespoň slovní název území, pokud je uváděna identifikace geografického území

¹⁸ Bude automaticky doplněn z Portálu veřejné správy (<http://portal.gov.cz>).

- **K: Název/označení území (N)** - název či označení geografického území, hodnota je určena výběrem ze seznamu
- **K: Datové zdroje (P)** - identifikace datových zdrojů, které obsahují či zpřístupňují data uvedená v katalogizačním záznamu, pro každý datový zdroj jsou uvedeny následující atributy
 - **K: Název zdroje (P)** - úplný název datového zdroje
 - **K: Popis zdroje (P)** - text výstižně charakterizující datový zdroj
 - **K: Typ zdroje (P)** - označení typu zdroje (např. soubor, webová služba apod.)
 - **K: URL zdroje (P)** - URL, na kterém lze získat data z datového zdroje (např. URL datového souboru, URL koncového bodu webové služby apod.)
 - **K: Jazyk zdroje (N)** - jazyk, který je používán v rámci datového zdroje, pokud je to relevantní
 - **K: Formát zdroje (P)** - označení formátu datového zdroje (např. konkrétní formát datového souboru nebo obecně použitý formát dat) v uživatelsky srozumitelné podobě
 - **K: Označení formátu zdroje (P)** - textové označení formátu zdroje, např. koncovka souboru, pokud se jako označení formátu používá (XLS, CSV apod.) nebo jiný text obecně používaný pro označení formátu (XML)
 - **K: URL dokumentace formátu zdroje (P)** - URL dokumentu, který dokumentuje formát datového zdroje (např. dokumentace XML schématu v případě, že formátem datového zdroje je XML). V případě, že je dokumentace obsažena v několika dokumentech, je uvedeno URL pro každý jednotlivý dokument.
 - **K: URL datového schématu zdroje (P)** - URL souboru s datovým schématem, který specifikuje formát zdroje ve strojovém jazyce tak, aby byla možná automatická validace dat datového zdroje, pokud takový jazyk existuje (např. se jedná o jazyk XML Schema v případě formátu XML, RDFS/OWL¹⁹ v případě formátu RDF). V případě, že je schéma zapsáno ve více souborech, je uvedeno URL pro každý jednotlivý soubor.
 - **MIME type (P)** - standardizované označení formátu dat²⁰
 - **MIME type inner (N)** - standardizované označení formátu dat, který je obsažen v jiném formátu dat (např. označení formátu dat, která jsou zkomprimovaná některým z kompresních algoritmů)
 - **K: Velikost souboru (N)** - velikost souboru v bytech v případě, že datovým zdrojem je datový soubor
 - **K: Datum poslední úpravy (N)** - datum, kdy byl datový zdroj naposledy upraven
- **Datum poslední modifikace záznamu (P)** - datum, kdy byl katalogizační záznam naposledy upraven, může být doplněno automaticky katalogizačním nástrojem
- **Autor poslední modifikace záznamu (P)** - označení osoby, která provedla poslední změnu v katalogizačním záznamu, může být doplněno automaticky katalogizačním nástrojem
- **K: Kontakt na autora poslední modifikace záznamu (P)** - kontakt na osobu, která provedla poslední změnu v katalogizačním záznamu

Rozšiřující prvky katalogizačního záznamu jsou:

- **K: Označení území pomocí geokoordinátů (N)** - množina geokoordinátů potřebná pro zobrazení daného území na mapovém podkladě, aby nebylo potřeba tuto množinu doplňovat ručně, je doporučeno, aby kurátor dat při tvorbě či aktualizaci katalogizačního záznamu pouze zvolil příslušné geografické území (např. ze seznamu nebo na mapě) a k doplnění množiny geokoordinátů by na základě této volby došlo automaticky²¹.

¹⁹ OWL - Ontology Web Language [59]

²⁰ Toto standardizované označení formátu dat je dobře strojově zpracovatelné, nicméně není vždy zcela srozumitelné neznalému člověku. Proto je doporučeno tento atribut doplňovat automaticky na základě hodnoty atributu Formát zdroje a zpřístupňovat jeho hodnotu pouze při strojovém zpracování katalogizačního záznamu nebo na vyžádání koncovému uživateli, kterému by ale ve výchozím zobrazení katalogizačního záznamu mohl tento atribut zůstat skrytý.

²¹ V době implementace rozšířené funkcionality Datového katalogu budou využity služby poskytované geoportálem INSPIRE nebo základním registrem RÚIAN.

1.5.4 Doporučení ohledně klasifikačních struktur (slovníků a taxonomií)

Pro klasifikaci dat popsaných v katalogizačních záznamech je doporučeno využít následujících klasifikačních struktur (slovníků/taxonomií):

- EUROVOC²² – víceoborová taxonomie, existuje ve všech úředních jazycích EU, může být využita pro označení oblastí/oborů činnosti, kterých se data týkají,
- CZ-NACE²³ – primárně popisuje ekonomické činnosti a může být např. využita pro označení dat příslušnými ekonomickými činnostmi, kterých se týkají.

1.6 Správa a provoz katalogu dat VS ČR

Jak již vyplývá z variant budování Datového katalogu vymezených v předcházející části dokumentu, Datový katalog a podpůrné softwarové nástroje budou muset být spravovány a provozovány z jednoho centrálního místa. Tato kapitola se tak v návaznosti na doporučenou variantu budování Datového katalogu věnuje určení odpovědností za správu a provoz Datového katalogu a odpovídá tak na otázku, kdo bude spravovat a provozovat katalog dat VS ČR.

Správa Datového katalogu

Správou Datového katalogu bude dle kompetenčního zákona (zák. č. 2/1969 Sb., o zřízení ministerstev a jiných ústředních orgánů státní správy České republiky) pověřeno Ministerstvo vnitra České republiky, protože dle tohoto zákona je ústředním orgánem státní správy pro oblast informačních systémů veřejné správy (§12, odst. 1, písm. o)) a také plní koordinační úlohu pro informační a komunikační technologie (§12, odst. 6). Správou obsahu Datového katalogu se bude zabývat redaktor.

Provoz Datového katalogu

Z hlediska možných přístupů k této otázce může být provozovatelem Datového katalogu samotný orgán veřejné správy, který je zároveň jeho správcem, nebo může správce Datového katalogu smluvně pověřit jeho provozem jiný (i soukromý) subjekt.

1.7 Zajištění kvality obsahu datového katalogu

Kvalita obsahu Datového katalogu je významný aspekt, který ovlivňuje využitelnost katalogizačních záznamů uživateli katalogu. Nekvalitní záznamy obsahující nepřesné, neúplné nebo dokonce zavádějící údaje mohou totiž využitelnost katalogu značně komplikovat. Tato kapitola je tak věnována doporučením ohledně metod, technik a nástrojů pro zajištění kvality záznamů v katalogu.

Problematika zajištění kvality obsahu datového katalogu je obsáhlé téma. V tomto dokumentu jsou tedy mechanismy a postupy zajištění kvality popsány pouze rámcově. Detailně by se zajištěním kvality v rámci provozu Datového katalogu měla zabývat politika pro zajištění kvality obsahu Datového katalogu, která by měla vzniknout v rámci jeho implementace.

1.7.1 Kvalitativní atributy obsahu katalogu a jednotlivých záznamů

Je třeba rozlišit celkovou kvalitu Datového katalogu a potom také na kvalitu jednotlivých záznamů v katalogu. Co se týče katalogu, je možné se zaměřit na kvalitativní atributy uvedené v tabulce 2. Jak je uvedeno v tabulce, těchto kvalitativních atributů je možné pomocí současných technických dosáhnout jen manuální kontrolou záznamů.

Tabulka 2: Kvalitativní atributy obsahu katalogu jako celku

Kvalitativní atribut	Popis	Způsob dosažení
Unikátnost záznamů	V katalogu se nevyskytuje záznam, který by popisoval stejná data jako jiný záznam, nebo data, která jsou nadmnožinou dat popisovaných jiným záznamem.	Přímá manuální kontrola redaktory.

²² <http://eurovoc.europa.eu/drupal/?q=cs>

²³ <http://apl.czso.cz/iSMS/en/klasstru.jsp?kodcis=80004&cisjaz=203>

Kvalitativní atribut	Popis	Způsob dosažení
Úplnost katalogu	V katalogu jsou katalogizovány všechny záznamy, jejichž přítomnost v katalogu vyplývá z kapitoly 1.3	Přímá manuální kontrola správcem katalogu, příp. redaktory.

Dále je nutné pro každý záznam v Datovém katalogu zajistit kvalitativní atributy uvedené v tabulce 3. Některých kvalitativních atributů je možné dosáhnout automatizovanými softwarovými prostředky. Jiných atributů je možné dosáhnout pouze manuální kontrolou redaktorů záznamů. Způsob kontroly je taktéž uveden v tabulce.

Tabulka 3: Kvalitativní atributy katalogizačních záznamů

Kvalitativní atribut	Popis	Způsob dosažení
Relevance záznamu	Záznam je z pohledu účelu Datového katalogu smysluplný. Tzn., že záznam popisuje nějaká otevřená data VS ČR.	Manuální kontrola nových záznamů redaktorem.
Správnost a úplnost vyplněných údajů	Všechny v záznamu vyplněné údaje jsou správné a jsou vyplněny všechny povinné údaje, včetně údajů o časové platnosti.	Částečná automatizovaná kontrola během vyplňování (kontrola vyplnění povinného údaje + kontrola jednodušších pravidel, např. správnost formátu data). Porušení je hlášeno kurátorovi před založením záznamu. Manuální kontrola nových záznamů redaktorem po jejich založení kurátorem. Porušení nahlášené redaktorem je hlášeno poskytovateli dat.
Platnost odkazu na data	Odkaz uvedený v záznamu vede na existující datový soubor či fungující datové API.	Automatická kontrola v konfigurovatelném časovém intervalu. Při zakládání záznamu je porušení hlášeno kurátorovi. Pokud je odkaz zneplatněn později, je porušení hlášeno poskytovateli dat a redaktorovi.
Shoda vyplněných údajů s odkazovanými daty	Údaje v záznamu odpovídají v daném okamžiku aktuální skutečnosti o katalogizovaných datech.	Manuální kontrola redaktorem záznamu.
Správnost zařazení do klasifikačních struktur (např. EUROVOC)	Význam katalogizovaných dat odpovídá významu konceptů ze zvolené klasifikační struktury přiřazených záznamu.	Manuální kontrola redaktorem záznamu.

Jak ukazuje tabulka 3, možnosti automatizované kontroly kvality záznamů jsou poměrně omezené. Manuální kontrola redaktory však může být poměrně nákladná. Je proto doporučeno doplnit manuální redakci dvěma způsoby, které popisují dvě následující podkapitoly.

1.7.2 Využití koncových uživatelů (veřejnosti)

Koncoví uživatelé, průzkumníci, z řad veřejnosti mohou být cenným zdrojem zpětné vazby o kvalitě obsahu Datového katalogu. Tito uživatelé si během procházení obsahu Datového katalogu mohou všimnout nedostatků v kvalitě katalogizačních záznamů, které používají. Mohou např. identifikovat, že popis dat není srozumitelný, nebo že neodpovídá skutečnému obsahu dat. Uživatelé také pravděpodobně snadno odhalí záznamy s chybějícími údaji nebo duplicitní záznamy. Proto je doporučeno, aby Datový katalog umožnil koncovým uživatelům odeslat zpětnou vazbu ohledně obsahu Datového katalogu a umožnil tak, aby se uživateli identifikované nedostatky dostaly k osobám odpovědným za jejich nápravu.

1.7.3 Využití metod strojového učení

V budoucnu je možné pro vyšší automatizaci zajištění kvality využít i metod vyvíjených v rámci aplikovaného výzkumu na inženýrských fakultách univerzit v ČR. Metody strojového učení mohou např. automatizovat zařazování záznamů do klasifikačních struktur či rozpoznávání duplicitních záznamů v katalogu.

1.8 Zajištění bezpečnosti provozu datového katalogu

Stejně jako je třeba řešit otázku zajištění kvality obsahu Datového katalogu, je třeba řešit otázku zajištění bezpečnosti jeho provozu. Aby mohl být Datový katalog bezpečně provozován, musí být vymezeny požadavky na bezpečnost Datového katalogu a následně musí být zajištěny jak určité technické předpoklady katalogizačního nástroje, tak musí být definovány odpovídající procesy pro zajištění splnění těchto bezpečnostních požadavků. V neposlední řadě musí být adekvátně vyškolen příslušný personál.

Problematika bezpečnosti informačních systémů je velmi obsáhlé a komplexní téma. V této kapitole se tak zaměříme zejména na vymezení základních bezpečnostních požadavků na Datový katalog, protože ty představují předpoklad pro jeho bezpečný provoz. Detailně by se zajištěním bezpečnosti v rámci provozu Datového katalogu měla zabývat bezpečnostní politika Datového katalogu, která by měla vzniknout v rámci jeho implementace.

1.8.1 Možná ohrožení Datového katalogu a cíle z hlediska bezpečnosti

Základním cílem Datového katalogu je poskytovat důvěryhodné, aktuální a přesné údaje o existujících otevřených datech veřejné správy. Zatímco zejména aktuálnost a přesnost údajů v Datovém katalogu je doménou především zajištění kvality jeho obsahu, zajištění důvěryhodnosti jeho obsahu je spolu se zajištěním integrity a dostupnosti tohoto obsahu pro uživatele doménou bezpečnosti Datového katalogu.

Hlavní hrozby ohrožující bezpečnost Datového katalogu jsou uvedeny v následující tabulce.

Tabulka 4: Hlavní hrozby pro bezpečnost Datového katalogu

ID	Hrozba
H1	Vložení neautorizovaného obsahu, např. vložení katalogizačních záznamů osobou, která k tomu není oprávněna
H2	Neautorizovaná úprava obsahu, např. pozměnění katalogizačního záznamu osobou, která k tomu není oprávněna
H3	Neautorizované odstranění obsahu, např. smazání katalogizačního záznamu osobou, která k tomu není oprávněna
H4	Neautorizované zpřístupnění katalogizačního záznamu, který má být zpřístupněn až po provedení redakce
H5	Změna obsahu katalogizačního záznamu tak, že obsahuje nevhodný obsah nebo obsahuje odkazy na takovýto nevhodný obsah
H6	Ztráta obsahu ať už v důsledku selhání software, hardware, cíleného útoku, lidské chyby či nepředvídatelné události
H7	Znepřístupnění obsahu ať už v důsledku selhání software, hardware, cíleného útoku, lidské chyby či nepředvídatelné události
H8	Omezení možnosti zasílat legitimní upozornění na existující otevřená data veřejné správy ze strany uživatelů Datového katalogu v důsledku zahlcení kanálu pro zasílání těchto upozornění nerelevantními zprávami (spamem)

Bezpečnostním cílem datového katalogu tak je chránit jeho obsah před neautorizovanými zásahy (zejména před neautorizovaným přidáváním, modifikací a odstraněním tohoto obsahu) a před neautorizovaným zamezením přístupu k tomuto obsahu. U katalogizačních záznamů, které mají být

zpřístupněny až po provedení redakce, je cílem zabránit neautorizovanému zpřístupnění těchto záznamů před provedením redakce.

1.8.2 Bezpečnostní požadavky a opatření pro jejich naplnění

Jednotlivé bezpečnostní požadavky, problémové oblasti, které řeší a příklady možných způsobů naplnění těchto požadavků jsou uvedeny v tabulce 5. K vymezení požadavků je třeba ještě doplnit, že Datový katalog bude obsahovat pouze údaje o publikovaných otevřených datech veřejné správy. Datový katalog tak nebude obsahovat ani odkazovat na žádná data podléhající utajení a nevztahují se na něj tak požadavky dané zákonem č. 412/2005 Sb., o ochraně utajovaných informací a o bezpečnostní způsobilosti, respektive vyhláškou č. 523/2005 Sb., o bezpečnosti informačních a komunikačních systémů a dalších elektronických zařízení nakládajících s utajovanými informacemi a o certifikaci stínících komor.

Tabulka 5: Bezpečnostní požadavky na Datový katalog

ID	Požadavek	Související hrozby	Zajištění
BZ1	Obsah Datového katalogu (katalogizační záznamy) mohou vkládat, měnit a odstraňovat pouze oprávněné osoby	H1, H2, H3, H5	Jednoznačná identifikace a autentizace uživatele Vydávání přístupových údajů pouze pověřeným osobám Přiřazování rolí uživatelům a řízení přístupu k objektům na základě oprávnění náležejících rolí Zabezpečená komunikace s Datovým katalogem Pozn.: Jednoznačná identifikace a autentizace uživatele není vyžadována v případě koncových uživatelů (viz kap. 1.4.5)
BZ2	Redakci katalogizačních záznamů mohou provádět pouze oprávněné osoby	H1, H2, H3, H4, H5	Jednoznačná identifikace a autentizace uživatele Vydávání přístupových údajů pouze pověřeným osobám Přiřazování rolí uživatelům a řízení přístupu k objektům na základě oprávnění náležejících rolí Zabezpečená komunikace s Datovým katalogem
BZ3	Záznamy, které musí projít redakcí, nesmí být zveřejněny před provedením redakce	H1, H2, H3, H4, H5	Rozlišování redigovaných a neredigovaných záznamů Pouze redaktor může rozhodnout o zpřístupnění redigovaného záznamu Všechny vytvořené záznamy jsou odeslány redaktorovi Změny v záznamech opět posuzuje redaktor Zabezpečená komunikace s Datovým katalogem
BZ4	Činnosti spojené se zajištěním běžného chodu a údržby Datového katalogu mohou provádět pouze oprávněné osoby	H6, H7	Jednoznačná identifikace a autentizace uživatele Vydávání přístupových údajů pouze pověřeným osobám Přiřazování rolí uživatelům a řízení přístupu k objektům na základě oprávnění náležejících rolí Zabezpečená komunikace s Datovým katalogem

ID	Požadavek	Související hrozby	Zajištění
BZ5	V případě události vedoucí k nedostupnosti obsahu Datového katalogu musí být možné obnovit jeho plnou činnost do 24 hodin od zjištění nedostupnosti obsahu Datového katalogu (mimo plánované odstávky)	H7	Přiměřené postupy a technologické zajištění zálohování obsahu datového katalogu a obnovy jeho obsahu Přiměřené záložní softwarové a hardwarové prostředky pro zajištění chodu Datového katalogu Přiměřené personální zajištění chodu a údržby Datového katalogu
BZ6	V případě události vedoucí ke ztrátě obsahu Datového katalogu musí být možné obnovit jeho obsah alespoň do stavu odpovídajícímu stavu jednu hodinu před nastalou událostí	H6	Přiměřené postupy a technologické zajištění zálohování obsahu datového katalogu a obnovy jeho obsahu
BZ7	V případě události vedoucí ke ztrátě obsahu Datového katalogu musí být možné určit, které katalogizační záznamy byly ztraceny	H6	Požizování záznamů o vzniku a změnách katalogizačních záznamů Zabezpečení těchto záznamů před neoprávněnou změnou či zničením
BZ8	Ochrana všech veřejně přístupných kanálů, kterými mohou uživatelé zasílat zpětnou vazbu, zprávy či žádosti do Datového katalogu před zneužíváním (spamem)	H8	Využití technologií pro ochranu před automatizovaným odesláním zpráv těmito kanály Možnost omezit či blokovat přístup k těmto kanálům z určených uzlů či klientů v síti internet Nasazení technologických prostředků pro omezení množství operací, které může jeden klient/uživatel provést za vymezený časový úsek Pozn.: Např. využití technologie CAPTCHA24 a jí podobné
BZ9	Možnost určit původce změn v obsahu Datového katalogu, původce akcí souvisejících s redakcí obsahu a původce akcí spojených se zajištěním běžného chodu a údržby Datového katalogu	H1, H2, H3, H4, H5, H6, H7, H8	Nepřetržité znamenávání událostí souvisejících s bezpečností Datového katalogu do auditních záznamů (logů) Zabezpečení těchto záznamů před neoprávněnou změnou či zničením

1.8.3 Další doporučení pro zajištění bezpečnosti Datového katalogu a jeho provozu

Kromě výše uvedených požadavků na bezpečnost Datového katalogu a z nich vyplývajících doporučení pro jejich naplnění se jeví vhodné zpracovat celkovou bezpečnostní politiku Datového katalogu, která zohlední skutečné podmínky, ve kterých bude Datový katalog provozován. Jako součást této bezpečnostní politiky by měly být vymezeny zodpovědnosti jednotlivých rolí a osob za bezpečnost Datového katalogu a měly by být vymezeny procesy pro řešení bezpečnostních incidentů. V rámci školení a vzdělávání uživatelů Datového katalogu by měly být uživatelé seznámeni i s obsahem této bezpečnostní politiky.

²⁴ CAPTCHA - Program pro generování testů, ve kterých mohou lidé snadno uspět, ale počítače nikoli. Tyto testy slouží k např. rozlišování, zda se určitou webovou službu snaží použít člověk nebo jiný počítačový program a může být použit k obraně proti automatickému přidávání nerelevantních příspěvků (spamu) [22]

Vzhledem k tomu, že veřejnosti bude umožněno odesílat zpětnou vazbu k Datovému katalogu a další zprávy (upozornění na data), je vhodné nasadit nejen technologické prostředky pro snížení rizika zneužívání těchto kanálů, ale umístit na webové stránky Datového katalogu upozornění před tímto zneužíváním.

Datový katalog bude přístupný na internetu. Je tedy vhodné aplikovat obecné zásady zabezpečení webových aplikací a služeb a katalogizační nástroj by také měl dodržovat obecně uznávané zásady pro bezpečnost webových aplikací (viz např. [55]).

Protože se Datový katalog může stát terčem útoků směřujících k omezení jeho fungování a znepřístupnění jeho obsahu, je vhodné zvážit nasazení přiměřené ochrany proti útokům typu DoS²⁵ a DDoS²⁶.

1.8.4 Přístup uživatelů do Datového katalogu a jejich oprávnění

V rámci veřejné správy je budován systém Jednotného identitního prostoru (JIP), který zajišťuje jednoznačnou identifikaci a autentizaci uživatelů do informačních systémů veřejné správy [50]. Systém JIP lze využít pro zajištění přístupu uživatelů z řad veřejné správy k Datovému katalogu a k jejich jednoznačné identifikaci a autentizaci. Využití JIP lze doporučit, aby nebylo třeba zavádět nový proces pro přidělování přístupových údajů k Datovému katalogu²⁷. Protože ale v současné době není systém JIP využíván všemi pracovníky VS ČR, lze po dobu, než využívání systému JIP dosáhne plánovaného rozsahu, pro zajištění jednoznačné identifikace a autentizace uživatelů Datového katalogu využít princip založený na kvalifikovaných certifikátech, který je používán v Informačním systému o datových prvcích (ISDP, viz vyhl. č. 469/2006 Sb.). Kvalifikovaný certifikát má přiděleno značné množství pracovníků VS ČR a využití kvalifikovaných certifikátů tak představuje další z možností, jak umožnit pracovníkům VS ČR přístup do Datového katalogu bez zavádění nového procesu pro přidělování přístupových údajů. V budoucnu by ale přístup do Datového katalogu měl být zajištěn s využitím Jednotného identitního prostoru.

Aby byl naplněn cíl Datového katalogu, měl by být přístup kurátorů dat k Datovému katalogu co nejjednodušší. V opačném případě by složitost přístupu mohla spíše kurátory od provádění katalogizace odradit. Proto je navrženo, aby bylo umožněno každému pracovníkovi VS ČR přistupovat k Datovému katalogu v roli kurátora dat, tj. každý pracovník VS ČR by tak byl osobou oprávněnou k vkládání, úpravě a odstraňování obsahu Datového katalogu (viz BZ1). Pověření kurátora dat poskytovatelem tak zůstává na úrovni organizačního postupu v rámci subjektu poskytovatele dat. Jednoznačná identifikace a autentizace těchto pracovníků musí být samozřejmě i nadále zajištěna, aby došlo k jejich odlišení od koncových uživatelů Datového katalogu. V případě redaktorů bude osobou oprávněnou k provádění redakce obsahu Datového katalogu (viz BZ2) pouze osoba pověřená správcem Datového katalogu, který také zajistí přidělení příslušné role těmto osobám v Datovém katalogu.

Každý kurátor bude oprávněn založit katalogizační záznam a u každého katalogizačního záznamu je také uvedeno, kdo je poskytovatelem tam uvedených dat (viz kap. 1.5.3). Kurátor dat nebude oprávněn upravovat nebo zneplatňovat libovolné katalogizační záznamy, ale pouze záznamy o datech poskytovatele, jehož zaměstnancem kurátor dat je. Zároveň každý kurátor, který je zaměstnancem poskytovatele dat²⁸, bude moci upravovat či zneplatnit libovolný záznam o datech tohoto poskytovatele. Během přihlášení tak musí dojít i k předání informace o tom, kdo je zaměstnavatelem kurátora.

1.9 Cílová softwarová architektura nástroje pro Datový katalog

Základní softwarová architektura nástroje pro datový katalog je poměrně jednoduchá a je znázorněna na obrázku 4. Samotný nástroj se skládá ze dvou komponent:

²⁵ DoS - Denial of Service - útok s cílem způsobit nedostupnost cíle útoku (např. webové stránky nebo služby) pro jeho legitimní uživatele [61]

²⁶ DDoS - Forma útoku DoS, ve kterém se více systémů snaží zaplavit svými požadavky cíl útoku a způsobit tak jeho nedostupnost pro legitimní uživatele [61]

²⁷ Jak již bylo uvedeno výše, koncoví uživatelé budou moci využívat Datový katalog bez nutnosti registrace a přihlášení a tudíž pro ně nebude třeba zajistit přidělování přístupových údajů.

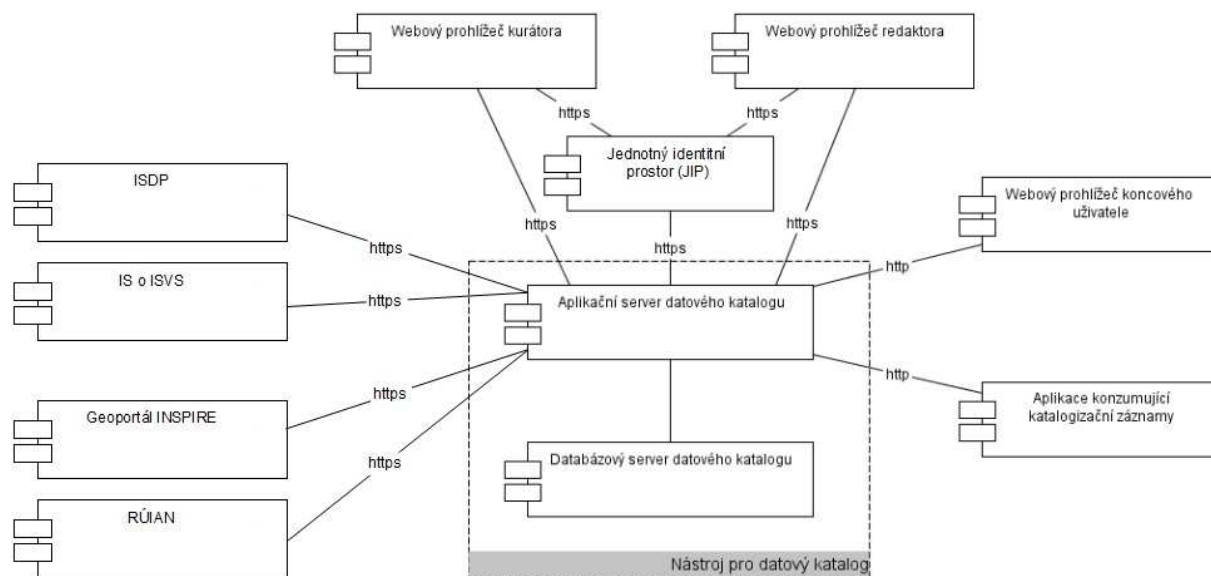
²⁸ Pokud je mezi kurátorem a poskytovatelem dat jiný vztah než zaměstnanecký, který zároveň opravňuje kurátora ke katalogizaci jeho dat a k přístupu k Datovému katalogu coby pracovníka poskytovatele, může i kurátor v takovémto vztahu k poskytovateli zakládat, upravovat a zneplatňovat záznamy o datech poskytovatele.

- databázový server datového katalogu - umožňuje ukládání, modifikaci a dotazování nad zadanými katalogizačními záznamy
- aplikační server datového katalogu - přijímá požadavky od uživatelů prostřednictvím protokolů HTTP či HTTPS, odesílá uživatelům výsledek požadavku v podobě HTML stránky a zajišťuje publikaci katalogizačních záznamů v podobě otevřených dat

Znázornění architektury na obrázku 1 také ukazuje další softwarové komponenty. Jedná se o webové prohlížeče uživatelů (tj. kurátorů, redaktorů a koncových uživatelů), které využívají pro přístup k datovému katalogu. Dále jsou znázorněny externí aplikace, které konzumují katalogizační záznamy v katalogu pro své potřeby. To je umožněno díky tomu, že samotný obsah katalogu, tj. katalogizační záznamy publikuje aplikační server jako otevřená data.

Architektura také znázorňuje vazbu na ISDP a IS o ISVS. Znázorněna je také vazba na geoportál INSPIRE a Základní registr územní identifikace, adres a nemovitostí. Dle navržené struktury katalogizačního záznamu by tyto systémy do budoucna mohly poskytovat geokординáty pro označení území, ke kterému se vztahují data popsána v katalogizačních záznamech.

Dále je znázorněna komponenta zajišťující přihlašování uživatelů (kurátorů a redaktorů) v cílovém stavu prostřednictvím Jednotného identitního prostoru.



Obrázek 1: Znázornění základní softwarové architektury nástroje pro katalogizaci dat

1.10 Požadavky na nástroj pro datový katalog

Na základě vyhodnocení doporučené varianty budování Datového katalogu a vyhodnocení doporučení ohledně řešení jednotlivých otázek diskutovaných v přecházejících kapitolách, jsou v této kapitole souhrnně definovány požadavky na nástroj pro Datový katalog. Kapitola tak odpovídá na stejně znějící otázku. Dále jsou v této kapitole uvedena doporučení ohledně pořízení nástroje pro Datový katalog, čímž kapitola odpovídá i na otázku, jak by měl být nástroj pro datový katalog pořízen.

V kapitole 2. plné verze Koncepce je uveden přehled existujících nástrojů pro katalogizaci dat. V návaznosti na doporučenou variantu řešení Datového katalogu a existenci samostatných nástrojů je doporučeno realizovat výběrové řízení na nástroj a jeho dodavatele. Doporučená výběrová kritéria a váhy²⁹ jednotlivých skupin kritérií:

²⁹ Pokud se ve skupině výběrových kritérií objevují obligatorní kritéria, váha se týká pouze fakultativních kritérií.

1. Cena (60 %)
 - a. cena pořízení
 - b. cena školení pro správce a redaktory Katalogu dat a on-line příruček
 - c. cena úprav nástroje
 - i. zajištění základních funkčních požadavků ad 2 a) - pokud bude třeba nástroj upravovat
 - ii. zajištění nadstavbových funkčních požadavků - ad 2 b) - pokud bude třeba nástroj upravovat
 - d. cena za provoz
 - i. za první rok provozu nástroje
 - ii. kalkulace ceny v dalších letech
2. Základní funkční požadavky - obligatorní kritéria
 - a. Jazyková verze v CZ
 - b. Vkládání, editace a zneplatňování záznamů, které mají strukturu vymezenou v kapitole 1.5.3
 - c. Publikace pouze vybraných (redaktorem schválených) záznamů
 - d. Vyhledávání a procházení obsahu Datového katalogu
 - e. Publikace obsahu datového katalogu ve formě otevřených a propojitelných dat (s využitím technologií Linked Data [5])
 - f. Možnost určit povinné a volitelné atributy záznamu v Datovém katalogu a vyžadovat vyplnění povinných atributů
 - g. Možnost využívat externích klasifikačních struktur (např. EUROVOC)
 - h. Možnost integrovat automatizované nástroje (např. nástroje pro kontrolu kvality)
 - i. Existence definovaného aplikačního programátorského rozhraní (API)
 - j. Možnost zadávání námětů na zvýšení kvality obsahu Datového katalogu od koncových uživatelů
 - k. Redakce záznamů v Datovém katalogu, včetně možnosti určení atributů, které je třeba opravit a notifikace kurátora a poskytovatele dat
 - l. Statistiky a analýzy záznamů v Datovém slovníku (četnosti záznamů podle jednotlivých prvků klasifikačních struktur, vývoj v čase a dle jednotlivých poskytovatelů dat)
 - m. Jednotná identifikace a autentizace uživatelů pomocí systému JIP / KAAS (jednotného identitního prostoru)
3. Nadstavbové funkční požadavky (20 %) - fakultativní kritéria
 - a. Možnost vkládat jako volitelné atributy - geokoordináty pro určení příslušnosti dat odkazovaných v záznamu v Datovém katalogu k územní oblasti (vazba na RÚIAN nebo geoportál INSPIRE)
 - b. Vizualizace dat (např. propojení s mapovými náhledy)
 - c. Možnost využívat externích klasifikačních struktur (např. EUROVOC) aniž by bylo nutné kopírovat číselníky do nástroje
 - d. Vyhledávání podle geografického umístění
 - e. Více jazyčnost nástroje - možnost zachytit u jednoho katalogizačního záznamu více jazykových verzí
4. Bezpečnost - obligatorní kritéria
 - a. viz kapitola 1.8.2
5. Výkonnost a škálovatelnost z hlediska objemu zpracovávaných záznamů (20 %)

1.11 Uživatelé datového katalogu a podmínky jeho použití

Tato kapitola se snaží odpovědět na otázku, kdo bude moci datový katalog využívat a za jakých podmínek. V návaznosti na zvolenou variantu budování Datového katalogu je v rámci této kapitoly také diskutováno zapojení veřejnosti do plnění datového katalogu. Kapitola tak také odpovídá na otázku, zda budou moci do datového katalogu přispívat občané ČR z řad veřejnosti.

Jednotlivé role související s tvorbou, redakcí, údržbou a provozem Datového katalogu jsou popsány v kapitole 1.4. Zde je pozornost soustředěna pouze na roli koncového uživatele Datového katalogu, tj. na roli, která primárně hledá potřebná otevřená data VS. V této roli mohou vystupovat jak soukromé osoby (primárně občané ČR), tak i právní subjekty a orgány veřejné správy.

Zpřístupnění záznamů v Datovém katalogu bude pro koncové uživatele bez poplatků a bez nutnosti registrace.

Datový katalog bude zpřístupněn v rámci domény gov.cz.

1.11.1 Zapojení veřejnosti do plnění datového katalogu

Jak vyplývá z vymezení jednotlivých rolí, veřejnost se nebude moci přímo podílet na vytváření katalogizačních záznamů. Nicméně uživatelé z řad veřejnosti (běžní uživatelé) budou moci jednak zasílat obecné zprávy, pomocí kterých budou moci poskytovat zpětnou vazbu k Datovému katalogu, a dále budou moci zasílat upozornění na publikovaná otevřená data. Díky tomu bude moci široká veřejnost upozorňovat na existenci publikovaných otevřených dat, která nejsou katalogizována v Datovém katalogu a alespoň tímto způsobem budou moci přispět k zlepšování úplnosti Datového katalogu.

1.12 Vzdělávání v oblasti katalogizace otevřených dat VS ČR

Aby mohly jednotlivé orgány veřejné správy efektivně provádět katalogizaci otevřených dat VS ČR, je třeba, aby příslušní pracovníci těchto orgánů byli seznámeni

- s principy otevřených dat VS ČR,
- s metodickými pokyny pro katalogizaci dat,
- s pravidly práce s katalogizačním nástrojem.

Vzdělávání v oblasti katalogizace dat VS se primárně zaměří na role správce Datového katalogu a redaktory. Tyto dvě role musí projít specializovaným školením, které je seznámí se všemi vykonávanými činnostmi a ovládáním katalogizačního nástroje. Školení pro správce a redaktory Datového katalogu zajistí dodavatel nástroje ve spolupráci s autory Metodiky katalogizace otevřených dat VS ČR.

Role poskytovatele dat, kurátora dat a koncového uživatele budou mít k dispozici on-line přístupné metodické příručky a návod pro práci s katalogizačním nástrojem.

1.13 Legislativní změny související s katalogizací otevřených dat

Tato kapitola se zaměřuje na řešení otázky, jaké budou potřebné či vhodné legislativní

změny pro fungování datového katalogu a efektivní katalogizaci otevřených dat. Legislativní aspekt katalogizace otevřených dat veřejné správy byl již v části Analýza současného stavu a varianty řešení v plné verzi Koncepce diskutován v souvislosti s licencemi pro otevřená data a ve vazbě na související legislativu. V této kapitole je tak diskutováno, zda bude efektivní zajištění katalogizace otevřených dat VS ČR vyžadovat podporu ze strany legislativy a případně, jakou podobu by tato podpora měla mít.

Ačkoli se tato Koncepce věnuje katalogizaci otevřených dat, je třeba ji řešit v širším kontextu problematiky otevřených dat jako takových. Aby bylo možné budovat Datový katalog, je třeba nejprve vymezit, jaká data veřejné správy jsou otevřená. Tímto vymezením se zabývala podrobněji kapitola 1.3, kde je také uvedeno, že jednou z podmínek otevřenosti dat veřejné správy je vymezení a jasné uvedení podmínek jejich užití (licence) s minimem omezení. Následující text tak svým zaměřením přesahuje otázky spojené čistě s katalogizací otevřených dat, ale věnuje se i otázce podmínek užití (licencování) otevřených dat veřejné správy a diskutuje i vymezení pojmu otevřená data veřejné správy v právních předpisech.

1.13.1 Podmínky užívání otevřených dat VS ČR

Jak již bylo uvedeno v předcházející části dokumentu, v ČR v současné době není uplatňován jasný a jednotný přístup k licencování dat veřejné správy. Jasné licencování či jasné uvedení právních podmínek užití dat je jedním z principů otevřených dat [54] a bylo by tedy vhodné, aby i pro otevřená data veřejné správy v ČR byly určeny jasné podmínky jejich užití.

Aby bylo vymezeno, co znamená „otevřenost“ v pojmu otevřená data ve vztahu k jejich užití, ale i v dalších oblastech, iniciovala nadace Open Knowledge Foundation vznik tzv. Definice otevřenosti (Open Definition), která blíže definuje 11 bodů, které by měly být splněny u podmínek použití dat, aby je bylo

možné považovat za otevřená data [36]. Dle [40] lze to nejdůležitější shrnout následujícím způsobem. Data jsou otevřená, pokud jsou:

- dostupná v celku³⁰, za cenu nepřesahující oprávněné náklady na jejich reprodukci, ve formě umožňující jejich úpravu a nejlépe stažitelná z internetu;
- dostupná za podmínek umožňujících jejich zpracování, další šíření a kombinování s jinými daty;
- dostupná za podmínek umožňujících jejich zpracování a další šíření každému a pro jakékoli použití (včetně komerčního).

Obecně by tedy podmínky využití dat veřejné správy, která mají být publikována jako otevřená data, měly vyhovovat výše uvedeným principům, respektive principům Definice otevřenosti.

Předběžná podoba zprávy evropského projektu LAPSI [47], který se zabývá právními otázkami opětovného použití informací veřejného sektoru, uvádí, že vzhledem k silné ochraně duševního vlastnictví a existenci zvláštního práva pořizovatele databáze v rámci Evropské unie, užívají členské státy zpravidla určitou podobu licence pro udělení oprávnění užívat data veřejné správy. Citovaná zpráva se také přiklání k názoru, že využívání mezinárodně uznávaných licenčních smluv je výhodnější z hlediska mezinárodního využití informací veřejného sektoru.

Jeví se tedy vhodné formulovat jednotný přístup ke stanovování podmínek využití otevřených dat veřejné správy a tam, kde je to možné a vhodné, poskytnout uživatelům dat licenci k jejich užití. Aby byl licenční přístup v rámci VS ČR sjednocen, je doporučeno vytvořit vzorové licence pro otevřená data VS ČR.

Vzhledem k tomu, že je doporučeno, aby i samotné záznamy v Datovém katalogu byly publikovány s využitím principů otevřených dat, je třeba zajistit nejen jejich technickou otevřenost, ale i právní. Z toho vyplývá, že podmínky využití katalogizačních záznamů by také měly odpovídat Definici otevřenosti a pokud se ukáže, že je vhodné poskytnout práva k využití katalogizačních záznamů ve formě licence, tak určit či vytvořit tuto licenci.

Aby se zamezilo situacím, kdy si uživatelé dat nejsou jisti, za jakých podmínek je mohou využívat, jeví se vhodné, aby u publikovaných dat veřejné správy bylo vždy jednoznačně a zřetelně uvedeno, za jakých podmínek je lze využít. Vzhledem k tomu, že ne všechna publikovaná data veřejné správy musí představovat otevřená data, je vhodné podmínky užití dat uvádět u všech dat veřejné správy, nejen u dat otevřených.

Licence Creative Commons představují sadu licencí pro díla chráněná autorským právem [13]. Tyto licence mají nejen svoje slovní znění, ale obsah práv a povinností vyplývající z jednotlivých licencí je vyjádřen i pomocí jednoduché sady piktogramů. Toto umožňuje uživatelům děl, které tyto licence využívají, snadno rozlišit, např. zda mohou či nemohou šířit odvozená díla. Na základě toho lze doporučit, aby bylo zvaženo vytvoření obdobné jednoduché sady piktogramů, které budou sloužit k označení dat veřejné správy v závislosti na tom, za jakých podmínek je možné je využívat. Lze si tak představit např. minimální sadu piktogramů pro rozlišení otevřených dat a dat, jejichž podmínky užití nevyhovují Definici otevřenosti („uzavřená data“). Piktogramy by bylo možné umístit např. na webové stránky orgánů veřejné správy vedle odkazů na stažení dat, aby uživatelé věděli, zda si stahují otevřená data či nikoli.

1.13.2 Pojem otevřených dat

Pojem otevřená data není v současnosti v platné legislativě vymezen. V Akčním plánu České republiky „*Partnerství pro otevřené vládnutí*“ [2] se nicméně hovoří o otevření 10 vybraných datových zdrojů veřejné správy, přičemž Akční plán rámcově pojem otevřená data vymezuje. Pokud by se Vláda České republiky rozhodla v rámci dalšího pokračování Partnerství pro otevřené vládnutí nebo v rámci jiných aktivit pro otevření dalších datových zdrojů veřejné správy, bylo by patrně vhodné, aby došlo nejen k harmonizaci přístupu k určování podmínek užití otevřených dat, ale aby pojem otevřená data jako takový byl vymezen a to s ohledem i na další atributy otevřených dat uvedené v kapitole 1.3. Vymezení pojmu otevřená data by rozhodnutí o otevření datového zdroje dalo konkrétní obsah, tj. bylo by zřejmé, že otevření datového zdroje znamená realizaci kroků nezbytných pro to, aby data z daného datového zdroje splňovala podmínky otevřených dat.

³⁰ Požadavek na dostupnost všech dat tvořících určitý celek je zde z důvodu, že za otevřená data se nepovažují případy, kdy je k volnému použití poskytnuta např. pouze ukázka či vybraná část dat [36]. Ve skutečnosti bude třeba vždy posoudit, co je tímto celkem a zda je opravdu možné a účelné zpřístupňovat všechna data z určité databáze (viz bod „úplnost“ v kap. 1.3)

Zákon č. 106/1999 Sb. upravuje podmínky přístupu k informacím veřejného sektoru a zpracovává Směrnici 2003/98/ES o opakovaném použití informací veřejného sektoru. V současné době je připravována novela této Směrnice [18]. Pokud bude tato novela přijata, vyžádá si její implementace novelu zákona č. 106/1999 Sb. Pracovní podoba návrhu novely Směrnice v anglickém jazyce [17] sice v textu samotných článků Směrnice nepoužívá pojem otevřená data, tento pojem je ale použit v rámci odůvodnění³¹. Jednou z možností, jak pojem otevřená data, respektive otevřená data veřejné správy, vymezit, je uvést jeho definici v zákoně č. 106/1999 Sb. V tomto případě je třeba zajistit, aby vymezení pojmu otevřená data nebylo v rozporu s výše uvedenou Směrnicí, respektive s jejím možným budoucím zněním.

1.13.3 Legislativní ukotvení katalogizace otevřených dat VS ČR

Má-li být katalogizace otevřených dat veřejné správy v centrálním Datovém katalogu povinná pro orgány veřejné správy včetně územních samosprávných celků, bude třeba, aby na obecné úrovni byla tato povinnost uložena zákonem obdobně, jako je tomu v případě povinnosti předávat údaje o číselnících a datových prvcích do ISDP a povinnosti předávat údaje o spravovaných informačních systémech veřejné správy do IS o ISVS, kdy tyto povinnosti ukládá příslušným orgánům veřejné správy zákon č. 365/2000 Sb., o informačních systémech veřejné správy a o změně některých dalších zákonů. Důvodem je zásada, že povinnosti lze ukládat toliko na základě zákona a že kompetence správních orgánů lze založit výlučně zákonem.

Protože byla k realizaci doporučena varianta s centrálním datovým katalogem, je vhodné, aby i přístup ke stanovování podmínek pro užití otevřených dat veřejné správy byl harmonizován. Pokud by se přístupy k určování práv a povinností při užívání dat veřejné správy často lišily mezi jednotlivými orgány veřejné správy, znesnadnilo by to využití otevřených dat veřejné správy, protože uživatelé by byli nuceni neustále zkoumat odchylky v úpravách podmínek užití dat. Z tohoto důvodu bylo výše doporučeno vytvoření jednotného přístupu ke stanovování podmínek užití otevřených dat veřejné správy. Stejně jako v případě povinnosti katalogizovat otevřená data, bude třeba harmonizaci přístupu k určování podmínek užívání otevřených dat i na úrovni územních samosprávných celků realizovat formou zákona.

Ve světě se oblast katalogizace otevřených dat veřejné správy dynamicky vyvíjí a lze očekávat, že i katalogizace otevřených dat VS ČR bude procházet vývojem. Proto by podrobnosti provádění katalogizace dat (metodika katalogizace) a podrobnosti aplikace harmonizovaného rámce pro určování podmínek užívání otevřených dat měly určovat prováděcí vyhlášky, aby nebylo třeba častých změn zákonných norem.

1.14 Vazba na připravovaný portál Publicdata.eu

Jak již bylo v předcházející části dokumentu uvedeno, připravovaný portál Publicdata.eu by měl představovat jednotné místo pro vyhledávání dat veřejné správy členských zemí Evropské unie. S ohledem na to je třeba se zabývat otázkou, jak bude zajištěna vazba na připravovaný portál Publicdata.eu. V návaznosti na doporučenou variantu budování Datového katalogu tak tato kapitola diskutuje možné přístupy k zajištění integrace s portálem Publicdata.eu a dává doporučení ohledně vhodných přístupů.

1.14.1 Katalogizační nástroj

Portál Publicdata.eu je budován na základě katalogizačního nástroje CKAN. Proto je z pohledu vazby na tento portál vhodné vybudovat Datový katalog VS ČR také na základě nástroje CKAN. Nicméně pokud bude Datový katalog VS ČR vybudován na základě jiného nástroje nebo bude vytvořen proprietární nástroj, bude vždy možné vybudovat vhodné rozhraní pro komunikaci s portálem Publicdata.eu, neboť i řada ostatních datových katalogů v rámci EU je vybudována na základě proprietárních nástrojů. Tuto skutečnost musí portál Publicdata.eu reflektovat.

Rozhraní bude muset vybudovat správce Datového katalogu ČR.

³¹ Do českého překladu pracovní verze novely Směrnice [18] byl pojem otevřená data (open data) přeložen jako zpřístupnění údajů.

1.14.2 Taxonomie pro klasifikaci katalogizačních záznamů

Pro zajištění interoperability s portálem Publicdata.eu, příp. jinými katalogy otevřených dat, je nutné zajistit interoperabilitu taxonomií určených pro klasifikaci katalogizačních záznamů. Pokud by si Katalog otevřených dat VS ČR zvolil svoji proprietární taxonomii, bylo by zajištění tohoto druhu interoperability velmi nákladné (jak prvotní provázání různých taxonomií, tak i pozdější údržba provázání). Z tohoto důvodu je doporučeno využít existující v EU běžně používané taxonomie, které jsou zmíněny v kap. 1.5.4.

1.15 Scénáře užití datové katalogu VS ČR

Scénáře užití datové katalogu VS ČR uceleně popisují zamýšlené použití Datového katalogu, které vychází z doporučení k řešení jednotlivých otázek spojených s katalogizací otevřených dat VS ČR. V rámci těchto scénářů jsou popsány situace, v rámci kterých budou jednotlivé kategorie uživatelů (viz kapitola 1.4 popisující jednotlivé role) využívat Datový katalog.

V plné verzi Koncepce jsou popsány detailně následující scénáře:

1. Pověření redaktora a přidělení příslušné role v Datovém katalogu
2. Zrušení pověření redaktora
3. Založení záznamu v Datovém katalogu (zahrnuje i redakci záznamu)
4. Úprava záznamu v Datovém katalogu (změna údajů v existujícím katalogizačním záznamu)
5. Zneplatnění záznamu v Datovém katalogu
6. Obnovení platnosti záznamu v Datovém katalogu
7. Procházení Datového katalogu a poskytnutí zpětné vazby
8. Vypořádání zpětné vazby a podnětů od koncových uživatelů

2 Plán realizace koncepce

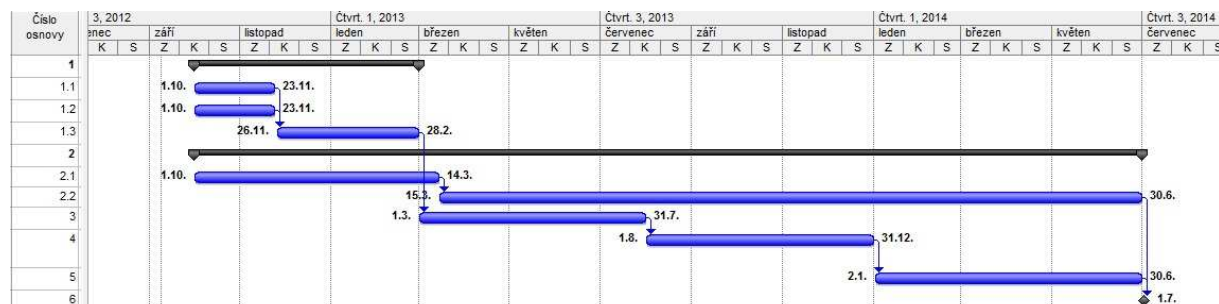
V této části je představen plán realizace Koncepce katalogizace otevřených dat VS ČR. Realizace Koncepce bude zajištěna v několika etapách. Kromě časového harmonogramu realizace těchto etap jsou v této části vymezeny i jejich hlavní výstupy, odhad jejich pracnosti a nákladů potřebných na jejich realizaci.

2.1 Etapy a hlavní skupiny činností

V této kapitole jsou vymezeny hlavní skupiny činností, které je třeba provést, aby byla realizována navržená Koncepce. Etapy mohou být realizovány formou samostatných projektů.

Tabulka 6: Harmonogram etap a hlavních skupin činností

ID	Etapa	Zahájení	Dokončení	Návaznost
1	Zjednodušený Datový katalog	1.10. 2012	28.2. 2013	
1.1	Nastavení nástroje	1.10. 2012	23.11. 2012	
1.2	Vytvoření Metodiky katalogizace	1.10. 2012	23.11. 2012	
1.3	Ověřovací provoz zjednodušeného Datového katalogu = Katalogizace oblastí určených v Akčním plánu	26.11.2012	28.2. 2013	1.1; 1.2
2	Příprava legislativních podmínek pro otevřená data³²	1.10. 2012	30.6.2014	
2.1	Příprava návrhu legislativních změn	1.10. 2012	14.3.2013	
2.2	Projednání a schvalování návrhu změn právních norem	15.3.2013	30.6.2014 ³³	2.1
3	Výběr nástroje, jeho dodavatele a provozovatele	1.3. 2013	31.7 2013	1.3
4	Implementace vybraného nástroje	1.8. 2013	31.12.2013	3
5	Ověřovací provoz plnohodnotného nástroje	2.1. 2014	30.6.2014	4
	Zahájení provozu	1.7.2014		2; 5



Obrázek 2: Harmonogram etap a hlavních skupin činností

³² Předpokládá se, že příprava legislativních podmínek bude realizována interně pracovníky příslušných orgánů veřejné správy. Není tak uváděna pracnost a náklady souvisejících skupin činností.

³³ Účinnost od 1.7.2014

2.2 Popis etap realizace koncepce

V následujícím textu je uveden popis jednotlivých etap realizace Koncepce.

2.2.1 Zjednodušený Datový katalog

Česká republika se v Akčním plánu „*Partnerství pro otevřené vládnutí*“ [2] zavázala realizovat katalog dat veřejné správy do 31.3.2013. Protože realizace plnohodnotného Datového katalogu, který by odpovídal všem doporučením Koncepce, bude vyžadovat více času, než bylo odhadováno v době tvorby Akčního plánu, jeví se vhodné začít katalogizační aktivitu alespoň realizací jeho zjednodušené formy (zjednodušený Datový katalog), jehož účelem by mělo být jak naplnění závazku Akčního plánu, tak ověření navržené Metodiky katalogizace otevřených dat VS ČR. Náplní této etapy je tak nastavení zvoleného nástroje pro Datový katalog, vytvoření Metodiky katalogizace otevřených dat VS ČR a ověřovací provoz.

2.2.1.1 Nastavení zvoleného nástroje

Pro realizaci zjednodušeného Datového katalogu se jeví vhodné využít některého z volně dostupných katalogizačních nástrojů, aby byly minimalizovány náklady spojené s jeho realizací. Jednou z možností je využití nástroje CKAN, jehož veřejně dostupná instance na adrese <http://cz.ckan.net> byla již použita akademickou sférou pro realizaci neoficiálního prototypu katalogu dat veřejné správy.

Zvolený nástroj pro zjednodušený Datový katalog bude třeba nastavit tak, aby vyhovoval minimálním požadavkům realizace tohoto zjednodušení. Realizace zjednodušeného Datového katalogu by se měla zaměřit na splnění závazku Akčního plánu v oblasti katalogizace dat veřejné správy a na ověření fungování navržené Metodiky katalogizace otevřených dat VS ČR. Nepředpokládáme tak, že v rámci zjednodušeného Datového katalogu budou realizována všechna doporučení Koncepce. Realizovány by tak měly být pouze ty úpravy a nastavení zvoleného katalogizačního nástroje, které budou nezbytné pro naplnění cíle Akčního plánu.

2.2.1.2 Vytvoření Metodiky katalogizace

Aby byla zajištěna konzistence a shodnost vytváření katalogizačních záznamů napříč subjekty veřejné správy, je vhodné rozpracovat postup katalogizace představený v této Koncepci do podrobnější metodiky, která bude určovat, jak by kurátoři dat měli při tvorbě katalogizačních záznamů postupovat a jak by měli postupovat redaktoři při redakci těchto záznamů. Metodika katalogizace otevřených dat VS ČR by tak měla představovat ucelenou sadu postupů a doporučení pro popis dat, jejich klasifikaci pomocí konceptů určených katalogizačních struktur, pro aplikaci dalších číselníků používaných v rámci katalogizačního záznamu a pro posuzování katalogizačních záznamů během redakce. Součástí metodiky by měl být také jasný návod, který umožní kurátorům dat určit, zda je třeba určitá data katalogizovat či nikoli.

2.2.1.3 Ověřovací provoz zjednodušeného Datového katalogu

V rámci ověřovacího provozu zjednodušeného Datového katalogu by mělo dojít ke katalogizaci dat, respektive datových zdrojů, které určuje Akční plán jako data, která mají být v horizontu jeho řešení otevřena, a tím by mělo dojít k naplnění jeho cílů v oblasti katalogizace dat. Dle [2] se jedná o následující datové zdroje:

- Obchodní rejstřík
- Insolvenční rejstřík
- Informační systém o veřejných zakázkách
- Výsledky voleb
- Registr aktivních legislativních prací - RALP (resortní nástroje podporující transparentnost výkonu státní správy a zapojení veřejnosti)
- Finanční statistika - státní dluh
- Finanční statistika - vládní finanční statistika
- ÚFIS - účetní záznamy a finanční údaje z CSÚIS
- Online přístup k údajům o financování politických stran
- Centrální registr dotací

Ověřovací provoz by také měl sloužit k ověření navržené Metodiky katalogizace otevřených dat VS ČR, protože katalogizace výše uvedených zdrojů by měla proběhnout dle navržené metodiky.

2.2.2 Příprava legislativních podmínek pro publikaci otevřených dat

Tato etapa je zaměřena na vytvoření legislativy v souladu s doporučeními uvedenými v kapitole 1.13. Měli být katalogizace otevřených dat povinná pro všechny orgány veřejné správy, bude třeba, aby tuto povinnost uložil orgánům veřejné správy zákon. V kapitole 1.13 je dále navrženo, aby přístup k určování podmínek využívání otevřených dat veřejné správy byl sjednocen pomocí jednotného přístupu k této otázce. Dále bylo doporučeno, aby byly jednotně určeny i podmínky využívání samotných katalogizačních záznamů v souladu s principy otevřených dat. Pro zajištění, aby podmínky užívání otevřených dat byly opravdu určovány jednotně v rámci veřejné správy, je třeba, aby harmonizace byla realizována formou zákona.

Již na začátku roku 2012 začala příprava novely zákona č. 106/1999 Sb., o svobodném přístupu k informacím, který s problematikou zveřejňování otevřených dat souvisí. Zároveň je připravována novela Směrnice 2003/98/ES o opakovaném použití informací veřejného sektoru [18]. Protože novela této Směrnice bude mít dopad na určování podmínek využívání dat veřejné správy, jejich opětovného použití a také na možnosti vybírání poplatků za zpřístupnění těchto dat, bude jednotný přístup k určování podmínek užití otevřených dat veřejné správy (jednotná licenční politika) formulována až po schválení novely Směrnice, případně až po přijetí z toho vyplývající novely zákona č. 106/1999 Sb.³⁴, do kterého bude novela Směrnice transponována.

Problematika svobodného přístupu k informacím, kterou se zabývá zákon č. 106/1999 Sb., souvisí s problematikou otevřených dat. Do budoucna by tak základní právní úprava zveřejňování otevřených dat veřejné správy mohla být upravena právě v tomto zákoně. Anglické znění pracovního návrhu novely Směrnice 2003/98/ES také v textu zdůvodnění uvádí, že novela reaguje na aktuální trend využívání otevřených dat [17]. Jak již bylo uvedeno, pokud bude tato novela přijata, bude pravděpodobně implementována právě novelou zákona č. 106/1999 Sb.

Kromě zákona č. 106/1999 Sb. by se problematika otevřených dat mohla dotknout i následujících zákonů:

- Zák. č. 365/2000 Sb., o informačních systémech veřejné správy a o změně některých dalších zákonů - v rámci tohoto zákona by mohla být definována povinnost publikovat data z IS o ISVS a ISDP ve formě otevřených dat a případně také povinnost zajistit v IS o ISVS vazbu mezi příslušným ISVS a katalogizačními záznamy o datech z tohoto systému v Datovém katalogu a mezi datovým prvkem v ISDP a katalogizačními záznamy o datech, kde je daný datový prvek použit. V souvislosti s tím mohou také být dotčeny prováděcí vyhlášky č. 469/2006 Sb., o informačním systému o datových prvcích a č. 528/2006 Sb., o informačním systému o informačních systémech veřejné správy, protože v rámci těchto vyhlášek lze rozšířit počet atributů vedených o ISVS.
- Zák. č. 123/1998 Sb., o právu na informace o životním prostředí - poskytování otevřených dat z oblasti životního prostředí by mělo být harmonizováno se základní právní úpravou poskytování otevřených dat v případě, že tato právní úprava nebude natolik obecná, aby se vztahovala i na otevřená data z oblasti životního prostředí.

2.2.3 Výběr nástroje, jeho dodavatele a provozovatele

Náplní této etapy je výběr katalogizačního nástroje pro plnohodnotný Datový katalog spolu s jeho dodavatelem a určení provozovatele tohoto katalogizačního nástroje. V rámci etapy bude vybírán katalogizační nástroj odpovídající požadavkům určeným v kapitole 1.9. Výběr katalogizačního nástroje a jeho dodavatele bude třeba provést v souladu s platnou legislativou. V závislosti na odhadované ceně zakázky může být nutné ji realizovat postupem daným zákonem č. 137/2006 Sb., o veřejných zakázkách.

Jako součást etapy by měla proběhnout analýza vhodného modelu provozu katalogizačního nástroje, na jejímž základě by měl být určen jeho provozovatel.

2.2.4 Implementace vybraného nástroje

Obsahem etapy je:

- implementace vybraného řešení,
- vytvoření podmínek na MV ČR,

³⁴ Nejedná se o novelu zákona č. 106/1999 Sb., na které se pracuje od začátku roku 2012, novela Směrnice si vyžádá další novelu tohoto zákona.

- příp. úprava Metodiky katalogizace otevřených dat dle plnohodnotného nastavení Datového katalogu,
- školení redaktorů.

Etapa se zaměřuje na implementaci vybraného katalogizačního nástroje, tj. dovozy, nastavení a provedení případných úprav nutných pro splnění všech definovaných požadavků. Současně se etapa zaměřuje na realizaci podmínek nezbytných pro zajištění provozu na MV ČR. V rámci implementace bude také třeba provést migraci katalogizačních záznamů z katalogizačního nástroje používaného v pro zjednodušený Datový katalog.

Aby bylo zajištěno, že nositelé jednotlivých rolí budou schopni s Datovým katalogem pracovat, bude třeba vytvořit dokumentaci Datového katalogu a také návody a školicí materiály pro kurátory, redaktory a koncové uživatele. Dokumentace a školicí materiály by měly být dodány jako součást dodávky katalogizačního nástroje.

Jako součást etapy bude také třeba provést výběr a školení redaktorů Datového katalogu. Někteří z těchto redaktorů už mohou být určeni a vyškoleni během realizace zjednodušeného Datového katalogu, nicméně lze předpokládat, že fungování plnohodnotného Datového katalogu bude klást větší požadavky na provádění redakce z důvodu většího množství katalogizovaných otevřených dat. Zajištění požadované úrovně redakce tak může vyžadovat určení a vyškolení dalších redaktorů. Nástroj zvolený pro realizaci plnohodnotného Datového katalogu se může lišit od toho, který bude použit pro jeho zjednodušenou variantu, což je další důvod, proč bude třeba redaktory Datového katalogu vyškolit.

2.2.5 Ověřovací provoz plnohodnotného nástroje

Stejně jako v případě zjednodušeného Datového katalogu je i v případě spuštění plnohodnotného Datového katalogu vhodné realizovat a následně vyhodnotit fázi ověřovacího provozu. Ověřovací provoz by měl sloužit k ověření fungování katalogizačního nástroje a také k ověření, že Metodika katalogizace otevřených dat VS ČR je vhodně nastavena. Poznatky získané v rámci ověřovacího provozu by měly být vyhodnoceny a měla by být formulována a realizována případná doporučení pro úpravy katalogizačního nástroje, způsobu práce s ním nebo pro úpravy Metodiky katalogizace.

2.2.6 Zahájení provozu

Zahájení provozu je milníkem, který završuje snahy o realizaci Koncepce katalogizace otevřených dat VS ČR a s ní spojeného Datového katalogu. Po dosažení tohoto milníku by měla být realizována všechna doporučení Koncepce a plnohodnotný Datový katalog by měl být funkční v plném rozsahu. Katalogizace otevřených dat by od tohoto okamžiku měla být běžnou součástí postupů zveřejňování otevřených dat ve veřejné správě.

2.3 Seznam výstupů

Tabulka 7: Seznam výstupů

ID	Výstup	ID etapy
V1	Nastavený nástroj pro zjednodušený Datový katalog	1
V2	Metodika katalogizace otevřených dat VS (včetně příruček pro kurátory, redaktory a koncové uživatele)	1

ID	Výstup	ID etapy
V3	Katalogizační záznamy pro data z následujících datových zdrojů (pokud dojde k jejich otevření, viz [2]): <ul style="list-style-type: none"> • Obchodní rejstřík • Insolvenční rejstřík • Informační systém o veřejných zakázkách • Výsledky voleb • Registr aktivních legislativních prací – RALP (resortní nástroje podporující transparentnost výkonu státní správy a zapojení veřejnosti) • Finanční statistika - státní dluh • Finanční statistika - vládní finanční statistika • ÚFIS - účetní záznamy a finanční údaje z CSÚIS • Online přístup k údajům o financování politických stran • Centrální registr dotací 	1
V4	Hodnotící zpráva realizace zjednodušeného Datového katalogu otevřených dat VS ČR	1
V5	Nový či upravený právní předpis, který obsahuje vymezení pojmu otevřená data	2
V6	Nový či upravený právní předpis, který obsahuje povinnost katalogizovat vybraná data veřejné správy ve formě otevřených dat	2
V7	Nový či upravený právní předpis, který určuje jednotný přístup ke stanovování podmínek využití otevřených dat veřejné správy (jednotná licenční politika)	2
V8	Vybraný katalogizační nástroj a jeho dodavatel a provozovatel	3
V9	Smlouvy na dodávku a provoz katalogizačního nástroje	3
V10	Nastavený a implementovaný nástroj pro Datový katalog	4
V11	Dokumentace k Datovému katalogu	4
V12	Upravená metodika a návody a školicí materiály pro kurátory, redaktory a koncové uživatele	4
V13	Migrované katalogizační záznamy ze zjednodušeného Datového katalogu	4
V14	Protokol o proškolení redaktorů	4
V15	Zpráva z vyhodnocení ověřovacího provozu Datového katalogu	4
V16	Protokol o zahájení produktivního provozu	5

2.4 Pracnost katalogizace

V této kapitole je prezentován expertní odhad počtu katalogizačních záznamů v Datovém katalogu a potřebného počtu redaktorů pro zvládnutí tohoto počtu záznamů. Počet záznamů v Datovém katalogu je samozřejmě závislý na počtu dat, respektive datových zdrojů, které publikuje veřejná správa. Protože ale přesný počet těchto dat není v současné době znám, není dále prezentován pouze jeden odhad počtu záznamů v Datovém katalogu, ale jsou prezentovány odhady pro tři varianty z hlediska počtu katalogizačních záznamů: minimální, střední a maximální. Dále uvedené odhadované počty představují počty katalogizačních záznamů, kterých by Datový katalog mohl dosáhnout v období tří let od spuštění provozu plnohodnotného Datového katalogu.

Protože není znám přesný počet dat, který veřejná správa publikuje, bylo třeba alespoň řádově rozdělit existující orgány veřejné správy do několika skupin z hlediska toho, jak velký počet záznamů v Datovém

katalogu se u jednotlivých orgánů veřejné správy očekává. Byly proto vymezeny následující třídy orgánů veřejné správy z hlediska očekávaného počtu katalogizačních záznamů:

- třída A - subjekty publikující největší počet dat a mající největší počet katalogizačních záznamů (řádově stovky záznamů),
- třída B - subjekty publikující středně velký počet dat a mající středně velký počet katalogizačních záznamů (řádově desítky záznamů),
- třída C - subjekty publikující relativně malý počet dat a mající nejmenší počet katalogizačních záznamů (řádově jednotky záznamů),
- třída 0 - zvláštní třída, do které byly zařazeny ty subjekty, které nemají vlastní webové stránky, na kterých by publikovaly svá data (viz níže).

Tabulka 8: Počty orgánů veřejné moci podle tříd očekávaného počtu katalogizačních záznamů

Třída	Počet orgánů veřejné moci
A	44
B	247
C	6296
0	714
Celkem	7301

Do třídy A byla zařazena ministerstva a ostatní ústřední správní orgány, kraje, Celní správa, Česká inspekce životního prostředí, Česká národní banka, Generální finanční ředitelství a Ústav zdravotnických informací a statistiky ČR. Do třídy B byly zařazeny obecní úřady obcí s rozšířenou působností, městské části Hlavního města Prahy Praha 1 – 22 a další orgány jako například Česká obchodní inspekce, Česká správa sociálního zabezpečení nebo Grantová agentura České republiky. Třídou C tvoří z větší části obce, které doplňují orgány veřejné správy nezařazené do třídy A nebo B, dále např. mateřské, základní a střední školy nebo profesní komory.

Jak již bylo uvedeno v samotném úvodu kapitoly, odhady počtu katalogizačních záznamů byly vypracovány pro tři varianty. Varianta označená MIN představuje variantu s nejmenším odhadovaným využitím Datového katalogu. V této variantě je odhadováno, že svá data katalogizuje třetina orgánů VS. Varianta MID je variantou středního využití katalogu, v této variantě svá data katalogizují všechny orgány VS, ale počet záznamů připadající na jeden orgán VS je spíše nižší. Poslední varianta MAX je variantou s maximálním využitím Datového katalogu, v této variantě všechny orgány veřejné správy katalogizují všechna svoje otevřená data. V následující tabulce 9 jsou uvedeny průměrné odhadované počty katalogizačních záznamů připadajících na jeden orgán veřejné správy příslušné třídy v každé z výše uvedených variantách.

Tabulka 9: Průměrné odhadované počty katalogizačních záznamů na jeden orgán veřejné správy

Třída	Varianta MIN	Varianta MID	Varianta MAX
A	33	150	300
B	3	15	55
C	0,3	1,5	5
0	0	0	0

V následující tabulce 10 jsou uvedeny celkové odhadované počty katalogizačních záznamů, které byly vypočítány jako součin průměrného počtu záznamů jednoho orgánu VS určité třídy a počtu orgánů VS v této třídě.

Tabulka 10: Celkové odhadované počty katalogizačních záznamů

Třída	Varianta MIN	Varianta MID	Varianta MAX
A	1 452	6 600	13 200
B	741	3 705	12 350
C	1 889	9 444	31 480
Celkem	4 082	19 749	57 030

Je uvažováno, že celkového počtu katalogizačních záznamů v Datovém katalogu bude dosaženo do tří let od spuštění běžného provozu plnohodnotného Datového katalogu. Následující tabulka uvádí průměrný počet katalogizačních záznamů v Datovém katalogu připadající na jeden rok, měsíc a pracovní den.

Je zřejmé, že rozdíly v celkovém počtu katalogizačních záznamů jsou mezi jednotlivými variantami značné. Zatímco v minimální variantě je odhadován celkový počet záznamů na necelé čtyři tisíce, v maximální variantě je to zhruba o padesát tisíc záznamů více. Pro srovnání, datový katalog Velké Británie³⁵ obsahoval v červenci roku 2012 necelých devět tisíc záznamů.

Tabulka 11: Průměrný počet katalogizačních záznamů v Datovém katalogu za jeden rok, měsíc a pracovní den (zaokrouhлено na celé záznamy)

Průměr za období	Varianta MIN	Varianta MID	Varianta MAX
Průměr za rok	1 361	6 583	19 010
Průměr za měsíc	113	549	1 584
Průměr za pracovní den ³⁶	5	25	72

Průměrná doba, kterou bude potřebovat redaktor na redakci jednoho katalogizačního záznamu včetně redakce jeho případných oprav, je odhadována na 30 minut³⁷. Průměrná doba, kterou bude potřebovat kurátor dat na vytvoření jednoho katalogizačního záznamu včetně jeho případných oprav, je odhadována na 90 minut. Jedná se o odhady průměrné pracnosti těchto činností. Ve skutečnosti bude náročnost vytváření záznamu a jeho redakce dána složitostí katalogizačního záznamu, respektive množstvím datových zdrojů v katalogizačním záznamu obsažených. Při vytváření katalogizačního záznamu může významnou část z odhadované doby trvání tvořit čas strávený vyhledáním a přiřazením vhodných konceptů klasifikačních struktur.

2.4.1 Pracnost pro redaktory obsahu Datového katalogu

Tabulka 12: Průměrná denní pracnost redakce obsahu Datového katalogu

	Varianta MIN	Varianta MID	Varianta MAX
Průměrná pracnost redakce za 1 pracovní den (člh)	2,6	12,5	36
FTE ³⁸ , při redakci do druhého dne	0,3	1,6	4,5

Celkový počet záznamů Datového katalogu, respektive jejich průměrný denní přírůstek, také určuje počet redaktorů, který je potřebný pro provedení redakce těchto záznamů v přijatelném čase. Tabulka 12

³⁵ <http://data.gov.uk/data>

³⁶ Počítáno s 22 pracovními dny v měsíci

³⁷ V uváděném čase je zahrnuta i případná aktualizace záznamu v souvislosti se změnou periodicky se měnících datových zdrojů.

³⁸ FTE - Full Time Equivalent

ukazuje, že při minimální variantě by postačil jeden redaktor na částečný úvazek na to, aby zvládl provést redakci pěti katalogizačních záznamů do druhého dne od jejich založení v Datovém katalogu. Provedení redakce došlých záznamů do druhého dne ve střední variantě již vyžaduje vykonání práce v objemu přesahujícím 1,5 úvazku za den. Zvládnutí denního přírůstku v maximální variantě by vyžadovalo čtyři redaktory na plný úvazek a jednoho redaktora na poloviční úvazek.

2.4.2 Pracnost pro kurátory dat

Tabulka 13: Průměrná denní pracnost vytváření katalogizačních záznamů za celou VS ČR

	Varianta MIN	Varianta MID	Varianta MAX
Průměrná pracnost redakce za 1 pracovní den (člh)	7,7	37,4	108
FTE	1	4,7	13,5

Tabulka 14: Průměrná celková pracnost vytváření záznamů za jeden orgán VS ČR zařazený do třídy A - C za období tří let (v člh)

	Varianta MIN	Varianta MID	Varianta MAX
Orgán třídy A	49,5	225	450
Orgán třídy B	4,5	22,5	75
Orgán třídy C	0,5	2,25	7,5

Na základě průměrného počtu katalogizačních záznamů připadajících na jeden orgán veřejné správy zařazený do třídy A až C (viz tabulka 9) byla vypočítána i průměrná pracnost potřebná na vytvoření těchto katalogizačních záznamů. Při odhadované době tvorby katalogizačního záznamu v délce 90 minut (viz výše) bude v minimální variantě orgán třídy A průměrně potřebovat vynaložit úsilí v objemu cca 50 člověkohodin práce na vytvoření 33 katalogizačních záznamů. V maximální variantě počet záznamů řádově narůstá a s tím se zvyšuje i celková pracnost, potřebná na jejich vytvoření. Ta pro orgán třídy A v maximální variantě činí 450 člověkohodin.

Na orgán třídy C připadá v minimální variantě průměrně 0,3 katalogizačního záznamu. V průměru tak bude orgán třídy C muset vynaložit cca půl hodinu práce jednoho člověka na vytvoření tohoto odhadované počtu katalogizačních záznamů. Nicméně je zřejmé, že pokud se konkrétní orgán třídy C rozhodne vytvořit katalogizační záznam, bude ho muset vytvořit celý a v takovém případě je třeba počítat s pracností jedné a půl hodiny práce jednoho člověka.

2.5 Odhad potřebných nákladů

Tato kapitola se věnuje odhadům nákladů na pořízení plnohodnotného Datového katalogu a jeho následný provoz. V kapitole jsou také obsaženy odhady nákladů spojených s vytvářením katalogizačních záznamů a prováděním redakce.

Pro získání informací o nákladnosti pořízení plnohodnotného Datového katalogu a jeho provozu bylo osloveno 10 subjektů s žádostí o poskytnutí informací (Request for information, RFI). Žadostí byli osloveni výrobci katalogizačních nástrojů nebo dodavatelé informačních systémů pro veřejnou správu. V termínu pro doručení odpovědí byly obdrženy tři vyplněné dotazníky.

2.5.1 Pořizovací náklady Datového katalogu

Zaslané dotazníky v rámci odpovědi na žádost o poskytnutí informací zahrnovaly jak řešení vývojem katalogizačního nástroje na míru, tak řešení úpravou již existujícího řešení. Rozpětí nákladů na pořízení Datového katalogu se v zaslaných dotaznících pohybovalo od 543 750 Kč (21 750 EUR, počítáno s kurzem 1 EUR = 25 Kč) za úpravy a implementaci existujícího open source řešení (bezplatné licence) až po 9 000 000 Kč bez DPH za vývoj a implementaci řešení na míru (včetně ceny licencí softwaru třetích stran). V dalším dotazníku byla cena úprav a implementace existujícího řešení vyčíslena na 3 040 000 Kč

bez DPH, nicméně v dotazníku nebyla uvedena odhadovaná cena licencí. Tato částka tak není konečnou částkou pořízení Datového katalogu.

Tým řešitelů Konceptce expertním odhadem stanovil náklady nasazení a implementace Datového katalogu v případě nasazení existujícího řešení na cca 840 tis. Kč bez DPH (bez ceny licencí). Současně byla týmem řešitelů odhadnuta pracnost a z nich odvozeny odhadované náklady na vývoj zcela nového zakázkového řešení Datového katalogu:

- návrh, vývoj, testování: 2,92 miliónu Kč bez DPH,
- nasazení, vč. provozního prostředí a školení redaktorů a přípravy on-line příručky pro kurátory: 1,7 miliónu Kč bez DPH.

2.5.2 Náklady provozu a údržby Datového katalogu

Odhadované náklady provozu a údržby Datového katalogu byly také součástí dotazníků v rámci žádosti o poskytnutí informací. V zaslaných dotaznících byly náklady údržby a provozu uvedeny zpravidla pouze pro základní úroveň podpory. Skutečné náklady by se lišily podle rozsahu podpory a dalších objednaných služeb.

V situaci, kdy by katalogizační nástroj byl nasazen u správce Datového katalogu, byly náklady provozu a údržby odhadnuty v rozsahu 99 000 Kč ročně (3 960 EUR, 1 EUR = 25 Kč) až 1 350 000 Kč ročně bez DPH. K uvedeným hodnotám je ale třeba podotknout, že ne vždy byl v dotazníku uveden rozsah podpory a uvedené hodnoty tak nelze považovat za porovnávací cen srovnatelné úrovně podpory. Tým řešitelů Konceptce odhadl roční provozní náklady na 0,67 miliónu Kč bez DPH.

V dotazníku byla také zkoumána možnost dodání katalogizačního nástroje v modelu Software jako služba (SaaS). Ve všech zaslaných dotaznících bylo uvedeno, že dodávka řešení tímto způsobem je možná. V jednom případě nebyla uvedena cena dodávky v modelu SaaS. Roční náklady se ve zbylých zaslaných dotaznících pohybovaly v rozmezí 300 000 Kč ročně (1000 EUR měsíčně, 1 EUR = 25 Kč) až 2 700 000 Kč ročně bez DPH (po uvedení do provozu by bylo třeba zaplatit ještě dalších 2,7 mil. Kč bez DPH). I zde je třeba podotknout, že přesná úroveň služeb nebyla v dotaznících uvedena a uvedené hodnoty tak nelze brát jako porovnání různých nabídek pro srovnatelnou úroveň služeb.

2.5.3 Náklady vytváření katalogizačních záznamů a provádění redakce

Odhad nákladů na vytváření katalogizačních záznamů a provádění redakce vychází z pracností spojených s těmito činnostmi, jejichž odhady jsou uvedeny v předcházející kapitole. Pro odhad těchto nákladů je dále uvažováno, že pracovník provádějící tvorbu katalogizačních záznamů nebo redakci obsahu Datového katalogu bude pobírat plat ve výši 25 000 Kč měsíčně včetně osobního ohodnocení (platová třída 12 podle nařízení vlády č. 564/2006 Sb.). Po zohlednění odvodů na sociálního, důchodové, nemocenské pojištění, příspěvek na státní politiku zaměstnanosti, povinné zdravotní pojištění a povinné úrazové pojištění činní náklady na tohoto pracovníka pro orgán veřejné správy 34 550 Kč měsíčně. Při uvažované pracovní době v délce osmi hodin denně a 22 pracovních dnech v měsíci vychází hodinové náklady a tohoto pracovníka 196,31 Kč/hod. Následující tabulka 15 uvádí celkové roční náklady spojené s vytvářením a redakcí katalogizačních záznamů.

Tabulka 15: Průměrné celkové roční náklady spojené s vytvářením a redakcí katalogizačních záznamů (Kč, zaokrouhloeno na tisíce)

	Varianta MIN	Varianta MID	Varianta MAX
Roční náklady na roli kurátora dat za celou VS ČR (Kč)	401 000	1 938 000	5 598 000
Roční náklady na roli redaktora (Kč)	134 000	646 000	1 866 000
Roční náklady celkem za obě role (Kč)	534 000	2 585 000	7 464 000

Obdobně jako byla v přecházející kapitole uvedena průměrná odhadovaná pracnost připadající na jeden orgán veřejné správy zařazený do třídy A až C, jsou v tabulce 16 uvedeny i průměrné odhadované náklady připadající na subjekty v jednotlivých třídách.

Tabulka 16: Průměrné náklady vytváření katalogizačních záznamů připadající na jeden subjekt VS ČR zařazený do třídy A - C (Kč)

Průměrné náklady připadající na jeden orgán VS (Kč)	Varianta MIN	Varianta MID	Varianta MAX
Orgán třídy A	9 717	44 169	88 339
Orgán třídy B	883	4 416	14 723
Orgán třídy C	88	441	1 472

I zde platí, že se jedná o průměrné náklady připadající na jeden orgán VS ČR dané třídy, který je odvozen od průměrného počtu katalogizačních záznamů připadající na tento subjekt. Při odhadovaných nákladech 196,31 Kč/hod a průměrné pracnosti vytváření katalogizačního záznamu 1,5 člh je třeba počítat s náklady ve výši cca 295 Kč na vytvoření každého celého katalogizačního záznamu.

2.6 Způsob provozování katalogizačního nástroje

Pro provoz katalogizačního nástroje připadají v úvahu následující varianty:

- provozovatelem Datového katalogu je jeho správce a pro provoz využívá vlastní ICT infrastrukturu,
- provozovatelem Datového katalogu je subjekt odlišný od jeho správce a pro provoz je využita ICT infrastruktura správce,
- provozovatelem Datového katalogu je subjekt odlišný od jeho správce a pro provoz poskytuje vlastní ICT infrastrukturu (model SaaS).

Výše uvedený výčet možností není úplným výčtem všech možných variant, ale představuje tři běžné možnosti provozu informačních systémů. Varianta provozu A představuje situaci, kdy správce Datového katalogu je zároveň jeho provozovatelem. Pro zjednodušení není rozlišováno, zda vlastní ICT infrastruktura správce použitá pro provoz Datového katalogu je skutečně ve vlastnictví správce, nebo je pronajata od třetího subjektu.

Varianta provozu B představuje situaci, kdy správce Datového katalogu pověří jeho provozem jiný subjekt, který ale neposkytuje pro provoz Datového katalogu ICT infrastrukturu. Ta je zajištěna správcem Datového katalogu. Variantou provozu B je i situace, kdy dodavatel katalogizačního nástroje dodá jeho správci i ICT infrastrukturu potřebnou pro jeho provoz, která se stane majetkem správce. Provozovatelem Datového katalogu v této variantě provozu by mohl být i jiný orgán veřejné správy, nicméně s větší pravděpodobností jím bude soukromý subjekt, zejména samotný dodavatel katalogizačního nástroje, protože ten bude s katalogizačním nástrojem nejvíce obeznámen a bude tedy disponovat odborníky se znalostmi a dovednostmi potřebnými pro provoz katalogizačního nástroje.

Poslední varianta provozu C představuje dodávku katalogizačního nástroje v modelu Software-as-a-Service (SaaS). Provoz katalogizačního nástroje zajišťuje pro správce provozovatel jako službu a dodává jak funkčnost katalogizačního nástroje, tak i ICT infrastrukturu pro jeho provoz.

Výhodou varianty A je, že správce Datového katalogu má plně pod kontrolou i samotný katalogizační nástroj. Nevýhodou je, že správce musí zajistit, jak veškerou potřebnou ICT infrastrukturu, tak i pracovníky se znalostmi a dovednostmi potřebnými pro provoz katalogizačního nástroje.

Výhodou varianty B je, že provoz katalogizačního nástroje zajišťují odborníci s potřebnými dovednostmi a znalostmi. Nevýhodou je, že správce musí zajistit potřebnou ICT infrastrukturu.

Výhodou varianty C je, že poskytovatel pro správce zajišťuje a provozuje jak samotný katalogizační nástroj, tak i ICT infrastrukturu pro jeho provoz. Za nevýhodu může být považována ztráta kontroly nad katalogizačním nástrojem. Správce by si v případě této varianty provozu měl ve smlouvě o poskytování služeb vyhradit veškerá práva k obsahu Datového katalogu a měl by také dbát na to, aby bylo vhodně ošetřeno získání obsahu Datového katalogu v případě ukončení smlouvy či úpadku provozovatele.

Na základě best practice získané při diskusi s pracovníky Ministerstva vnitra bude správcem Datového katalogu pro jeho provoz pravděpodobně zvolena stejná varianta jako, v případě systémů IS o ISVS a

ISDP. Jedná se o výše popsanou variantu provozu A s tím, že dodavatel katalogizačního nástroje bude zajišťovat technickou podporu tohoto nástroje.

2.7 Doporučení pro výběr dodavatele a provozovatele katalogizačního nástroje

Přestože odhadované ceny v odpovědích na žádost o poskytnutí informací (viz kap. 2.5) je třeba brát toliko jako informativní, ve dvou ze tří zaslaných dotazníků tyto ceny převyšují hranici zakázky malého rozsahu podle zákona č. 137/2006 Sb., o veřejných zakázkách. Pro výběr dodavatele, a případně provozovatele, katalogizačního nástroje je tak doporučeno realizovat veřejnou zakázku dle zák. č. 137/2006 Sb.

2.8 Přehled odhadovaných nákladů jednotlivých etap realizace Koncepce

V následující tabulce 17 jsou prezentovány odhady potřebných nákladů jednotlivých etap realizace Koncepce. Náklady a pracnost etapy „*Příprava legislativních podmínek pro otevřená data*“ nejsou uváděny, protože tato etapa bude realizována interně pracovníky příslušných orgánů veřejné správy. Náklady ostatních etap byly určeny expertním odhadem provedeným autory Koncepce, s výjimkou odhadů pro etapu č. 4 a ročních provozních nákladů, kde v odhadovaném rozpětí byly zohledněny i informace z přijatých odpovědí oslovených subjektů na zaslané žádosti o poskytnutí informací.

Tabulka 17: Odhadované náklady jednotlivých etap realizace Koncepce

ID	Etapa	Odhadované náklady (v tis. Kč, bez DPH)
1	Zjednodušený Datový katalog	234
2	Příprava legislativních podmínek pro otevřená data	-
3	Výběr nástroje, jeho dodavatele a provozovatele	215
4	Implementace vybraného nástroje	4 620 - 9 000
5	Ověřovací provoz plnohodnotného nástroje	384
	Celkem	5 453 - 9 833
	Roční provozní náklady údržby	670 – 1 350
	Roční provozní náklady pro role správce a kurátora dat pro varianty MIN a MID	MIN = 535 (134 bez kurátorů dat) MID = 2584 (646 bez kurátorů dat)

3 Použité zkratky

Zkratka	Význam
API	Application Programming Interface
AutZ	autorský zákon, zák. 121/2000 Sb.
CAPTHA	Completely Automated Public Turing Test To Tell Computers and Humans Apart
CSV	Comma Separated Values
ČSÚ	Český statistický úřad
ČÚZK	Český úřad zeměměřičský a katastrální
člh	člověkohodina
DDoS	Distributed Denial of Service
DoS	Denial of Service
FTE	Full Time Equivalent
GML	Geography Markup Language
HTML	Hypertext Markup Language
ISDP	Informační systém o datových prvcích
IS o ISVS	Informační systém o informačních systémech veřejné správy
ISVS	Informační systém veřejné správy
JIP	Jednotný identitní prostor
JSON	JavaScript Object Notation
JSONP	JSON-with-padding
MFF UK	Matematicko-fyzikální fakulta Univerzity Karlovy v Praze
MV	Ministerstvo vnitra
OData	Open Data Protocol
OGDI	(Microsoft) Open Government Data Initiative
orgán VS	orgán veřejné správy
PDF	Portable Document Format
REST	Representational State Transfer
RFI	Request for information
RDF	Resource Description Framework
RTF	Rich Text Format
RÚIAN	Základní registr územní identifikace, adres a nemovitostí
SaaS	Software as a Service
SHP	Shapefile
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VS ČR	Veřejná správa České republiky
XML	eXtensible Markup Language

4 Zdroje

1. About datacatalogs.org, <http://datacatalogs.org/about>
2. Akční plán České republiky „Partnerství pro otevřené vládnutí“ (2012), http://dl.dropbox.com/u/3675785/OSF/5053_12%20-%20III%20-%20material%20-%20akcni%20plan%20OGP.doc
3. Arnošt, D., Hora, P., Hroch, M., Mates, J., Seige, V., Slánský, D.: Business Intelligence příručka manažera. TATE International, s.r.o (2007)
4. Autodesk, Inc.: Autodesk DWG (2012), <http://usa.autodesk.com/adsk/servlet/pc/index?id=6703438&siteID=123112>
5. Bauer, F., Kaltenböck, M.: Linked Open Data: The Essentials. Edition mono/monochrom, Vienna (2011)
6. Berners-Lee, T.: Linked Data - Design Issues (2006), <http://www.w3.org/DesignIssues/LinkedData.html>
7. Berners-Lee, T., Fielding, R., Masinter, L.: Uniform Resource Identifier (URI): Generic Syntax (2005), <https://tools.ietf.org/html/rfc3986>
8. Bolychevsky, I., Peltan, D.: LOD2 Deliverable 9.1.1 First release of the Publicdata.eu Website and Tools (2011), https://docs.google.com/document/d/1Yw7H9gblPFOD02pgYIC_M492N8WI2c-r4BxiZHz1m9Q/export?format=pdf&id=1Yw7H9gblPFOD02pgYIC_M492N8WI2c-r4BxiZHz1m9Q
9. Both, W., Schieferdecker, I.: Berliner Open Data-Strategie [Berlínská strategie pro otevřená data]. Fraunhofer-Informationszentrum Raum und Bau IRB, Stuttgart (2012)
10. Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., Yergeau, F.: Extensible Markup Language (XML) 1.0 (Fifth Edition) (2008), <http://www.w3.org/TR/2008/REC-xml-20081126/>
11. Clark, K. G., Feigenbaum, L., Torres, E.: SPARQL Protocol for RDF (2008), <http://www.w3.org/TR/rdf-sparql-protocol/>
12. Creative Commons, <http://creativecommons.org/>
13. Creative Commons: About The Licenses, <http://creativecommons.org/licenses/>
14. Cyganiak, R., Maali, F., Peristeras, V.: Self-Service Linked Government Data with dcat and Gridworks. In: Proceedings of the 6th International Conference on Semantic Systems, pp. 37:1–37:3. ACM, New York (2010)
15. Český úřad zeměměřičský a katastrální: Struktura a popis výměnného formátu RÚIAN (VFR) (2012), http://www.cuzk.cz/GenerujSoubor.ashx?NAZEV=10-VFR_STRUKTURA_POPIS_0_6_PDF
16. Environmental Systems Research Institute, Inc.: ESRI Shapefile Technical Description (1998), <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>
17. European Commission: Proposal for a Directive of the European Parliament and the Council amending Directive 2003/98/EC on the re-use of public sector information (2011), http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/proposal_directive.pdf
18. Evropská komise: Návrh Směrnice Evropského parlamentu a Rady, kterou se mění směrnice 2003/98/ES o opakovaném použití informací veřejného sektoru (2011), http://ec.europa.eu/information_society/policy/psi/docs/pdfs/opendata2012/revision_of_PSI_Directive/proposal_directive_CS.pdf
19. Evropská komise: Sdělení Komise Evropskému parlamentu, Radě, Evropskému hospodářskému a sociálnímu výboru a Výboru regionů: Veřejně přístupné údaje, hnací síla inovací, růstu a transparentní správy (2011), http://ec.europa.eu/information_society/policy/psi/docs/pdfs/opendata2012/open_data_communication/cs.pdf
20. Freed, N., Borenstein, N.: Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types (1996), <http://www.ietf.org/rfc/rfc2046.txt?number=2046>
21. Glickman, Y., Flügge, M., Pterneas, V.: ENGAGE Deliverable D7.7.6: Data Provision Methods and Components (2012), http://www.engage-project.eu/engage/wp/wp-content/uploads/downloads/2012/02/ENGAGE_D7.7.6-Data-Provision-Methods-Components_v1.0.pdf
22. Google: Telling Humans and Computers Apart Automatically (2012), <http://www.google.com/recaptcha/captcha>
23. Horzinková, E., Novotný, V.: Základy organizace veřejné správy v ČR. 2. upravené vydání. Plzeň: Vydavatelství a nakladatelství Aleš Čeněk, Plzeň (2010)
24. Institut pro veřejnou správu Praha: Slovník nejčastěji používaných pojmů ve veřejné správě (2012), <http://svs.institutpraha.cz/index.php?page=slovník&id=508>

25. Introducing JSON, <http://www.json.org/>
26. Kučera, J., Chlapek, D., Mynarz, J.: Czech CKAN Repository as Case Study in Public Sector Data Cataloging. In: Systémová integrace, vol. 19, is. 2, pp. 95–107 (2012).
27. McBride, B.: Resource Description Framework (RDF): Concepts and Abstract Syntax (2004), <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>
28. Microsoft Corporation: Data Market (2012), <https://datamarket.azure.com/>
29. Microsoft Corporation: Open Data Protocol (OData) Specification (2012), <http://www.odata.org/media/30001/%5Bms-odata%5D.pdf>
30. Microsoft Corporation: Open Government Data Initiative (2010), <http://ogdisdk.cloudapp.net/>
31. Microsoft Corporation: Rich Text Format (RTF) Specification Version 1.9.1 (2008), <http://download.microsoft.com/download/2/f/5/2f599e18-07ee-4ec5-a1e7-f4e6a9423592/Word2007RTFSpec9.docx>
32. Microsoft Corporation: Windows Azure 90-day free trial (2012), <http://www.windowsazure.com/en-us/pricing/free-trial/>
33. Microsoft Corporation: Windows Azure Pricing Calculator (2012), <http://www.windowsazure.com/en-us/pricing/calculator/>
34. Mouzakitīs, S. et al: ENGAGE Deliverable D7.7.1: Analysis Report of Public Sector Data and Knowledge Sources (2011), <http://www.engage-project.eu/engage/wp/wp-content/plugins/download-monitor/download.php?id=4>
35. OGD DataLab: Open Government Data Initiative v5, <https://github.com/openlab/DataLab>
36. Open Definition, <http://opendefinition.org/okd/>
37. Open Geospatial Consortium: Geography Markup Language (2012), <http://www.opengeospatial.org/standards/gml/>
38. Open Geospatial Consortium: KML (2012), <http://www.opengeospatial.org/standards/kml/>
39. Open Knowledge Foundation: CKAN Information Brochure, <http://ckan.org/files/2012/02/CKAN-Information-brochure.pdf>
40. Open Knowledge Foundation: The Open Data Manual (2011), <http://opendatahandbook.org/>
41. Open Knowledge Foundation: Linked Data and RDF (2012), <http://docs.ckan.org/en/latest/linked-data-and-rdf.html>
42. Open Knowledge Foundation: Pricing, <http://ckan.org/solutions/pricing/>
43. Özses, S., Ergül, S.: Cross-domain communications with JSONP, Part 1: Combine JSONP and jQuery to quickly build powerful mashups (2009), <http://www.ibm.com/developerworks/library/wa-aj-jsonp1/>
44. Peters, D.: Version 4.0 – License Draft Ready for Public Comment! (2012), <http://creativecommons.org/weblog/entry/32157>
45. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF (2008), <http://www.w3.org/TR/rdf-sparql-query/>
46. Raggett, D., Le Hors, A., Jacobs, I.: HTML 4.01 Specification (1999), <http://www.w3.org/TR/1999/REC-html401-19991224/>
47. Ricolfi, M. et al: LAPSI Position paper No 4: The „Licensing“ of public sector information, http://www.lapsi-project.eu/lapsifiles/lapsi_licensing_paper_adv_draft.pdf
48. Rodriguez, A.: RESTful Web services: The basics (2008), <https://www.ibm.com/developerworks/webservices/library/ws-restful/>
49. Shafranovich, Y.: Common Format and MIME Type for Comma-Separated Values (CSV) Files (2005), <http://tools.ietf.org/html/rfc4180>
50. Správa základních registrů: Jednotný identitní prostor (JIP) (2012), <http://www.szrcr.cz/slovník-pojmu/jip-jednotny-identitni-prostor>
51. Socrata, Inc.: Getting Started with the Socrata Open Data API, <http://dev.socrata.com/getting-started/>
52. Socrata Inc.: Socratize your datasets. For FREE (2012), <http://www.socrata.com/solutions/socrata-plans-and-pricing/try-it-free>
53. Socrata, Inc.: The Socrata Open Data Platform (2012), <http://www.socrata.com/solutions/solution-tour/>
54. Sunlight Foundation: Ten Principles for opening up government information (2010), <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>

55. The Open Web Application Security Project (OWASP): A Guide to Building Secure Web Applications and Web Services (2005), <http://netcologne.dl.sourceforge.net/project/owasp/Guide/2.0.1/OWASPGuide2.0.1.pdf>
56. Vickery, G.: Review of recent studies on PSI re-use and related market developments (2011)
57. Vláda České republiky: Usnesením Vlády České republiky ze dne 4. dubna 2012 č. 243 o Akčním plánu České republiky Partnerství pro otevřené vládnutí (2012), [http://racek.vlada.cz/usneseni/usneseni_webtest.nsf/0/304FBE579E2DCA01C12579D5002E19DD/\\$FILE/243%20uv120404.0243.pdf](http://racek.vlada.cz/usneseni/usneseni_webtest.nsf/0/304FBE579E2DCA01C12579D5002E19DD/$FILE/243%20uv120404.0243.pdf)
58. Voříšek, J. a kol: Principy a modely řízení podnikové informatiky. 1. vydání. Praha: Nakladatelství Oeconomica (2008)
59. W3C OWL Working Group: OWL 2 Web Ontology Language Document Overview (2009), <http://www.w3.org/TR/owl2-overview/>
60. Wikipedia: Application Programming Interface (2012), http://en.wikipedia.org/wiki/Application_programming_interface
61. Wikipedia: Denial-of-service attack (2012), http://en.wikipedia.org/wiki/Denial-of-service_attack
62. Wikipedia: Portable Document Format (2012), http://en.wikipedia.org/wiki/Portable_Document_Format

4.1 Odkazované předpisy

4.1.1 České právní předpisy

Nařízení vlády č. 564/2006 Sb., o platových poměrech zaměstnanců ve veřejných službách a správě

Vyhl. č. 523/2005 Sb., o bezpečnosti informačních a komunikačních systémů a dalších elektronických zařízení nakládajících s utajovanými informacemi a o certifikaci stínících komor

Vyhl. č. 469/2006 Sb., o formě a technických náležitostech předávání údajů do informačního systému o datových prvcích a o postupech Ministerstva informatiky a jiných orgánů veřejné správy při vedení, zápisu a vyhlášení datových prvků v informačním systému o datových prvcích (vyhláška o informačním systému o datových prvcích)

Vyhl. č. 528/2006 Sb., o formě a technických náležitostech předávání údajů do informačního systému, který obsahuje základní informace o dostupnosti a obsahu zpřístupněných informačních systémů veřejné správy (vyhláška o informačním systému o informačních systémech veřejné správy)

Zák. č. 2/1969 Sb., o zřízení ministerstev a jiných ústředních orgánů státní správy České republiky

Zák. č. 123/1998 Sb., o právu na informace o životním prostředí

Zák. č. 106/1999 Sb., o svobodném přístupu k informacím

Zák. č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů

Zák. č. 365/2000 Sb., o informačních systémech veřejné správy a o změně některých dalších zákonů

Zák. č. 500/2004 Sb., správní řád

Zák. č. 412/2005 Sb., o ochraně utajovaných informací a o bezpečnostní způsobilosti

Zák. č. 137/2006 Sb., o veřejných zakázkách

Zák. č. 300/2008 Sb., o elektronických úkonech a autorizované konverzi dokumentů

Zák. č. 111/2009 Sb., o základních registrech

4.1.2 Evropské předpisy

Směrnice Evropského parlamentu a Rady 96/9/ES o právní ochraně databází

Směrnice Evropského parlamentu a Rady 2003/98/ES o opakovaném použití informací veřejného sektoru

Směrnice Evropského parlamentu a Rady 2007/2/ES o zřízení Infrastruktury pro prostorové informace v Evropském společenství (INSPIRE)