

Koncepce katalogizace otevřených dat VS ČR

(plná 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	Úvod	7
2	Analýza současného stavu a varianty řešení	10
2.1	Souhrnný popis současného stavu publikace dat veřejné správy	10
2.1.1	Typy a formáty aktuálně poskytovaných dat	10
2.1.2	Legislativa související s poskytováním dat	15
2.1.3	Zhodnocení současných praktik zveřejňování dat VS ČR	17
2.2	Současné praktiky katalogizace dat (ve světě)	18
2.2.1	Příklady datových katalogů ve světě	18
2.2.2	Vyhodnocení používaných praktik	20
2.3	Přehled nástrojů pro katalogizaci dat	22
2.3.1	Obecné možnosti volby nástroje	23
2.3.2	Vlastnosti katalogizačního nástroje	24
2.3.3	Používané specializované katalogizační nástroje	25
2.3.4	Další katalogizační nástroje	28
2.4	2.4 Přístupy k plnění, provozu a správě datového katalogu VS ČR	28
2.4.1	Přístupy k plnění datového katalogu VS ČR	28
2.4.2	Přístupy k provozu a správě datového katalogu VS ČR	30
2.4.3	Varianty budování Datového katalogu	31
3	Koncepce katalogizace otevřených dat VS ČR	39
3.1	Úvod od Koncepce	39
3.2	Přínosy katalogu otevřených dat VS ČR	40
3.2.1	Uspřádání přístupu k datům veřejné správy	41
3.2.2	Vytvoření předpokladu pro snazší opětovné použití dat veřejné správy ČR	41
3.2.3	Vytvoření předpokladu pro využívání otevřených propojitelných	41
3.2.4	Vytvoření předpokladu pro dosažení vyšší transparentnosti veřejné správy	42
3.3	Katalogizovaná data	42
3.3.1	Vymezení otevřených dat a rozsahu katalogizace	42
3.3.2	Příklady aplikace definice otevřených dat na existující data	44
3.4	Role a místa zodpovědná za katalogizaci otevřených dat VS ČR	45
3.4.1	Správce Datového katalogu	45
3.4.2	Provozovatel Datového katalogu	46
3.4.3	Poskytovatel dat	46

3.4.4	Kurátor dat.....	46
3.4.5	Redaktor	47
3.4.6	Koncový uživatel.....	47
3.5	Postup katalogizace a související doporučení.....	47
3.5.1	Postup katalogizace.....	47
3.5.2	Pravidla katalogizace	48
3.5.3	Struktura katalogizačního záznamu	48
3.5.4	Doporučení ohledně klasifikačních struktur (slovníků a taxonomií)	51
3.6	Správa a provoz katalogu dat VS ČR.....	51
3.7	Zajištění kvality obsahu datového katalogu	52
3.7.1	Kvalitativní atributy obsahu katalogu a jednotlivých záznamů	52
3.7.2	Využití koncových uživatelů (veřejnosti).....	53
3.7.3	Využití metod strojového učení	54
3.8	Zajištění bezpečnosti provozu datového katalogu.....	54
3.8.1	Možná ohrožení Datového katalogu a cíle z hlediska bezpečnosti.....	54
3.8.2	Bezpečnostní požadavky a opatření pro jejich naplnění.....	55
3.8.3	Další doporučení pro zajištění bezpečnosti Datového katalogu a jeho provozu ...	57
3.8.4	Přístup uživatelů do Datového katalogu a jejich oprávnění	58
3.9	Cílová softwarová architektura nástroje pro Datový katalog	59
3.10	Požadavky na nástroj pro datový katalog a způsob pořízení nástroje	60
3.11	Uživatelé datového katalogu a podmínky jeho použití.....	61
3.11.1	Zapojení veřejnosti do plnění datového katalogu	62
3.12	Vzdělávání v oblasti katalogizace otevřených dat VS ČR.....	62
3.13	Legislativní změny související s katalogizací otevřených dat	62
3.13.1	Podmínky užívání otevřených dat VS ČR	62
3.13.2	Pojem otevřených dat	64
3.13.3	Legislativní ukotvení katalogizace otevřených dat VS ČR	64
3.14	Vazba na připravovaný portál Publicdata.eu	65
3.14.1	Katalogizační nástroj.....	65
3.14.2	Taxonomie pro klasifikaci katalogizačních záznamů	65
3.15	Scénáře užití datové katalogu VS ČR	65
3.15.1	Pověření redaktora a přidělení příslušné role v Datovém katalogu.....	66
3.15.2	Zrušení pověření redaktora.....	66
3.15.3	Založení záznamu v Datovém katalogu	66
3.15.4	Úprava záznamu v Datovém katalogu.....	67

3.15.5	Zneplatnění záznamu v Datovém katalogu	68
3.15.6	Obnovení platnosti záznamu v Datovém katalogu	69
3.15.7	Procházení Datového katalogu a poskytnutí zpětné vazby	70
3.15.8	Vypořádání zpětné vazby a podnětů od koncových uživatelů	70
4	Plán realizace koncepce	72
4.1	Etapy a hlavní skupiny činností	72
4.2	Popis etap realizace koncepce	73
4.2.1	Zjednodušený Datový katalog	73
4.2.2	Příprava legislativních podmínek pro publikaci otevřených dat	74
4.2.3	Výběr nástroje, jeho dodavatele a provozovatele	75
4.2.4	Implementace vybraného nástroje	75
4.2.5	Ověřovací provoz plnohodnotného nástroje	76
4.2.6	Zahájení provozu	76
4.3	Seznam výstupů	76
4.4	Pracnost katalogizace	77
4.4.1	Pracnost pro redaktory obsahu Datového katalogu	81
4.4.2	Pracnost pro kurátory dat	82
4.5	Odhad potřebných nákladů	82
4.5.1	Pořizovací náklady Datového katalogu	82
4.5.2	Náklady provozu a údržby Datového katalogu	83
4.5.3	Náklady vytváření katalogizačních záznamů a provádění redakce	83
4.6	Způsob provozování katalogizačního nástroje	84
4.7	Doporučení pro výběr dodavatele a provozovatele katalogizačního nástroje	85
4.8	Přehled odhadovaných nákladů jednotlivých etap realizace Koncepce	85
5	Slovník základních pojmů	87
6	Použité zkratky	92
7	Zdroje	94
7.1	Odkazované předpisy	97
7.1.1	České právní předpisy	97
7.1.2	Evropské předpisy	98
Příloha A – Varianty budování datového katalogu		99
Příloha B – Žádost o poskytnutí informací		100
Otevřená výzva na poskytnutí informací - nástroj pro datový katalog		100
Příloha 1: Žádost o poskytnutí informace - Dotazník		101
A) Kontakt na zpracovatele odpovědí na otázky		101

B) SW nástroj pro Katalogizaci dat.....	101
C) Licence, pracnost a ceny.....	101
Příloha 2: Specifikace požadavků na katalogizační nástroj.....	102
Request for Information	106
Appendix 1: Request for information - Questionnaire.....	107
A) Contact information	107
B) Data cataloguing software tool	107
C) Licensing, required effort and price	107
Appendix 2: Data cataloguing tool requirements specification	108

1 Úvod

V rámci tohoto dokumentu je popsána Koncepce katalogizace otevřených dat veřejné správy České republiky. Tato koncepce 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].

Veřejná správa v ČR i ve světě (nebude-li řečeno jinak, je v tomto dokumentu označením veřejná správa myšlena veřejná správa ČR) vytváří velké množství dat, která mohou být potenciálně využita subjekty soukromého i veřejného sektoru pro různé účely. Jako příklady využití dat veřejné správy lze uvést zpracování studií a analýz, propojování dat z různých zdrojů a jejich obohacování nebo vytváření aplikací, které umožňují uživatelům s těmito daty nejrůznějším způsobem pracovat a vytvářet z nich informace. Orgány veřejné správy České republiky již dnes aktivně publikují velké množství dat z různých tematických oblastí, která jsou často dostupná prostřednictvím internetu na webových stránkách jednotlivých orgánů veřejné správy. Další data mohou zájemci získat na základě žádosti v souladu se zákonem č. 106/1999 Sb., o svobodném přístupu k informacím.

Koncepce představená v tomto dokumentu se zaměřuje na tzv. otevřená data veřejné správy (tento dokument se samozřejmě zaměřuje na otevřená data veřejné správy ČR a takto je tento pojem chápán, není-li řečeno jinak, i v dalším textu, nicméně pojem otevřená data veřejné správy, v angličtině Open Government Data, je používán i v zahraničí). Koncept otevřených dat veřejné správy zapadá do širšího kontextu snah o posílení transparentnosti a otevřenosti veřejné správy zastřešovaných pod pojmem „Otevřené vládnutí“ (Open Government) [5]. Přistoupením k Partnerství pro otevřené vládnutí (viz výše) se k těmto snahám přihlásila i Česká republika.

Obecně lze otevřená data dle [40] charakterizovat jako data, jejichž podmínky užití neomezuje jejich uživatelé ve způsobu použití dat a opravňují uživatele k jejich dalšímu šíření, pokud při tomto využití a šíření bude uveden autor dat a pokud i ostatní uživatelé budou mít stejná oprávnění s daty nakládat (tj. nedojde k omezení těchto práv například tím, že by uživatel dále šířící otevřená data omezil jejich užití pouze na nekomerční účely). Jednotlivé licence k otevřeným datům mohou ale ducha jejich výše uvedené charakteristiky implementovat různým způsobem a je tak třeba se s těmito podmínkami před užitím dat seznámit, zejména s ohledem na to, jak licence upravuje další šíření nikoli původních dat, ale dat z těchto dat odvozených (např. jako výsledek analýzy původních dat nebo obohacením o data z jiných zdrojů). Konkrétní licence může, ale nemusí vyžadovat, aby i odvozená data byla také šířena jako otevřená data. Tyto podmínky tak ovlivňují možné modely komerčního využití otevřených dat.

Přímo pro otevřená data veřejné správy formulovala nadace Sunlight Foundation deset principů pro posuzování, zda jsou data veřejné správy otevřená a vhodně zpřístupněná veřejnosti. Podle těchto principů jsou data veřejné správy považována za otevřená, pokud jsou [54]:

1. úplná,
2. primární,
3. zveřejněná bez zbytečného odkladu,
4. snadno dostupná,
5. strojově čitelná,
6. neomezuující nikoho ani žádný ze způsobů přístupu,
7. používající standardy s volně dostupnou specifikací (otevřené standardy),

8. jasně licencovaná za podmínek umožňujících jejich využití s minimem možných omezení,
9. stále dostupná,
10. dostupná uživatelům při vynaložení minima možných nákladů na jejich získání.

Otevřená data veřejné správy ČR tedy představují data publikovaná některým orgánem veřejné správy ČR odpovídající výše uvedeným principům. Bližší komentář, jak jsou tyto principy chápány ve vztahu k otevřeným datům VS ČR, je uveden v kapitole 3.3. Zde jen uvedme, že podle těchto principů [54] by otevřená data veřejné správy měla být zveřejněna tak, aby obsahovala pokud možno všechny údaje o určitém subjektu (samozřejmě s výjimkou osobních údajů a dalších dat, jejichž zveřejnění ve formě otevřených dat brání zvláštní zákony, např. dat významných pro zajištění bezpečnosti státu). Otevřená data by měla pocházet z primárních zdrojů, což umožní provádění jejich nezávislých analýz a ověření závěrů analýz již provedených. V praxi ale nemusí být vždy možné zajistit přístup k datům z primárních informačních systémů (např. díky legislativní úpravě), proto pokud existuje sekundární informační systém garantovaný veřejnou správou poskytující určitá data, jeví se adekvátní poskytnutí těchto dat prostřednictvím tohoto systému. Pro usnadnění pochopení dat a tím i jejich využití, by vedle samotných dat měla být zveřejněna i data popisující tato data (metatada). Hodnota některých typů dat klesá v čase a data by tak měla být zveřejňována bez zbytečného odkladu. Bez ohledu na zvolený způsob přístupu k datům by data měla být snadno dostupná. Pokud pro to neexistuje důvod, neměl by přístup k datům být omezen pouze na určitou skupinu osob nebo aplikací. Toto omezení může např. vzniknout díky povinnosti registrace u určitého orgánu veřejné správy. Strojová čitelnost dat usnadňuje jejich další využití, ať už se jedná o jejich analýzy nebo o vývoj aplikací. Např. často užívaný formát PDF je vhodný pro zobrazování dat ve formě čitelné pro člověka, ale už je mnohem hůře zpracovatelný v případě, kdy je třeba ze souboru ve formátu PDF automatizovaně získávat údaje z jednotlivých sloupců v něm uložené tabulky. Použité formáty dat by také měly být otevřené, tj. takové, že jejich specifikace je volně dostupná. To umožňuje, aby data v tomto formátu byla zpracovatelná v různých aplikacích a uživatelé tak nejsou nuceni pro využití dat využívat aplikace jednoho výrobce. Zřetelné uvedení podmínek užití dat, která respektují otevřenost definovanou v [40], respektive v [36], posiluje právní jistotu uživatelů dat, že data mohou legálně používat. Dostupnost otevřených dat veřejné správy by po čase neměla být omezena, aby nedocházelo ke znehodnocení úsilí těch, kteří se na tato data odkazují. Nicméně je vždy třeba zvážit, jaké jsou technologické a finanční možnosti zajištění dostupnosti dat v dlouhodobém časovém horizontu, a proto by měla vzniknout doporučení pro efektivní zajištění tohoto požadavku v praxi. Poplatky za využívání dat mohou působit jako překážka jejich využívání. Samozřejmě je nutné u každého případu individuálně posoudit, jaké náklady jsou se zpřístupněním dat ve formě otevřených dat spojeny¹. Nicméně otevřená data by měla být dostupná uživatelům při vynaložení minima možných nákladů na jejich získání.

Aplikace principů otevřených dat umožňuje, aby data vytvářená a publikovaná veřejnou správou byla lépe využitelná pro různé účely vzpomínané výše. Ačkoli by se na první pohled mohlo zdát, že otevřená data mají význam zejména pro odbornou a laickou veřejnost, aplikace principů otevřených dat má význam i pro samotnou veřejnou správu, která tak může optimalizovat vzájemnou výměnu a sdílení dat.

¹ Např. Směrnice 2003/98/ES o opakovaném využití informací veřejného sektoru v článku 6 stanovuje zásadu, že celkový příjem z vybírání poplatků za poskytování a povolování opakovaného použití dokumentů nesmí přesahovat "*náklady na jejich shromažďování, vytváření, reprodukci a šíření se započítáním přiměřeného zisku*". V bodu 14 odůvodnění je se pak k problematice přiměřeného zisku uvádí, že ten by měl být stanoven s ohledem na "*požadavky samofinancování dotyčného subjektu veřejného sektoru*".

Aby mohla být otevřená data publikovaná veřejnou správou efektivně využita, je třeba, aby bylo možno tato data snadno vyhledat. Pokud se potenciální zájemci nedozví, že určitá data existují, nemohou je snadno využít. Jedním z nástrojů pro usnadnění přístupu k otevřeným datům veřejné správy je tak **datový katalog**, který představuje evidenci existujících dat publikovaných veřejnou správou jako otevřená data. Datový katalog je jednotným místem pro vyhledávání otevřených dat veřejné správy a pro získávání údajů o těchto datech (tj. získávání metadat). Díky tomu datový katalog usnadňuje nejen vyhledávání otevřených dat veřejné správy, ale díky poskytování metadat, jako jsou například autor, datum vytvoření a publikace, používané datové formáty, schémata a podobně, usnadňuje potenciálním uživatelům i pochopení dat.

Účelem **Koncepce katalogizace otevřených dat veřejné správy ČR** popsané v tomto dokumentu je vymezit strategický přístup ČR k vybudování katalogu otevřených dat veřejné správy ČR (dále jen Datový katalog) a k procesu katalogizace těchto dat. Tento dokument je rozdělen do tří částí: Analýza současného stavu a varianty řešení, Koncepce katalogizace otevřených dat VS ČR a Plán realizace koncepce.

První část nazvaná Analýza současného stavu a varianty řešení obsahuje popis aktuálního stavu publikace dat VS ČR a shrnuje praktiky uplatňované v katalogizaci otevřených dat ve světě. Dále je v této části obsažen přehled nástrojů pro katalogizaci dat a jsou zde představeny možné přístupy k plnění Datového katalogu a k jeho správě a provozu.

Ve druhé části dokumentu je obsažena samotná Koncepce katalogizace otevřených dat VS ČR. Tato koncepce přináší odpovědi na základní otázky spojené s katalogizací otevřených dat. Koncepce by tak mimo jiné měla určit, jaká data katalogizovat, kdo by měl katalogizaci provádět a jak by měl při katalogizaci postupovat. V této části jsou také popsány scénáře užití Datového katalogu.

Poslední část obsahuje Plán realizace koncepce. Tento rámcový plán určuje základní kroky při budování katalogu otevřených dat VS ČR a jejich věcnou návaznost. Součástí je i harmonogram realizace navržených kroků a odhadnutý rámcový rozpočet realizace koncepce.

2 Analýza současného stavu a varianty řešení

2.1 Souhrnný popis současného stavu publikace dat veřejné správy

Ačkoli VS ČR v současné době nemá zaveden oficiální Datový katalog, již dnes publikuje data, která různými způsoby zpřístupňuje veřejnosti. Cílem této kapitoly je souhrnně popsat současný stav publikace dat VS ČR a zhodnotit dopad tohoto stavu na přístup zájemců k datům VS ČR.

2.1.1 Typy a formáty aktuálně poskytovaných dat

Bez existence uceleného katalogu dat veřejné správy je obtížné celkově zhodnotit, jaké typy dat VS ČR publikuje a v jakých formátech. Nicméně část autorského kolektivu se podílela na interním grantovém projektu Vysoké školy ekonomické v Praze², v rámci kterého proběhla experimentální katalogizace dat VS ČR (neoficiální experimentální katalog dat VS ČR je k dispozici na internetové adrese <http://cz.ckan.net>). Další část autorského kolektivu se podílela na projektech realizovaných na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze zaměřených na zpracování dat veřejné správy, případně přímo pracují v rámci VS ČR.

Z hlediska toho, o čem data vypovídají (obsah dat), lze říci, že prakticky všechna ministerstva publikují data, která se týkají činností jejich resortu. Data týkající se vykonávaných agend publikují i územní samosprávné celky. Velké množství dat z různých tematických oblastí publikuje Český statistický úřad. V samostatné sekci uvedené níže uvádíme příklady významných datových zdrojů VS ČR.

VS ČR publikuje svá data v celé řadě formátů. Častou formou jsou dokumenty ve formátech PDF³, MS Word (doc/docx) nebo RTF⁴, zejména v případě, kdy jsou data uvedena jako součást zprávy či ročenky. Lze se ale setkat i s případy, kdy jsou v těchto formátech publikovány i pouhé tabulky s daty. Data, která mají povahu tabulek, jsou také někdy publikována jako soubory tabulkového kalkulátoru (často MS Excel – xls/xlsx) nebo soubory hodnot oddělených čárkami (CSV⁵). Tabulky s daty bývají také uveřejňovány na webových stránkách orgánů veřejné správy v podobě HTML⁶ tabulek (tabulka je součástí webové stránky). Data z některých informačních systémů veřejné správy lze získat prostřednictvím jejich aplikačního rozhraní (API). V takovém případě pak data často bývají získávána ve formátech založených na XML⁷. Např. geodata jsou publikována ve formátu GML⁸.

V úvodu bylo zmíněno, že jedním z významných atributů otevřených dat je jejich strojová čitelnost, která umožňuje jejich další automatizované zpracování. Následující tabulka hodnotí vhodnost jednotlivých formátů dat s přihlédnutím k různým charakteristikám strojové čitelnosti. Především je hodnocena existence standardizovaného prostředku pro zápis dat v jednotné strukturované podobě (např. možnost zápisu řádků tabulky ve struktuře, která je zpracovatelná v rámci různých SW aplikací). Také je hodnocena nezávislost formátu na aplikaci, tj. univerzálnost použití dat zapsaných v daném formátu na konkrétní aplikaci nebo typu aplikace

² Grant IGA VŠE 10/2011

³ PDF - Portable Document Format

⁴ RTF - Rich Text Format

⁵ CSV - Comma Separated Values

⁶ HTML - Hypertext Markup Language

⁷ XML - eXtensible Markup Language

⁸ GML - Geography Markup Language; značkovací jazyk založený na XML pro popis geografických informací vyvíjený organizací Open Geospatial Consortium [37]

(např. tabulkový editor). Dále je hodnocena existence standardizovaného prostředku, který umožňuje popsat strukturu a případně i sémantiku (význam) dat ve strojově čitelné podobě. To je důležité pro zajištění správné práce s publikovanými daty. Dále je přihlédnuto k možnosti vytvářet strojově čitelná propojení mezi souvisejícími daty. Propojení umožňují při strojovém zpracování přímo kombinovat související data z různých zdrojů.

Tabulka 1: Vlastnosti formátů dat

Formát	Nezávislost na aplikaci	Zápis ve strukturované podobě	Popis struktury dat	Popis sémantiky dat	Vytváření propojení	Vhodnost použití 1 (nejlepší) - 5 (nejhorší)
PDF	Ne	Ne	Ne	Ne	Ne	5
DOC(X), RTF	Ne	Ne	Ne	Ne	Ne	5
TXT	Ano	Ne	Ne	Ne	Ne	5
HTML	Ano	Částečně	Ne	Ne	Ne	4
XLS(X)	Ne	Částečně	Ne	Ne	Ne	4
CSV	Ano	Ano	Částečně	Ne	Ne	3
JSON	Ano	Ano	Částečně	Ne	Ne	3
XML	Ano	Ano	Ano	Ne	Ne	2
OData	Ano	Ano	Ano	Částečně	Částečně	2
RDF ⁹	Ano	Ano	Ano	Ano	Ano	1

Tabulka 2: Vhodnost použití jednotlivých formátů dat

Formát	Komentář	Vhodnost použití 1 (nejlepší) - 5 (nejhorší)
PDF	Je vhodný pouze pro publikaci dokumentů určených pro následné zpracování člověkem (tj. čtení). Není vhodný pro následné strojové zpracování, neboť není určen pro jednotný strukturovaný zápis dat. Je závislý na aplikaci (PDF prohlížeč či editor).	5
DOC(X), RTF	Z pohledu zveřejňování dat mají tyto formáty podobné vlastnosti jako formát PDF a nejsou tedy vhodné. Je závislý na aplikaci (prohlížeč či editor tzv. „rich text“ dokumentů, tj. dokumentů s formátovaným textem a multimédií).	5

⁹ RDF - Resource Description Framework

Formát	Komentář	Vhodnost použití 1(nejlepší) - 5 (nejhorší)
TXT	Má podobné vlastnosti jako předchozí. Není ale závislý na aplikaci a umožňuje jen zápis „čistých“ textů bez formátování a multimédií.	5
HTML	Je často používaný VS ČR pro zveřejňování dat na různých portálech v podobě čitelné pro člověka. Částečná strojová čitelnost je umožněna jen v případě, kdy jsou využity strukturované prvky formátu pro jednotný strukturovaný zápis dat (např. využití HTML tabulek). Z nich je možné „vytěžit“ data pomocí specializovaného software. Často ale jednotná struktura chybí a data tak není možné strojově využít. V případě jednotné struktury ji není možné popsat pomocí standardizovaných prostředků. Je nezávislý na aplikaci.	4
XLS(X)	Je také často používaný VS ČR pro zveřejňování dat na různých portálech. Umožňuje další zpracování dat člověkem v tabulkovém prohlížeči či editoru (typu MS Excel). Není ale obecně vhodný pro další strojové zpracování ¹⁰ . I když umožňuje strukturovaný zápis dat, jsou pomocí něj často vytvářeny složité tabulkové struktury, ve kterých jsou navíc využívány různé výpočty a makra. Ty nemají smysl mimo samotný tabulkový editor. Navíc není možné pomocí standardizovaných prostředků popsat strukturu a sémantiku dat a není možné související data propojovat.	4
CSV	Jedná se o standardizovaný formát pro zápis tabulkových dat ve strojově čitelné podobě. Je univerzálně použitelný nezávisle na tabulkovém editoru. Struktura dat je částečně popsána přímo uvnitř CSV souborů ve strojově zpracovatelné podobě, ale nebyl zatím vyvinut standardizovaný jazyk pro popis této struktury. Není možné popisovat sémantiku dat a neexistuje standardizovaný prostředek pro propojování souvisejících dat.	3
JSON	Jedná se o formát používaný především pro výměnu strukturovaných dat mezi serverovými a klientskými částmi webových aplikací. Struktura je popsána uvnitř JSON zdrojových kódů ve strojově zpracovatelné podobě, ale nebyl zatím vyvinut standardizovaný jazyk pro popis této struktury. Není možné popisovat sémantiku dat a neexistuje standardizovaný prostředek pro propojování souvisejících dat.	3

¹⁰ Tabulkové editory pracující s formáty XLS(X) zpravidla umožňují export dat z tohoto formátu do formátu CSV, který je lépe strojově zpracovatelný. U složitých tabulkových struktur využívajících výpočty nebo dokonce makra nemusí být možné exportovat úplný obsah souboru ve formátu XLS(X).

Formát	Komentář	Vhodnost použití 1(nejlepší) - 5 (nejhorší)
XML	Umožňuje zveřejňování strukturovaných dat (ne jenom tabulkových ale i hierarchických dat) ve strojově čitelné podobě. Strukturu je navíc možné popsat pomocí standardizovaného prostředku (např. jazyka XML Schema). Je často používaný datovými API VS ČR. Je nezávislý na aplikaci, ve které je použit. Neumožňuje snadné propojování souvisejících dat a popis sémantiky.	2
OData	Jedná se o protokol (celý název je Open Data Protokol), který umožňuje standardizovaným způsobem webovým klientům číst a upravovat data na webových serverech. Kombinuje jednodušší protokol AtomPub s formátem JSON. Využívá standardního prostředku pro popis struktury vyměřovaných dat a částečně i jejich sémantiky. Umožňuje také propojovat související data, avšak pouze v rámci jednoho datového zdroje.	2
RDF	Jedná se o grafový datový model umožňující zachytit libovolná strukturovaná data ve strojově čitelné podobě. Nabízí standardizovaný prostředek pro popis struktury i sémantiky dat. Je založen na známých principech webu, s jejichž pomocí umožňuje i plné propojování souvisejících dat z různých zdrojů.	1

2.1.1.1 Příklady datových zdrojů VS ČR

VS ČR poskytuje řadu dat na svých portálech. Pro představu uvádíme několik příkladů v následující tabulce.

Tabulka 3: Příklady významných datových zdrojů VS ČR

Portál	Popis dat	Formát
ESF ČR (http://www.esfcr.cz)	Poskytuje data o projektech financovaných strukturálním fondem ESF ČR. Data jsou poskytována ve formátu HTML v jednotné struktuře, která umožňuje vytěžování.	HTML
Informační systém o veřejných zakázkách (http://www.isvzus.cz)	Poskytuje povinně zveřejňované údaje o nadlimitních a podlimitních veřejných zakázkách VS ČR. Data jsou uveřejňovány v jednotné HTML struktuře, kterou lze po poměrně časově náročném programování strojově vytěžovat. Obsahuje často nevalidní data (neexistující IČO, nesmyslné částky, ...).	HTML

Portál	Popis dat	Formát
Veřejná databáze ČSÚ (http://vdb.czso.cz)	Umožňuje exportovat statistická data do formátů XLS a XML. XML export nereflektuje logickou strukturu zveřejňovaných dat. Reflektuje tabulkovou strukturu, což může v určitých případech komplikovat zpracování.	XLS, XML
Administrativní registr ekonomických subjektů (ARES) (http://www.info.mfcr.cz/ares/ares.html.cz)	Jedná se o XML datové API umožňující přístup k datům v několika zdrojových registrech (např. v obchodní, rejstříku). XML struktura reflektuje logickou strukturu zveřejňovaných dat a proto je strojové zpracování poměrně snadné. Data ale nejsou zcela aktuální a úplná.	XML
Veřejné rozhraní k Centrálnímu systému účetních informací státu (ÚFIS) (http://www.info.mfcr.cz/ufis/)	Jedná se o XML datové API umožňující přístup k vybraným údajům účetních a finančních informací státu. XML struktura reflektuje logickou strukturu zveřejňovaných dat.	XML
Informační systém výzkumu, experimentálního vývoje a inovací (VaVal) (http://www.isvav.cz)	Informační systém poskytuje data o výzkumných aktivitách, projektech, záměrech, výsledcích a veřejných soutěžích. Data umožňuje exportovat do formátu XLS	HTML, XLS
Portál veřejné správy (http://portal.gov.cz)	Informační systém veřejné správy v současnosti poskytuje popisy nejčastějších životních situací, informace o orgánech VS a zákonech ČR	HTML, částečně XML (životní situace)
Základní registr územní identifikace, adres a nemovitostí (www.ruian.cz)	Jeden ze základních registrů. Poskytuje referenční údaje o území, adresách a nemovitostech.	VFR ¹¹ (GML/XML)

2.1.1.2 Současné využití katalogizace ve VS ČR

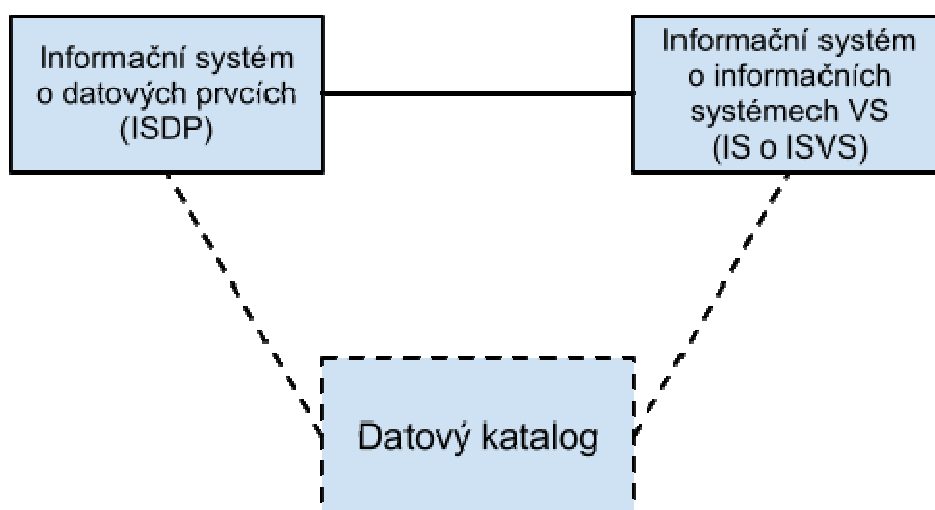
Principy katalogizace již VS ČR aplikuje v oblasti geografických dat v podobě Národního geoportálu INSPIRE¹². Portál katalogizuje jak data, tak i datové služby, které orgány veřejné

¹¹ VFR je výměnný formát RÚIAN, tento formát je podмноžinou standardu GML [15]

¹² Národní geoportál INSPIRE: <http://geoportal.gov.cz>.

správy poskytují veřejnosti. Katalogizovaná data i služby mohou být zpoplatněny. Geoportál také data integruje a poskytuje uživatelům prostřednictvím mapové služby.

Dalšími katalogy na úrovni státní správy jsou Informační systém o datových prvcích (ISDP)¹³ a Informační systém o informačních systémech veřejné správy (IS o ISVS)¹⁴. ISDP slouží jako katalog datových prvků, které jsou informační systémy povinny používat při vzájemné výměně dat. IS o ISVS slouží jako katalog informačních systémů veřejné správy. Oba katalogy udržují sadu metadat a umožňují dle nich vyhledávat. ISDP a IS o ISVS se vzájemně vhodně doplňují. Zatímco IS o ISVS eviduje informační systémy VS ČR, ISDP částečně eviduje, s jakými datovými prvky informační systémy pracují. Je zřejmé, že vhodným doplněním a logickým uzavřením by byl katalog dat, který by evidoval data, které informační systémy zveřejňují a z jakých datových prvků jsou složena. To schematicky znázorňuje obrázek 1.



Obrázek 1: Katalogizace informačních systémů, veřejných dat a datových prvků VS ČR

Vzájemně propojené systémy IS o ISVS, ISDP a Datový katalog tak tvoří podstatné části katalogu veřejné správy. Pokud data katalogizovaná v Datovém katalogu pocházejí z některého informačního systému veřejné správy, údaje o tom, z jakého ISVS pocházejí a jaké datové prvky využívají, jsou významnými metadaty katalogizovaných dat. Je tak třeba zdůraznit vzájemnou provázanost těchto systémů. V Datovém katalogu by nemělo docházet k duplicitnímu zadávání metadat o ISVS a datových prvcích, ale záznamy v Datovém katalogu by měly obsahovat odkazy na relevantní záznamy v IS o ISVS a v ISDP. Obdobně by v IS o ISVS mohly být uvedeny odkazy na katalogizační záznamy dat, která jsou výstupem ISVS.

Ačkoli se otevřená data zaměřují na data samotná a na jejich další využití, informace v Datovém katalogu o tom, ze kterého ISVS data pocházejí, mohou mít pro uživatele význam i v tom, že některé ISVS poskytují i veřejně dostupné služby a rozhraní, pomocí kterých lze získat užitečné prezentace dat (např. mapové vizualizace v RÚIAN).

2.1.2 Legislativa související s poskytováním dat

V úvodu byla otevřená data obecně vymezena jako data, která mohou jejich uživatelé volně užívat a dále šířit, pokud zůstane uveden autor dat a nedojde k omezení dalších uživatelů v používání dále šířených či zpracovaných dat [40]. Otevřená data tak nejsou záležitostí pouze

¹³ <https://www.sluzby-isvs.cz/ISDP>

¹⁴ <https://www.sluzby-isvs.cz/ISoISVS>

technickou, ale i právní, neboť je třeba zajistit, aby jejich uživatelé měli potřebná oprávnění k užití dat v rozsahu, který odpovídá požadované otevřenosti dat. V této sekci jsou tak uvedeny nejdůležitější právní normy, které se týkají otevřených dat a jejich využívání.

Data mohou podléhat různé právní ochraně v závislosti na typu dat a kontextu jejich použití. Z hlediska užití dat se ochranou dat, respektive databází, zabývá zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon, AutZ). Autorský zákon umožňuje ochranu databáze autorským právem, pokud jsou splněny podmínky uvedené v § 2 odst. 2 AutZ. Podle § 2 odst. 6 AutZ není údaj sám o sobě autorským dílem. Obsah databází se statistikami, přehledy či záznamy, které se v oblasti otevřených dat veřejné správy vyskytují velmi často, tak pravděpodobně nebude chráněn autorským právem. I když je samozřejmě v každém případě třeba posoudit, zda jsou u dané databáze naplněny pojmové znaky nutné pro její ochranu prostřednictvím autorského práva.

Databáze je podle AutZ chráněna také tzv. zvláštním právem pořizovatele databáze, které je zakotveno ve Směrnici 96/9/ES o právní ochraně databází a v českém právním řádu je upraveno v §§ 88 – 94 AutZ. Databáze chráněná tímto zvláštním právem je vymezena v § 88 AutZ jako „*soubor nezávislých děl, údajů nebo jiných prvků, systematicky nebo metodicky uspořádaných a individuálně přístupných elektronickými nebo jinými prostředky, bez ohledu na formu jejich vyjádření*“. Podle § 94 odst. 2 AutZ je přiměřenou aplikací § 3 písm. a) z ochrany zvláštním právem pořizovatele databáze vyjmuta databáze, která je součástí právního předpisu.

Kromě toho, že autorský zákon upravuje rozsah ochrany databází, upravuje i licence a licenční smlouvy jako právní nástroj, pomocí kterého autor poskytuje nabyvateli licence oprávnění k výkonu práva užití díla, respektive oprávnění k výkonu práv plynoucích z obsahu zvláštního práva pořizovatele databáze. S ohledem na licence k informacím veřejného sektoru se předběžná podoba doporučení LAPSI (Evropská tematická síť pro otázky právních aspektů poskytování informací veřejného sektoru) [47] přiklání k názoru, že využívání mezinárodně uznávaných licenčních smluv je výhodnější z hlediska mezinárodního využití informací veřejného sektoru. Takovými mezinárodními licencemi jsou např. licence Creative Commons [12]. V současné době je připravována nová verze těchto licencí, která by měla plně zohledňovat i zvláštní právo pořizovatele databáze [44].

V předcházejícím odstavci byl použit pojem informace veřejného sektoru, který souvisí s pojmem otevřená data a vychází ze Směrnice 2003/98/ES o opakovaném použití informací veřejného sektoru. Tato Směrnice hovoří o dokumentu jako o nosiči informací veřejného sektoru, nicméně tímto dokumentem je podle článku 2 „*obsah na jakémkoli nosiči (psaný či tištěný na papíře či uložený v elektronické formě nebo jako zvuková, vizuální nebo audiovizuální nahrávka) nebo část takového obsahu*“. Definice dokumentu je tak dosti široká a lze ji vztáhnout i na data veřejného sektoru, respektive na data veřejné správy, která je součástí veřejného sektoru. Směrnice 2003/98/ES byla do českého právního řádu transponována do zákona č. 106/1999 Sb., o svobodném přístupu k informacím. Cílem Směrnice je umožnit, aby informace veřejného sektoru byly fyzickými a právními osobami využívány i pro jiné účely, než pro který byly původně pořízeny, a to bez ohledu na to, zda jsou tyto účely komerční či nekomerční.

Směrnice 2003/98/ES principům otevřených dat nebrání, nicméně např. v oblasti formátů nezahrnuje požadavek na strojovou čitelnost uvedený v [54]. Již dnes ale existuje pracovní návrh změny Směrnice 2003/98/ES [18], který reaguje jak na požadavek na strojovou čitelnost dat veřejné správy členských zemí EU, tak i na fakt, že původní implementace Směrnice v členských zemích EU jsou rozdílné, zejména v oblasti vybírání poplatků za poskytování informací, v čemž je spatřována překážka jejich využívání. Nový návrh Směrnice tak zavádí povinnost poskytovat informace ve strojově čitelné podobě, pokud je to možné a vhodné, a

informace by měly být využitelné pro komerční i nekomerční účely, pokud není výslovně stanoveno jinak. Poplatky mají být omezeny na výši, která se rovná mezním nákladům na reprodukci a šíření informací s tím, že ve výjimečných případech lze vybírat poplatky vyšší.

V souvislosti s výše uvedeným uveďme ještě, že zatímco koncept otevřených dat je postaven na aktivním zpřístupňování dat veřejné správy k dalšímu využití (a Směrnice 2003/98/ES je s tímto v souladu), koncept svobodného přístupu k informacím je postaven spíše na poskytování informací veřejného sektoru na žádost potenciálních uživatelů těchto dat [9]. Nicméně zákon č. 106/1999 Sb., o svobodném přístupu k informacím, v § 4 umožňuje poskytování informací jak na žádost, tak i zveřejněním, tedy aktivním zpřístupněním dat. Tento zákon určuje i typy informací, které musí být zveřejněny povinně. Mezi takovéto informace patří i informace v evidencích a rejstřících, které jsou podle zvláštního zákona každému přístupné (viz § 5 zák. č. 106/1999 Sb.).

Kromě připravované změny obecného přístupu k poskytování informací veřejného sektoru v rámci EU existuje již dnes v některých oblastech zvláštní právní úprava, která umožňuje ve větší či menší míře volné užití určitého typu dat veřejné správy. V této souvislosti jmenujme alespoň Směrnici 2007/2/ES o zřízení Infrastruktury pro prostorové informace v Evropském společenství (INSPIRE), která je do českého právního řádu transponována v zákoně č. 123/1998 Sb. o právu na informace o životním prostředí. Tato Směrnice ukládá členským státům EU povinnost vytvářet metadata k prostorovým datům (geodatům), provozovat infrastrukturu pro prostorové informace a poskytovat služby založené na prostorových datech prostřednictvím geoportálu INSPIRE¹⁵. Podle článku 11 této Směrnice a podle § 11c odstavce 4 zák. č. 123/1998 Sb. jsou služby umožňující vyhledávání a službu umožňující zobrazování a prohlížení prostorových dat zpřístupněny veřejnosti bezplatně. Nicméně podle tohoto paragrafu „data zpřístupněná prohlížečnými službami založenými na prostorových datech mohou být ve formě zabraňující opětovnému využití pro obchodní účely“.

Na závěr uveďme ještě zákon č. 111/2009, o základních registrech, ve kterém je základní registr územní identifikace, adres a nemovitostí (RÚIAN)¹⁶ podle § 30 odst. 2 veřejným seznamem a spolu s ním je podle § 39 odst. 3 veřejným registrem i informační systém územní identifikace. Podle § 47 jsou zveřejňovány „změny v územně správním členění České republiky, vzájemné vazby a skladebnost územních prvků a územně evidenčních jednotek, a to i v grafické formě a adresy adresních míst na území České republiky“. Podmínky poskytování údajů z Registru územní identifikace jsou pak uvedeny v § 62, který mimo jiné stanoví, že „Registr územní identifikace je veřejně přístupný způsobem umožňujícím dálkový přístup. Údaje z tohoto registru se poskytují v elektronické podobě, včetně vytěžování jeho obsahu; na nosičích dat se poskytují za úplaty, nejedná-li se o údaje pro tvorbu územně analytických podkladů“.

2.1.3 Zhodnocení současných praktik zveřejňování dat VS ČR

Jak ukazují předchozí kapitoly, VS ČR již pro veřejnost publikuje řadu dat. Současný přístup však není koordinován a nejsou dány jednoznačné postupy pro zveřejňování dat. Výjimku tvoří oblast geodat, kde je přístup koordinován mezinárodně pomocí Směrnice 2007/2/ES (INSPIRE). Současný stav vede k několika problémům:

¹⁵ Národní geoportál INSPIRE: <http://geoportal.gov.cz>, Evropský geoportál INSPIRE: <http://inspire-geoportal.ec.europa.eu/>

¹⁶ Základní registr územní identifikace adres a nemovitostí: <http://www.szrcr.cz/registr-uzemni-identifikace-adres-a-nemovitosti>

- Je obtížné zjistit, jaká data VS ČR publikuje a kde. Neexistuje nástroj, který by umožnil zveřejňovaná data katalogizovat a vyhledávat v záznamech vzniklého katalogu na základě různých metadat (datum publikace, autor dat, kategorie, atd.).
- Data jsou zveřejňována v řadě různých formátů a to často takových, které neumožňují opakované strojové zpracování (např. PDF, DOC, HTML, atd.).
- Data nejsou publikována pod vhodnou licencí, což znemožňuje veřejnosti data volně využívat, vyhledávat souvislosti a nalezené souvislosti zaznamenávat.

2.2 Současné praktiky katalogizace dat (ve světě)

Ve světě jsou budovány datové katalogy na různých úrovních veřejné správy a jsou pro ně využívány různé nástroje. V této kapitole jsou prezentovány praktiky a přístupy, které se ve světě používají pro katalogizaci otevřených dat veřejné správy. Vzhledem k faktu, že se oblast katalogizace otevřených dat veřejné správy velmi dynamicky vyvíjí, není cílem této kapitoly poskytnout vyčerpávající analýzu současného stavu katalogizace otevřených dat veřejné správy, ale poskytnout základní přehled významných světových datových katalogů a na jejich základě formulovat inspirativní doporučení pro katalogizaci otevřených dat VS ČR. Vzhledem k zaměření kapitoly na zahraničí, nevztahují se zde použité pojmy veřejná správa a (otevřená) data veřejné správy k České republice.

2.2.1 Příklady datových katalogů ve světě

V řadě zemí světa dnes probíhají aktivity zaměřené na otevřená data a jejich katalogizaci. Podrobně se aktivitami zaměřenými na publikování a katalogizaci dat veřejné správy zabývá např. výzkumná zpráva projektu ENGAGE [34]. V této sekci tak uvádíme pouze příklady významných datových katalogů, které ve světě v rámci těchto iniciativ vznikly.

Aktivity a iniciativy orientované na otevřená data veřejné správy a s nimi spojené datové katalogy je možné z hlediska zaměření rozdělit na lokální, národní a mezinárodní. Lokální datové katalogy se zaměřují na otevřená data měst či regionů, mezi které lze počítat i samostatné územní celky států s federativním zřízením. Národní datové katalogy se zaměřují na otevřená data z celého státu. Nadnárodní datové katalogy se pak zaměřují na otevřená data pocházející z více států. Zdroj [34] poukazuje na fakt, že i jednotlivé instituce jako statistické úřady na svých webových stránkách publikují velké množství dat. Bylo by tak možné hovořit o datových katalozích jednotlivých institucí. I když už se nemusí zcela striktně jednat o data veřejné správy, dle citovaného datového zdroje realizují svoje iniciativy v oblasti otevřených dat i nadnárodní instituce jako je Organizace spojených národů nebo Světová banka. Existují tak i datové katalogy nadnárodních institucí.

Dalším z hledisek, podle kterých lze datové katalogy členit, je hledisko oficiality takového datového katalogu. Ne všechny katalogy otevřených dat veřejné správy jsou oficiální, tedy vytvářené přímo veřejnou správou. Příklady oficiálních národních a lokálních datových katalogů a jimi používaných nástrojů jsou uvedeny v tabulce 4, respektive v tabulce 5. Že se jedná o oficiální katalogy je odvozeno buď z informací uvedených přímo na webových stránkách datového katalogu, nebo z toho, že datový katalog je umístěn na webové adrese, která patří některému z úřadů veřejné správy daného státu.

Ve sloupci „Nástroj“ je uveden název nástroje, který je použit pro datový katalog. Pokud je v tomto sloupci uvedeno heslo „*Proprietární*“, znamená to, že dostupné zdroje naznačují, že se jedná o nástroj vytvořený na míru (i když je třeba připustit, že na pozadí webového portálu zpřístupňujícího záznamy datového katalogu může být v provozu i některý ze specializovaných katalogizačních nástrojů, ale pokud to není na datovém portálu nebo v jiném zdroji uvedeno, nelze nástroj rozpoznat).

Tabulka 4: Příklady oficiálních národních datových katalogů

Stát	Katalog	Nástroj
Belgie	http://data.gov.be	Proprietární
Dánsko	http://data.digitaliser.dk/	Proprietární
Estonsko ¹⁷	http://opendata.riik.ee/	Proprietární
Finsko	http://data.suomi.fi	Proprietární
Francie	http://www.data.gouv.fr/	Proprietární
Nizozemsko	http://data.overheid.nl	CKAN
Portugalsko	http://www.dados.gov.pt/	Proprietární
Rakousko	http://www.data.gv.at/	CKAN
Slovensko	http://data.gov.sk/	CKAN
Španělsko	http://datos.gob.es	Proprietární
USA	http://www.data.gov/	Socrata
Velká Británie	http://catalogue.data.gov.uk/	CKAN

Tabulka 5: Příklady oficiálních lokálních datových katalogů

Stát	Region	Katalog	Nástroj
Finsko	Helsinky	http://www.hri.fi/en/	CKAN
Francie	Paříž	http://opendata.paris.fr/	Proprietární
Kanada	Regina	http://www.regina.ca/residents/open-government/data/	OGDI DataLab
Německo	Berlín	http://daten.berlin.de/	CKAN
Španělsko	Zaragoza	http://datos.zaragoza.es	Proprietární
USA	Stát Oregon	http://data.oregon.gov	Socrata
USA	Chicago	http://data.cityofchicago.org	Socrata
Velká Británie	Kent	http://picandmix.org.uk/categories/	Proprietární

¹⁷ Datový katalog ve výstavbě

Pro datové katalogy se používají jak proprietární katalogizační nástroje vytvořené na míru potřebám dané katalogizační aktivity tak i specializované katalogizační nástroje CKAN (Comprehensive Knowledge Archive Network), Socrata a Microsoft Open Government Data Initiative DataLab. Tyto nástroje budou představeny v následující kapitole. Je také zřejmé, že v některých zemích existují jak oficiální národní katalog i oficiální lokální datové katalogy (např. Španělsko, Velká Británie). Berlínská strategie pro otevřená data v souvislosti s tím uvádí, že datový katalog Berlína by měl být integrován s datovým katalogem Spolkové republiky Německo a také s datovými katalogy dalších německy mluvících zemí [9].

Následující tabulka 6 obsahuje příklady neoficiálních datových katalogů. Tyto katalogy jsou zpravidla dostupné na webové adrese, která nepatří orgánu veřejné správy v dané zemi, případně se u nich nepodařilo ověřit, že se jedná o oficiální datový katalog. Mezi zeměmi s neoficiálním katalogem dat veřejné správy je Česká republika. Tento neoficiální datový katalog byl naplněn katalogizačními záznamy v rámci výzkumného projektu Vysoké školy ekonomické v Praze.

Tabulka 6: Příklady neoficiálních datových katalogů

Země	Region	Typ	Katalog	Nástroj
Česká republika	N/A	Národní	http://cz.ckan.net	CKAN
Irsko	N/A	Národní	http://ie.ckan.net/	CKAN
Itálie	N/A	Národní	http://it.ckan.net/	CKAN

Již byly uvedeny příklady lokálních i národních datových katalogů. Jako příklad nadnárodního datového katalogu lze zmínit portál Publicdata.eu, v rámci kterého vzniká evropský katalog dat veřejné správy, a který je vytvářen v rámci výzkumného projektu LOD2¹⁸ [8]. Tento portál také využívá katalogizační nástroj CKAN a předběžná podoba tohoto portálu je již dnes přístupná na webové adrese <http://www.publicdata.eu/>. Do tohoto katalogu byla zařazena data z neoficiálního datového katalogu České republiky.

Portál Publicdata.eu je zatím pouze výzkumným projektem, nicméně Sdělení Evropské komise „Veřejně přístupné údaje, hnací síla inovací, růstu a transparentní správy“ z prosince roku 2011 uvádí jako klíčová opatření pro oblast veřejně přístupných údajů vznik portálu pro data Komise a dalších orgánů a agentur EU a také zřízení celoevropského datového portálu umožňujícího přístup k datům z celé EU [19]. Podle tohoto Sdělení se předpokládá spuštění tohoto portálu na jaře roku 2013.

Na závěr je vhodné ještě zmínit katalog datacatalogs.org¹⁹. Nejedná se o katalog otevřených dat veřejné správy, ale o katalog takovýchto katalogů [1]. Lze ho tedy použít pro hledání existujících datových katalogů.

2.2.2 Vyhodnocení používaných praktik

U katalogů dat veřejné správy a s nimi spojených katalogizačních aktivit realizovaných ve světě lze vysledovat následujících sedm hlavních praktik související s katalogizací dat veřejné správy:

- využívání samostatných katalogizačních nástrojů,
- katalogizace dat na různých úrovních veřejné správy,

¹⁸ Projekt EU ICT FP7 vedený pod číslem 257943, <http://lod2.eu/>

¹⁹ <http://www.datacatalogs.org/>

- propojování katalogů na různých úrovních,
- potřeba definovaných procesů a legislativně podpořených procesů tvorby a údržby katalogu,
- existence jednoznačné metodiky určující co katalogizovat,
- použití jednotných taxonomií pro klasifikaci dat,
- jasné vymezení práv a povinností při využívání otevřených dat.

2.2.2.1 Využívání samostatných katalogizačních nástrojů

I když tomu tak není ve všech případech, většina analyzovaných katalogů dat veřejné správy využívá pro vedení katalogizačních záznamů samostatné nástroje. V rámci států Evropské unie je pro tyto účely často využíván nástroj CKAN. V některých případech jsou tyto samostatné katalogizační nástroje integrovány do webových portálů, což umožňuje upravit vzhled a způsob prezentace obsahu datového katalogu tak, aby lépe vyhovoval celkovému zaměření katalogizační iniciativy. Portály také umožňují v rámci katalogizační iniciativy využívat i dalších funkcí, které nejsou přímo podporované v katalogizačních nástrojích, jako je například diskusní fórum nebo zveřejňování zpráv a článků o otevřených datech a jejich využívání.

Doporučení pro ČR: Pokud VS ČR nemá informační systém, který by disponoval funkcí nutnou pro podporu katalogizace dat, jeví se vhodné využít samostatný katalogizační nástroj. Katalogizační nástroj by měl být volen tak, aby umožnil případnou integraci s Portálem veřejné správy (případně s jiným portálem či portály subjektů pověřených správou katalogizačního nástroje). Integrace s Portálem veřejné správy by měla přinést sjednocení přístupu k datům s přístupem k jiným informacím veřejné správy České republiky.

2.2.2.2 Katalogizace dat na různých úrovních veřejné správy

Ve světě probíhá katalogizace dat veřejné správy na různých úrovních – na centrální úrovni v podobě celostátních katalogů, ale i na úrovni místních samospráv, měst nebo federálních celků. Statistické úřady často na svých webových stránkách publikují značné množství dat a tyto webové stránky by tak bylo možné označit za datový katalog samostatné instituce.

Doporučení pro ČR: Přístup ke katalogizaci otevřených dat VS ČR je třeba navrhnout s ohledem na uspořádání a vnitřní fungování VS ČR. Měla by ale být zvažována jak varianta centrálního katalogu pro celou veřejnou správu ČR, tak i varianta založená na oddělených datových katalozích vybraných orgánů veřejné správy.

2.2.2.3 Propojování katalogů na různých úrovních

Tam, kde existují oddělené datové katalogy na různých úrovních veřejné správy, se tyto katalogy integrují a propojují, aby bylo umožněno vyhledávání napříč těmito katalogy. Vznikající portál Publicdata.eu a plánovaný oficiální celoevropský datový portál (viz [19]) jsou příkladem, že tato integrace může probíhat i přes hranice. V případě této přeshraniční integrace je třeba zajistit propojení datových katalogů nejen po stránce technické, ale i z hlediska jazyka, aby popisy dat byly srozumitelné i lidem mimo stát, kde data původně vznikla.

Doporučení pro ČR: Katalog otevřených dat VS ČR je třeba budovat s ohledem na možné požadavky integrace datových katalogů, ať už by se jednalo o nutnost integrovat oddělené datové katalogy více orgánů veřejné správy nebo o integraci s celoevropským datovým portálem. Celoevropský datový portál byl zatím oznámen ve Sdělení Evropské komise [19], nicméně vývoj v této oblasti je třeba sledovat, aby požadavky na integraci celoevropského datového portálu a Datového portálu byly včas identifikovány a zapracovány. Dále při budování Datového katalogu by bylo vhodné vyhodnotit, zda bude Datový katalog napojen na již vznikající portál Publicdata.eu a jak bude tato integrace případně realizována. Integrace s tímto

portálem může být zdrojem cenných zkušeností pro budoucí integraci s celoevropským datovým portálem.

2.2.2.4 Potřeba definovaných procesů a legislativně podpořených procesů tvorby a údržby katalogu

Nestačí mít k dispozici funkční katalogizační nástroj. Aby mohla katalogizace dat veřejné správy efektivně probíhat, je třeba formulovat příslušné procesy pro katalogizaci těchto dat a stanovit odpovědnosti za provádění těchto procesů. Katalogizaci dat veřejné správy a s ní spojené procesy a kroky může být nutné promítnout do legislativy nebo do vnitřních procesů subjektů provádějících katalogizaci.

Doporučení pro ČR: Analyzovat požadavky katalogizace dat na legislativu. Formulovat procesy katalogizace otevřených dat VS ČR a příp. návrhy na úpravy relevantní legislativy.

2.2.2.5 Existence jednoznačné metodiky určující co katalogizovat

Ne všechna data veřejné správy mohou být publikovaná jako otevřená data. Důvodem jsou např. požadavky na ochranu osobních údajů nebo ochrana údajů z důvodu jejich významu pro bezpečnost státu a jejich možné utajení. Metodika jasně určující, jaká data mohou být publikovaná jako otevřená data a katalogizovaná v katalogu otevřených dat VS ČR, by měla usnadnit práci osobám vykonávajícím katalogizaci dat.

Doporučení pro ČR: Vytvořit metodiku pro publikování otevřených dat VS ČR a jejich katalogizaci v katalogu otevřených dat VS ČR.

2.2.2.6 Použití jednotných taxonomií pro klasifikaci dat

Vyhledatelnost dat z určité oblasti se zlepší, pokud pro klasifikaci dat i pro klasifikaci záznamů o těchto datech budou používány jednotné taxonomie. Příkladem takovéto harmonizace taxonomií je použití slovníku EUROVOC pro portál Publicdata.eu [8].

Doporučení pro ČR: Definovat jednotné taxonomie pro klasifikaci katalogizačních záznamů a také formulovat způsob jejich užívání při katalogizaci otevřených dat VS ČR. Identifikovat taxonomie, které budou využity v rámci celoevropského datového portálu, protože jejich využití může být potřeba zohlednit při integraci s tímto portálem. Analyzovat možnosti využití slovníku EUROVOC pro Datový katalog.

2.2.2.7 Jasně vymezení práv a povinností při využívání otevřených dat

Jasně vymezení práv a povinností při využívání otevřených dat posiluje právní jistotu jejich uživatelů. Proto ve světě provozované katalogy dat veřejné správy uvádějí v rámci informací o datech i údaj o tom, jaká je licence těchto dat nebo jaké jsou podmínky jejich využití.

Doporučení pro ČR: Jasně vymezit pojem otevřená data z hlediska českého právního řádu a v rámci VS ČR definovat jednotný přístup k licencování otevřených dat VS ČR. V Datovém katalogu uvádět údaje o licencích otevřených dat VS ČR.

2.3 Přehled nástrojů pro katalogizaci dat

Jak již vyplývá z předcházející kapitoly, pro datové katalogy jsou používány různé nástroje. V této kapitole jsou nejprve prezentovány obecné možnosti, které připadají v úvahu při budování datového katalogu, dále jsou prezentovány obecné vlastnosti katalogizačního nástroje a následně jsou zde stručně představeny využívané softwarové nástroje pro datové katalogy.

Přestože na obecné úrovni budou v této kapitole stručně prezentovány i jiné možnosti pořízení katalogizačního nástroje, budou blíže představeny pouze specializované katalogizační nástroje,

kteře jsou běžně využívány pro katalogy otevřených dat ve světě, protože právě využívání těchto samostatných specializovaných nástrojů se v současné době zdá být běžnou praktikou katalogizace otevřených dat veřejné správy.

2.3.1 Obecné možnosti volby nástroje

V této sekci jsou pouze stručně uvedeny obecné možnosti pořízení/vybudování softwarového nástroje pro Datový katalog, které jsou následující:

- výběr a nasazení specializovaného katalogizačního nástroje,
- výběr a nasazení jiného nástroje, který bude upraven pro plnění úlohy datového katalogu,
- vývoj katalogizačního nástroje na míru,
- úprava některého ze stávajících informačních systémů veřejné správy.

První z uvedených možností je výběr a nasazení specializovaného katalogizačního nástroje. V rámci této možnosti je podle potřeb VS ČR a v souladu s požadavky vyplývajícími z této Koncepce vybrán specializovaný softwarový nástroj, který je přímo navržen pro podporu katalogizace dat. Tento nástroj je nakonfigurován dle definovaných požadavků a nasazen pro použití. Hlavní výhodou této varianty je, že se vybírá nástroj, který je primárně určen pro katalogizaci dat a základní funkčnost pro tuto aktivitu tak není třeba do nástroje doprogramovat. Nevýhodou je, že těchto nástrojů je v současné době na trhu pouze omezené množství. Protože se jedná o typizovaný nástroj, je třeba počítat s nutností úprav dle požadavků VS ČR.

Druhou možnost představuje výběr a nasazení jiného nástroje, který bude upraven pro plnění úlohy datového katalogu. Jedná se tedy o výběr a nasazení jiného typizovaného řešení, jež není specializovaným katalogizačním nástrojem, ale které lze upravit tak, aby plnilo roli katalogizačního nástroje. Může se například jednat o úpravu nástroje pro správu obsahu²⁰ (CMS) nebo o úpravu nástroje pro webové portály typu Wikipedia. Výhodou je, že těchto nástrojů je na trhu značné množství. Jsou k dispozici nástroje od renomovaných výrobců i nástroje vyvíjené jako open source software dostupné zdarma. Nevýhodou varianty je její časová a finanční náročnost na úpravu zvoleného nástroje do požadované podoby katalogizačního nástroje.

Třetí variantou je vývoj vlastního katalogizačního nástroje na míru. Výhodou je, že vzniklý nástroj by plně splňoval požadavky na datový katalog VS ČR. Nevýhodou je časová a finanční náročnost na jeho vytvoření. Vzhledem k existenci několika nástrojů určených přímo pro katalogizaci dat a především řady CMS nástrojů či nástrojů typu Wikipedia, které lze pro potřeby katalogizace upravit, jeví se vhodnější zvážit nejprve vybudování Datového katalogu spíše pomocí těchto variant, než vývojem katalogizačního nástroje na míru.

Poslední navrhovanou variantou je úprava některého ze stávajících informačních systémů veřejné správy tak, aby kromě své stávající funkce plnil i roli datového katalogu. Výhodou varianty je, že by se jednalo o již existující nástroj, pro který je vybudována potřebná infrastruktura. Samozřejmě by bylo třeba posoudit, zda by tato infrastruktura byla dostačující i po rozšíření systému. Nevýhoda spočívá v tom, že úpravy stávajícího systému mají své meze a v rámci těchto mezí by nemuselo být možné plnohodnotný datový katalog realizovat.

²⁰ Mezi známé volně dostupné open source CMS nástroje patří nástroje Drupal (<http://drupal.org/>) a Joomla (<http://www.joomla.org/>).

2.3.2 Vlastnosti katalogizačního nástroje

Na základě dosavadních zkušeností autorů s katalogizací otevřených dat a s katalogizačním nástrojem CKAN a po vyhodnocení funkcí poskytovaných dalším katalogizačním nástrojem Socrata lze říci, že specializované katalogizační nástroje poskytují funkce v oblastech zachycených na následujícím obrázku. Aby mohl katalogizační nástroj samostatně fungovat, je třeba, aby disponoval nejen specifickými funkcemi pro podporu katalogizace otevřených dat, ale měl by disponovat i určitými obecnými funkcemi, které lze najít i u jiných nástrojů a systémů. Mezi tyto funkce patří funkce ve skupinách Bezpečnost, Ukládání dat, Statistiky a analýzy a případně i Vizualizace dat a Podpora spolupráce, protože tyto funkce mohou být poskytovány jinými specializovanými nástroji.

Dále uvedený popis skupin funkcí katalogizačních nástrojů přeložen a upraven dle [26].



Obrázek 2: Funkce katalogizačního nástroje (přeloženo z [26])

Funkce pro katalogizaci a publikaci katalogizačních záznamů tvoří spolu s funkcemi pro vyhledávání a procházení obsahu datového katalogu jádro jeho funkcí. Tyto funkce slouží pro vkládání a úpravu katalogizačních záznamů a pro zpřístupnění těchto záznamů uživatelům datového katalogu. Nástroj pro datový katalog by měl umožnit označování záznamů tématy z dohodnuté taxonomie (např. pomocí značek – tagů). Nástroj by také měl uchovávat záznamy o změnách provedených v záznamech a měl by umožnit návrat k předcházejícím verzím záznamů.

Funkce pro vyhledávání a procházení obsahu datového katalogu jsou nezbytné, aby uživatelé datového katalogu mohli získat katalogizační záznamy, které potřebují. Uživatelé by tak měli mít k dispozici plnotextové vyhledávání i vyhledávání podle kategorií a skupin, do kterých jsou

členěny záznamy katalogu. Pokročilé vyhledávání by mělo zahrnovat vyhledávání podle jednotlivých atributů dat, jako je například typ souboru s daty apod. Některé katalogizační nástroje také umožňují vyhledávat data podle jejich příslušnosti ke geografické oblasti, např. pomocí označení požadované oblasti na zobrazené mapě.

Funkce pro řízení kvality obsahu katalogu slouží pro zajištění platnosti a úplnosti údajů o otevřených datech uvedených v datovém katalogu. Tyto funkce tak např. umožňují kontrolu platnosti uváděných odkazů na datové zdroje nebo kontrolu chybějících údajů v záznamech datového katalogu.

Funkce pro ukládání dat a metadat zajišťují uložení obsahu datového katalogu. Datový katalog musí být schopen primárně ukládat katalogizační záznamy, které obsahují metadata o otevřených datech veřejné správy. Některé nástroje pro datový katalog umožňují i vkládání samotných dat do datového katalogu.

Funkce pro vizualizaci dat pomáhají uživatelům datového katalogu lépe pochopit tam katalogizovaná data. Existující katalogizační nástroje zpravidla umožňují zobrazení obsahu dat ve formě tabulek nebo případně i zobrazení dat na mapovém podkladě, pokud se data vztahují k nějakým geografickým lokacím.

Statistické a analytické funkce slouží pro získání základního přehledu o obsahu datového katalogu, jako jsou například informace o tom, kolik katalog obsahuje záznamů, na kolik datových zdrojů záznamy odkazují nebo jaká je frekvence změn v záznamech datového katalogu. Pokud jsou poskytovány pokročilejší funkce pro statistiku a analýzy, bývá umožněno, aby uživatelé prováděli vlastní analytické dotazy týkající se obsahu datového katalogu.

Bezpečnostní funkce poskytované nástrojem pro datový katalog by měly zahrnovat alespoň funkce pro zálohování obsahu datového katalogu a pro provedení obnovy jeho obsahu ze zálohy. Katalogizační nástroj by měl disponovat také mechanismy pro řízení přístupu k obsahu datového katalogu, aby bylo možné vymezit různé skupiny uživatelů s různou úrovní přístupu, díky čemuž je možné rozlišit uživatele oprávněné vkládat a měnit obsah katalogu od těch, kteří jsou oprávněni obsah katalogu pouze prohlížet.

Funkce pro podporu spolupráce jsou zpravidla reprezentovány funkcemi pro poskytování zpětné vazby ohledně obsahu datového katalogu. Může se tak jednat o možnost umisťovat ke katalogizačním záznamům zprávy, nebo možnost hodnotit záznamy katalogu na předem určené škále (líbí/nelíbí; kvalita hodnocená pomocí hvězdiček na škále od nuly do pěti apod.).

Katalogizační nástroj může disponovat i dalšími funkcemi. Zmiňme alespoň vícejazyčné uživatelské rozhraní a vícejazyčné verze katalogizačních záznamů, které pomáhají využití obsahu datového katalogu v mezinárodním prostředí. Vícejazyčná metadata v datovém katalogu by bylo vhodné zajistit s ohledem na evropský portál Publicdata.eu [8]. Kromě těchto jazykových funkcí je pro technickou integraci datových katalogů a pro automatizované využívání jeho obsahu vhodné, aby katalogizační nástroj disponoval aplikačním rozhraním (API).

2.3.3 Používané specializované katalogizační nástroje

Popisem existujících nástrojů pro správu otevřených dat se zabývá zdroj [21], přičemž jsou zde popisovány nástroje CKAN, Socrata a Ontowiki. Zařazení nástrojů CKAN a Socrata potvrzuje poznatky o katalogizačních nástrojích používaných v zahraničních iniciativách zaměřených na otevřená data. Další zmíněný nástroj, Ontowiki, je dle [21] specializovaný nástroj pro práci s daty ve formátu RDF. Vzhledem k tomuto zaměření nebude tento nástroj dále uvažován. V rámci iniciativy společnosti Microsoft Corporation nazvané Open Government Data Initiative [30] vzniká další nástroj pro katalogizaci otevřených dat veřejné správy: OGD DataLab [35]. V dalším textu kapitoly budou nástroje CKAN, Socrata a OGD DataLab blíže představeny.

2.3.3.1 CKAN

Comprehensive Knowledge Archive Network (CKAN) je řešení pro datové katalogy a datové portály [39]. CKAN je vyvíjen jako open source software neziskovou organizací Open Knowledge Foundation. Jak již bylo uvedeno v předešlém textu, CKAN je využíván pro řadu datových katalogů veřejné správy ve světě a je také využit pro připravovaný evropský portál Publicdata.eu. Početná komunita uživatelů tohoto nástroje je také sdružena okolo portálu TheDatahub.org.

Tabulka 7: Základní údaje o nástroji CKAN

Výrobce	Open Knowledge Foundation (OKFN)
Odkaz	http://ckan.org
Licence	GNU Affero General Public Licence v3 ²¹
Dostupnost	Nástroj samotný je zdarma pro libovolné využití, lze pořídit i jako placenou službu od OKFN [42]
Lokalizace do češtiny	Dostupná
API	Dostupné. API je založené na architektuře REST, datové zprávy využívají formát JSON nebo JSONP. S využitím rozšíření lze k obsahu datového katalogu přistupovat pomocí protokolu a dotazovacího jazyka SPARQL, data jsou pak získána ve formátu RDF [41].

CKAN je od počátku vyvíjen jako specializovaný nástroj pro datový katalog a disponuje tak pro podporu této oblasti širokou paletou funkcí, které jsou průběžně vylepšovány a doplňovány. Pro nástroj CKAN je k dispozici i řada rozšíření, která rozšiřují nástroj o další funkce a vývoj vlastního rozšíření je také jedna z možností, jak přizpůsobit CKAN podmínkám konkrétní organizace. Nástroj CKAN také disponuje aplikačním rozhraním (API), které lze využít pro přístup k jeho obsahu z ostatních aplikací nebo pro integraci tohoto nástroje s ostatními informačními systémy. Díky četnému využití pro portály otevřených dat veřejné správy ve světě se nástroj CKAN jeví být vyspělým nástrojem pro datový katalog, který je díky zmíněným vlastnostem i dostatečně flexibilní.

2.3.3.2 Socrata

Socrata je platforma pro správu a publikaci dat veřejné správy, která je vyvíjená a poskytována společností Socrata Inc. se sídlem v USA [53]. Dle [21] je Socrata nejrozšířenějším nástrojem pro otevřená data v rámci vlády USA.

Tabulka 8: Základní údaje o nástroji Socrata

Výrobce	Socrata, Inc.
Odkaz	http://www.socrata.com/
Licence	Proprietární

²¹ <http://www.gnu.org/licenses/agpl-3.0.html>

Dostupnost	Neplacená a placená služba, neplacená služba je omezená z hlediska dostupných funkcí, objemu dat a pouze jeden uživatel může spravovat a upravovat obsah datového katalogu [52]
Lokalizace do češtiny	Nedostupná
API	Dostupné. API je založeno na architektuře REST, při využití API lze pracovat s daty ve formátech JSON, XML, RDF, CSV, XLS [51].

Nástroj Socrata je orientován nejen na katalogizaci vlastních dat a datových zdrojů, ale také na prezentaci obsahu dat. Nabízí tak rozsáhlé funkce, které umožňují uživatelům procházet a analyzovat obsah katalogizovaných datových zdrojů. Subjekty publikující data také mohou uživatelům nabídnout například vizualizaci dat na mapovém podkladě. Stejně jako CKAN i Socrata disponuje aplikačním rozhraním pro interakci s dalšími aplikacemi. Socrata využívá model Software-as-a-Service (SaaS, software jako služba) a primárně je dostupná jako služba placená (lze si vybrat z několika cenových tarifů, pro údaje o cenách je třeba kontaktovat představitele společnosti Socrata Inc.). K dispozici je i služba zdarma, která je ale omezena z hlediska funkcí, objemu spravovaných dat i tím, že obsah katalogu může spravovat a vytvářet pouze jeden uživatel.

2.3.3.3 OGD I DataLab

OGDI DataLab je nástroj pro katalogizaci dat veřejné správy, jehož vývoj byl iniciován společností Microsoft Corporation v rámci iniciativy Open Government Data Initiative (OGDI) [30].

Tabulka 9: Základní údaje o nástroji OGD I DataLab

Výrobce	Microsoft Corporation ve spolupráci s komunitou
Odkaz	http://ogdisdk.cloudapp.net/ , https://github.com/openlab/DataLab
Licence	Microsoft Public License (Ms-PL)
Dostupnost	Nástroj samotný je k dispozici zdarma, pro jeho provoz je třeba využívat služeb platformy Microsoft Windows Azure. K dispozici jsou různé tarify [33], platformu je možné zdarma vyzkoušet [32].
Lokalizace do češtiny	Nedostupná
API	Dostupné. API je založeno na architektuře REST, využívá Open Data Protocol (OData), při práci s geodaty je využíván formát KML, dále je možné pracovat s daty ve formátech JSON a JSONP [35].

Stejně jako nástroj Socrata, i OGD I DataLab se zaměřuje nejen na samotnou katalogizaci dat, ale i na přístup k datům a jejich vizualizaci. Disponuje tak funkcemi pro procházení dat, jejich vizualizaci včetně vizualizace na mapových podkladech, analýzu dat a stahování dat [35]. Nástroj OGD I DataLab také disponuje aplikačním rozhraním (API). Toto rozhraní mimo jiné využívá Open Data Protocol (viz kapitola 2.1.1), který je přímo navržen pro práci s datovými zdroji. Nástroj samotný je vyvíjen jako open source software a je určen pro platformu Microsoft Windows Azure [35].

2.3.4 Další katalogizační nástroje

Kromě výše představených specializovaných katalogizačních nástrojů existují i další nástroje, které mohou vystupovat v roli datového katalogu, případně nejsou primárně orientované jako samostatný datový katalog, ale určitou formu katalogu v sobě zahrnují. Příklady takovýchto dalších nástrojů jsou uvedeny v následujících odstavcích.

Pro potřeby aplikací Business Intelligence a řízení dat se v organizacích využívají i katalogizační nástroje zaměřené na katalogizaci jednotlivých datových entit v databázích a datových zdrojích organizace. Tyto katalogizační nástroje bývají součástí nástrojů pro řízení kvality dat a správu metadat a mají velký význam pro identifikaci a řízení tzv. master dat – klíčových číselníků či referenčních dat [3].

Jako určitá forma katalogu dat vystupuje Data Market platformy Microsoft Windows Azure [28]. Data Market slouží k nabídce dat a datových zdrojů a to zdarma nebo za poplatek. Obdobně jako elektronické obchody, které obsahují katalog nabízených produktů a služeb, Data Market obsahuje katalog nabízených dat.

2.4 2.4 Přístupy k plnění, provozu a správě datového katalogu VS ČR

Světové zkušenosti ukazují, že pro budování datových katalogů jsou aplikovány různé přístupy, a proto i v ČR je třeba zvážit, jaký způsob budování Datového katalogu je vhodný. V této kapitole jsou nejprve vymezeny jednotlivé aspekty, které je třeba při budování Datového katalogu vzít v úvahu a následně je na základě těchto aspektů vymezeno několik základních variant vybudování katalogu otevřených dat VS ČR. V závěru kapitoly jsou navrženy varianty vyhodnoceny a jsou formulována doporučení pro budování Datového katalogu.

2.4.1 Přístupy k plnění datového katalogu VS ČR

2.4.1.1 Přístupy k vkládání záznamů

Aspekt vkládání souvisí s tím, kdo by měl být oprávněn vkládat a upravovat záznamy Datového katalogu (plnění Datového katalogu). Přístupy k plnění datového katalogu jsou následující.

K vkládání a úpravě záznamů datového katalogu může být oprávněn pouze jeden pověřený subjekt. Tento subjekt by tak musel katalogizovat otevřená data všech ostatních orgánů veřejné správy. Musel by tak zmapovat místa, kde ostatní orgány veřejné správy publikují svá otevřená data, pravidelně je sledovat a vytvářet katalogizační záznamy pro nově publikovaná data, případně aktualizovat existující záznamy v případě změn.

Druhým přístupem je situace, kdy pro plnění datového katalogu je oprávněno více pověřených subjektů. Ty by tak katalogizovaly otevřená data, která jsou v rámci VS ČR publikována. V rámci implementace lze uvažovat o tom, zda by měla být katalogizací pověřena omezená množina orgánů veřejné správy nebo zda by byl každý orgán veřejné správy povinen katalogizovat svá otevřená data.

Posledním přístupem je situace, kdy Datový katalog může plnit kdokoli včetně veřejnosti. Subjekty by tak nebyly pro plnění Datového katalogu autorizovány.

Tabulka 10: Výhody a nevýhody přístupů k vkládání záznamů

	Jeden pověřený subjekt	Více pověřených subjektů	Kdokoli
Výhody	(+) garance jednotného způsobu vyplňování katalogizačních záznamů (+) snazší zajištění (školení) dodržování jednotné metodiky	(+) menší nároky na kapacity u každého pověřeného subjektu (+) větší znalost katalogizovaných dat a tudíž přesnější záznamy	(+) rozložení pracovní na velkou skupinu uživatelů (crowdsourcing) (+) záznam o zajímavých datech může vložit i subjekt mimo VS
Nevýhody	(-) vysoké kapacitní nároky na jeden subjekt (-) menší znalost katalogizovaných dat (katalogizátor není expert ne všechna katalogizovaná data) a nebezpečí nepřesnosti záznamů	(-) horší garance dodržování jednotné metodiky (-) nebezpečí rozdílných způsobů katalogizace shodných dat	(-) katalogizovaná i nevhodná data (-) zahlcení katalogu (-) negarantovaná jednotná metodika (-) nebezpečí neúplných nebo záměrně zkreslených záznamů

2.4.1.2 Rozsah redakce

Aspekt rozsahu redakce se zabývá otázkou, zda a případně na jaké části obsahu Datového katalogu by měla probíhat redakce. Redakcí se rozumí posuzování katalogizačních záznamů z hlediska toho, zda jsou vytvářeny pro skutečně publikovaná otevřená data, zda jsou u katalogizačních záznamů vyplněny všechny povinné atributy, zda jsou dodržena pravidla používání taxonomií pro klasifikaci dat a záznamů nebo zda nedochází ke katalogizaci nevhodného obsahu. Redakce obsahu Datového katalogu je tak prostředkem zajištění kvality jeho obsahu.

Samostatnou otázkou je správnost údajů uvedených v záznamech Datového katalogu. Za správnost údajů v Datovém katalogu by měly být odpovědné ty orgány veřejné správy, které vytvářejí záznamy v datovém katalogu. Pokud by za správnost údajů měl zodpovídat subjekt odpovědný za redakci obsahu Datového katalogu, musel by být oprávněn provádět kontrolu správnosti uvedených údajů u orgánů veřejné správy, které vytvářejí a publikují otevřená data, což by mohlo být značně náročné.

Další samostatnou otázkou spojenou s redakcí obsahu Datového katalogu je podrobnost prováděné redakce, tedy jak důkladně jsou jednotlivé atributy katalogizačních záznamů redigovány, zda jsou redigovány všechny atributy katalogizačních záznamů, nebo jen některé. Redakce každého záznamu může být prováděna pouze zevrubně s cílem odhalit zejména opravdu závažné nedostatky v katalogizačních záznamech, nebo může být prováděna opravdu důkladně s cílem zajistit co možná nejvyšší přesnost a úplnost katalogizačních záznamů. V rámci tohoto aspektu nebude hloubka či podrobnost redakce dále hodnocena. Hloubka prováděné redakce by měla být stanovena v metodice katalogizace otevřených dat veřejné správy.

Možné přístupy k redakci datového katalogu z hlediska rozsahu této redakce jsou následující.

Prvním přístupem je úplná redakce, kdy redakci podléhá veškerý obsah datového katalogu. Veškeré nové záznamy katalogu i změny existujících záznamů musí být nejprve podrobeny redakci. Pouze ověřené záznamy jsou zobrazeny uživatelům Datového katalogu.

Druhým přístupem je částečná redakce, kdy redakci podléhá pouze část obsahu Datového katalogu. Tato část může být vymezena z věcného hlediska, kdy redakci podléhá pouze určitý typ dat, respektive záznamů o těchto datech, nebo redakci podléhají záznamy pouze od určitého typu jejich původců. Záznamy Datového katalogu jsou označeny podle toho, zda prošly redakcí (ověřené záznamy) či nikoli (neověřené záznamy). Rozlišení záznamů umožňuje zpřístupnit uživatelům Datového katalogu všechny záznamy. Přístup částečné redakce lze aplikovat i tak, že záznamy datového katalogu jsou v něm zpřístupněny okamžitě po jejich vytvoření/změně ve stavu neověřených záznamů a následně postupně tak, jak procházejí redakcí, přecházejí záznamy do stavu ověřených záznamů. Rychlost zveřejnění je v některých případech důležitým aspektem využitelnosti dat. Je tedy nutné tuto variantu redakce uvažovat.

Posledním přístupem je situace, kdy redakce obsahu Datového katalogu není vůbec aplikována. Tento přístup se může zdát nesmyslný. Ovšem ve spojení s přístupem umožňujícím vkládat nové záznamy pouze autorizovaným subjektům jistě smysl má.

Tabulka 11: Výhody a nevýhody přístupů k rozsahu redakce

	Úplná redakce	Částečná redakce	Bez redakce
Výhody	(+) vyšší kvalita všech záznamů	(+) vyšší kvalita záznamů v redigované části (+) redigované a neredigované záznamy jsou odlišeny (+) uživatelé katalogu vidí všechny záznamy (i neredigované)	(+) odpadá časová náročnost (+) odpadá finanční náročnost
Nevýhody	(-) nutnost určit redaktora (-) neredigované záznamy nejsou přístupné uživatelům katalogu (-) vysoká náročnost časová (-) vysoká náročnost finanční	(-) nutnost určit redaktora (-) náročnost časová (-) náročnost finanční	(-) nižší kvalita záznamů (-) nemožnost zajistit kvalitu záznamů

2.4.2 Přístupy k provozu a správě datového katalogu VS ČR

Při budování datového katalogu je třeba zvážit, zda bude budován „odspoda“ či „odshora“, zda použít jednotnou nebo heterogenní softwarovou platformu datového katalogu, kdo bude oprávněn vkládat záznamy do datového katalogu, zda a jak bude probíhat redakce obsahu katalogu a také, jak bude provozován nástroj pro datový katalog. Níže jsou popsány možné přístupy k jednotlivým aspektům a jsou uvedeny jejich výhody a nevýhody.

2.4.2.1 Distribuovanost / centralizace datového katalogu

Distribuovanost datového katalogu souvisí s počtem provozovaných katalogů a jejich distribucí mezi jednotlivé orgány veřejné správy. Možnými přístupy k distribuovanosti jsou přístup „odshora“, kdy by vznikl jeden centrální katalog pro celou VS ČR a přístup „odspoda“, kdy by existovalo více oddělených datových katalogů pro vybrané orgány veřejné správy.

Jednou z možností, jak implementovat přístup „odspoda“ je vytvoření jednoho zvláštního datového katalogu, do kterého by byly automatizovaně přenášeny záznamy z jednotlivých ostatních katalogů. Tím by bylo dosaženo stavu, kdy lze vyhledávat otevřená data veřejné správy z jednoho místa, ale jeho obsah by byl tvořen obsahem katalogů spravovaných vybranými orgány veřejné správy, které by si mohly upravit postupy plnění podle svých potřeb.

Tabulka 12: Výhody a nevýhody přístupů k distribuovanosti datového katalogu

	Centrální katalog	Oddělené katalogy
Výhody	(+) centralizace (+) snazší kontrola kvality (+) snazší přístup veřejnosti do katalogu	(+) přizpůsobení potřebám jednotlivých orgánů veřejné správy (+) lépe odpovídá architektuře webu (je decentralizovaná)
Nevýhody	(-) vyšší nároky na technologické zabezpečení (-) vyšší náklady	(-) odlišné struktury metadat (jejich hodnot, např. jiné využití dohodnutých značek) (-) horší přístup veřejnosti (-) horší kontrola kvality a vzájemné konzistence

2.4.2.2 Heterogennost / homogennost katalogizačních nástrojů

Heterogennost katalogizačních nástrojů je spojena se softwarovým nástrojem pro datový katalog a otázkou, zda by tento softwarový nástroj měl být jednotný v rámci VS ČR (homogenní prostředí), nebo zda by mělo být jednotlivým orgánům veřejné správy umožněno vybrat si vlastní nástroj (heterogenní prostředí). Z tohoto již také vyplývá, že o heterogenním prostředí je smysluplná uvažovat pouze v případě, že je u aspektu A aplikován přístup „odspoda“, tedy oddělené katalogy spravované vybranými orgány veřejné správy. V případě, že je aplikován přístup centrálního datového katalogu, bude používán pouze jeden softwarový nástroj pro datový katalog a prostředí tak bude homogenní.

Tabulka 13: Výhody a nevýhody přístupů k homogenosti katalogizačních nástrojů

	Homogenní prostředí	Heterogenním prostředí
Výhody	(+) snazší propojení mezi oddělenými katalogy po technické stránce (+) snazší údržba v rámci VS (+) úspory z rozsahu v rámci VS jako celku - výhodnější ceny, školení, centrální metodika používání	(+) nezávislost na jedné platformě (+) nezávislost na jednom dodavateli v rámci celé VS (+) snazší přizpůsobení podmínkám jednotlivých orgánů veřejné správy
Nevýhody	(-) závislost na jednom výrobcí dodavateli (-) závislost na jedné platformě	(-) obtížnější technologická integrace (-) vyšší finanční náklady integrace

2.4.2.3 Provoz katalogizačního nástroje

Aspekt provozu katalogizačního nástroje se týká otázky, kdo bude katalogizační nástroj provozovat. Z hlediska možných přístupů k této otázce může být provozovatelem katalogizačního nástroje samotný orgán veřejné správy nebo soukromý subjekt. Tento aspekt ale nebude mít vliv na posuzování variant v této části dokumentu a bude podrobněji popsán až v části 3. Plán realizace koncepce.

2.4.3 Varianty budování Datového katalogu

V této kapitole jsou prezentovány základní varianty, které pro budování Datového katalogu připadají v úvahu. Variantou budování Datového katalogu se rozumí adekvátní kombinace přístupů v rámci výše popsaných aspektů budování datového katalogu. V této kapitole nejsou

analyzovány všechny existující kombinace přístupů k jednotlivým aspektům, ale pouze takové kombinace, které byly autory vyhodnoceny jako vhodné pro další analýzu.

Protože je cílem kapitoly také formulovat doporučení pro budování Datového katalogu, jsou jednotlivé varianty budování zhodnoceny z hlediska bezpečnosti Datového katalogu, z hlediska finanční náročnosti na vybudování a následný provoz, z hlediska dopadu varianty na kvalitu obsahu Datového katalogu a také z hlediska propojení s jinými evidenčními systémy (IS o ISVS a ISDP). Propojitelnost s jinými evidenčními systémy je zde chápána jak z technologického hlediska, tak i z hlediska obsahového, protože je třeba zajistit nejen technologické propojení jednotlivých systémů, ale i to, aby byla sjednocena struktura záznamů a používané taxonomie.

Autoři dokumentu provedli hodnocení variant tak, že každý hodnotil jednotlivé varianty ve výše zmíněných dimenzích přiřazením hodnoty na stupnici 1 až 4, kde 1 znamená nejlepší výsledek a 4 znamená nejhorší výsledek. Detailní vyhodnocení variant je uvedeno v příloze A. V dalším textu je u každé varianty uveden pouze součet udělených bodů (součet se vzhledem k počtu hodnotitelů mohl pohybovat v rozmezí 5 až 20 bodů, čím méně bodů, tím lepší hodnocení varianty).

Finanční náročnost Datového katalogu ve fázi provozu ovlivní kromě rozsahu prováděné redakce také její hloubka či podrobnost. Při vyhodnocování finančních nároků provozu jednotlivých variant byla ve všech variantách uvažována stejná hloubka prováděné redakce.

Vymezení jednotlivých variant z hlediska kombinace přístupů v jednotlivých aspektech budování datového katalogu ilustruje následující tabulka. Aspekt provozu není zohledněn, protože u každé z variant se příslušný orgán veřejné správy může rozhodnout, zda bude katalogizační nástroj provozovat sám, nebo zda provoz svěří jinému subjektu.

Tabulka 14: Varianty budování datového katalogu

Varianta/aspekt	Distribuvanost	Heterogenost	Vkládání	Redakce
A	Centrální katalog	Homogenní prostředí	Více pověřených subjektů	Úplná redakce
B	Centrální katalog	Homogenní prostředí	Více pověřených subjektů	Částečná redakce
C	Oddělené katalogy	Homogenní prostředí	Více pověřených subjektů	Úplná redakce ²²
D	Oddělené katalogy	Heterogenní prostředí	Kdokoli	Úplná redakce
E	Oddělené katalogy	Heterogenní prostředí	Více pověřených subjektů	Úplná redakce
F	Oddělené katalogy	Heterogenní prostředí	Více pověřených subjektů	Částečná redakce

²² Každý subjekt pověřený a oprávněný vkládat záznamy do Datového katalogu provádí úplnou redakci na datovém katalogu, který spravuje.

2.4.3.1 Varianta A

V této variantě je navržen vznik jednoho centrálního datového katalogu pro celou veřejnou správu, do kterého by jednotlivé orgány veřejné správy vkládaly záznamy o svých otevřených datech. Orgán veřejné správy pověřený správou Datového katalogu by prováděl redakci veškerého obsahu Datového katalogu. Vzhledem k tomu, že se jedná o jeden centrální Datový katalog, byl by pro něj vybrán jeden softwarový nástroj.

Jak již vyplývá z vyhodnocení výhod a nevýhod jednotlivých přístupů, výhoda této varianty spočívá v centralizaci. Otevřená data VS ČR jsou katalogizována na jednom místě a uživatelé je mohou z tohoto místa vyhledávat. V centrálním katalogu je také snazší kontrolovat kvalitu záznamů, zajištění této kvality je také podpořeno úplnou redakcí obsahu katalogu.

Nevýhodou varianty jsou vyšší nároky na její zajištění po technické stránce, kdy Datový katalog musí být dimenzován na využívání celou veřejnou správou i všemi potenciálními uživateli. Zajištění úplné redakce celého obsahu Datového katalogu může být náročné jak z hlediska času potřebného pro provedení redakce, tak i z hlediska vynaložených finančních prostředků.

Tabulka 15: Zhodnocení varianty A

	Bezpečnost	Finanční náročnost vybudování	Finanční náročnost provozu	Kvalita záznamů katalogu	Propojitelnost s jinými evidenčními systémy
Hodnocení	5	8	13	5	5

2.4.3.2 Varianta B

Tato varianta se podobá Variantě A. I zde je navržen vznik jednoho centrálního Datového katalogu, do kterého by orgány veřejné správy vkládaly záznamy o svých otevřených datech. Oproti Variantě A by ale redakce probíhala pouze na části obsahu tak, že by záznamy katalogu byly ihned zpřístupněny uživatelům Datového katalogu a orgán pověřený jeho správou by postupně zajišťoval jejich redakci. Softwarové prostředí Datového katalogu by i v tomto případě bylo homogenní.

Výhodu této varianty v porovnání s Variantou A lze spatřovat v tom, že by nedocházelo k prodlevě mezi založením záznamu v Datovém katalogu a jeho zpřístupněním jeho uživatelům způsobenou redakcí. Pokud není uvažována prodleva mezi zpřístupněním určitých dat a založením záznamu o nich do Datového katalogu, dozvěděli by se potenciální uživatelé o těchto datech dříve, než v případě Varianty A. Výhodou pro orgán veřejné správy pověřený správou datového katalogu a redakcí jeho obsahu je zde výhoda, že redakci je možné provádět postupně a nároky na čas, finance a personální zajištění redakce tak mohou být nižší než v případě Varianty A.

Samozřejmě je zde riziko, že se k uživatelům budou dostávat méně kvalitní záznamy než v případě Varianty A, kde jsou zpřístupňovány pouze záznamy ověřené v rámci redakce. Nicméně ověřené a neověřené záznamy budou jasně označené a uživatelé se budou moci podle tohoto údaje rozhodovat, jak k záznamům přistoupit. Uživatelé nebudou muset čekat, až dojde k redakci a záznam o datech, jim bude zpřístupněn ihned po zveřejnění.

Tabulka 16: Zhodnocení varianty B

	Bezpečnost	Finanční náročnost vybudování	Finanční náročnost provozu	Kvalita záznamů katalogu	Propojitelnost s jinými evidenčními systémy
Hodnocení	5	8	8	13	9

2.4.3.3 Varianta C

Zatímco předcházející dvě varianty představovaly budování Datového katalogu „odshora“, tedy od centrálního Datového katalogu, představuje tato varianta budování Datového katalogu „odspoda“. V rámci varianty je navržen vznik samostatných datových katalogů u vybraných orgánů veřejné správy. Svůj datový katalog by mohl mít každý orgán veřejné správy, nicméně předpokládáme, že zejména pro malé orgány VS (malé obce) by vybudování takového katalogu mohlo být příliš náročné, tudíž je možné, aby se více orgánů veřejné správy spojilo a společně sdílely jeden datový katalog. Zároveň je doporučeno, aby vznikl i centrální Datový katalog, který by ale nebyl plněn přímo, nýbrž automatizovaně tak, že by do něj byly přenášeny záznamy z jednotlivých dílčích datových katalogů. Centrální datový katalog by sloužil zejména jako jednotný bod pro vyhledávání. Pro usnadnění jeho automatizovaného plnění je v této variantě navrženo, aby všechny oddělené datové katalogy včetně centrálního používaly homogenní softwarové prostředí, tj. všechny by používaly stejný softwarový katalogizační nástroj. Tato varianta zahrnuje úplnou redakci obsahu prováděnou tak, že každý správce dílčího datového katalogu je pověřen redakcí jeho celého obsahu. Centrální datový katalog nemá samostatnou redakci, protože úplná redakce obsahu jednotlivých dílčích datových katalogů se jeví být dostatečná pro zajištění kvality obsahu Datového katalogu.

Datové katalogy jsou využívány prostřednictvím internetu a výhodou této varianty je, že lépe odpovídá architektuře současného webu, která je také distribuovaná. I když předpokládáme, že by pro vytváření záznamů datového katalogu měla být vydána metodická doporučení, umožňuje tato varianta, aby si jednotlivé orgány veřejné správy přizpůsobily datový katalog a procesy jeho naplňování svým potřebám. Nároky na redakci obsahu datového katalogu budou také distribuovány mezi jednotlivé dílčí datové katalogy.

Vznik oddělených datových katalogů s sebou přináší obtíže při vyhledávání obsahu napříč datovými katalogy, které by ale měly být minimalizovány vznikem automatizované plněného centrálního katalogu. I přes úplnou redakci na jednotlivých dílčích datových katalogích je v případě této varianty větší riziko, že záznamy datového katalogu nebudou zcela konzistentní a že např. klasifikační taxonomie budou používány rozdílně u jednotlivých orgánů veřejné správy.

Tabulka 17: Zhodnocení varianty C

	Bezpečnost	Finanční náročnost vybudování	Finanční náročnost provozu	Kvalita záznamů katalogu	Propojitelnost s jinými evidenčními systémy
Hodnocení	11	12	14	12	14

2.4.3.4 Varianta D

Tato varianta se podobá Variantě C. Varianta je postavena na oddělených datových katalogích, kde jejich správci budou provádět úplnou kontrolu jejich obsahu. I v tomto případě předpokládáme vznik automatizované plněného centrálního katalogu sloužícího pro vyhledávání. Tato varianta se ale od Varianty C liší v přístupech ke dvěma aspektům.

Prvním rozdílem je, že prostředí může být z hlediska používaných katalogizačních nástrojů heterogenní. Každý orgán veřejné správy si tak může vybrat takový katalogizační nástroj, který bude vyhovovat jeho potřebám. Bude ale třeba zajistit, aby tyto katalogizační nástroje byly schopny automatizovaně předávat záznamy do centrálního datového katalogu v požadované struktuře. Dále bude třeba zajistit, aby všechny katalogizační nástroje pracovaly stejně se všemi dohodnutými slovníky a taxonomiemi, aby nejen struktura katalogizačních záznamů odpovídala struktuře v centrálním katalogu, ale aby i standardizované části katalogizačních záznamů používaly stejné pojmy, klasifikační třídy apod. Z tohoto hlediska se jeví jako vhodné, aby jeden

orgán veřejné správy byl odpovědný za vydávání závazných standardů pro datové katalogy ostatních orgánů VS.

Druhým rozdílem je, že v této variantě je navrženo, aby i veřejnost mohla přidávat a upravovat záznamy v datovém katalogu. Díky tomu může záznam o zajímavých datech vložit i osoba mimo veřejnou správu. Výhodou tohoto přístupu je i fakt, že se pracnost plnění datového katalogu může potenciálně rozložit mezi velké množství osob přispívajících do datového katalogu svými záznamy nebo jejich opravami. Varianta s sebou nicméně nese riziko, že metodika tvorby katalogizačních záznamů nebude dodržována. Nebezpečím také je, že by datové katalogy mohly obsahovat i záznamy, které nebudou odkazovat na otevřená data veřejné správy, ale na nevhodný obsah. Datový katalog by také mohl být zahlcen záznamy s malou či žádnou hodnotou pro potenciální uživatele (spam). Případně by se mohly vyskytnout i záznamy neúplné nebo záměrně zkreslené. Proto se jeví nezbytné kombinovat tento přístup s úplnou redakcí obsahu Datového katalogu, protože pak bude katalog obsahovat pouze ověřené záznamy. Redaktoři by spam a záznamy odkazující na nevhodný obsah měli zamítnout.

Tabulka 18: Zhodnocení varianty D

	Bezpečnost	Finanční náročnost vybudování	Finanční náročnost provozu	Kvalita záznamů katalogu	Propojitelnost s jinými evidenčními systémy
Hodnocení	13	14	18	17	19

2.4.3.5 Varianta E

Tato varianta je další variantou budování Datového katalogu „odspoda“. V rámci varianty je navržen vznik více oddělených datových katalogů u vybraných orgánů veřejné správy, přičemž každý z těchto orgánů VS by si mohl zvolit katalogizační nástroj dle potřeby. Softwarové prostředí by tak bylo heterogenní. K vytváření obsahu datového katalogu by bylo oprávněno více subjektů a redakce obsahu datového katalogu by byla úplná. Každý orgán veřejné správy, který by spravoval některý z oddělených datových katalogů, by zároveň prováděl úplnou redakci celého jeho obsahu. Pro usnadnění vyhledávání v těchto datových katalogích je doporučeno vytvořit centrální datový katalog, do kterého by byly automatizovaně přenášeny záznamy z jednotlivých dílčích datových katalogů.

Výhodou varianty je skutečnost, že umožňuje jednotlivým orgánům veřejné správy, aby si zvolily takový katalogizační nástroj, který bude nejlépe vyhovovat jejich potřebám. Úplná redakce by měla pozitivně působit na kvalitu obsahu Datového katalogu, díky distribuovanosti Datového katalogu a heterogenosti softwarového prostředí je zde riziko nekonzistence záznamů Datového katalogu a nedostupnosti softwarových nástrojů pro zajištění kvality katalogizačních záznamů u jednotlivých katalogizačních nástrojů (pro různé katalogizační nástroje mohou být k dispozici rozdílné softwarové nástroje pro zajištění kvality obsahu datového katalogu).

Tabulka 19: Zhodnocení varianty E

	Bezpečnost	Finanční náročnost vybudování	Finanční náročnost provozu	Kvalita záznamů katalogu	Propojitelnost s jinými evidenčními systémy
Hodnocení	13	14	15	14	18

2.4.3.6 Varianta F

Tato varianta se podobá předcházející variantě. Opět je zde navržen vznik více oddělených datových katalogů podporovaných různými katalogizačními nástroji dle volby příslušných

orgánů veřejné správy (heterogenní prostředí). Pro snazší vyhledávání je i zde doporučen vznik centrálního datového katalogu plněného automatizovaně přenášením katalogizačních záznamů z jednotlivých dílčích datových katalogů. I v této variantě by k vkládání a úpravě záznamů datového katalogu bylo oprávněno více subjektů. Rozdílem oproti předcházející variantě je uplatnění pouze částečné redakce obsahu datového katalogu. Redakci by opět prováděl subjekt spravující datový některý z oddělených datových katalogů, ale pouze na části jeho obsahu.

Stejně jako v předcházejícím případě je výhodou této varianty možnost výběru katalogizačního nástroje podle potřeb jednotlivých orgánů veřejné správy. Další výhodou je zrychlení přístupu uživatelů katalogu ke katalogizačním záznamům, protože díky částečné redakci by byly založené záznamy přístupné uživatelům ihned, ale s označením, že se jedná o dosud neredigované záznamy.

Nevýhody opět spočívají v riziku nekonzistence katalogizačních záznamů mezi jednotlivými oddělenými datovými katalogy a nestejně dostupnosti softwarových nástrojů pro podporu zajištění kvality obsahu datového katalogu. Díky pouze částečné redakci je v této variantě vyšší riziko nekvalitních záznamů v Datovém katalogu.

Tabulka 20: Zhodnocení varianty F

	Bezpečnost	Finanční náročnost vybudování	Finanční náročnost provozu	Kvalita záznamů katalogu	Propojitelnost s jinými evidenčními systémy
Hodnocení	13	14	11	15	19

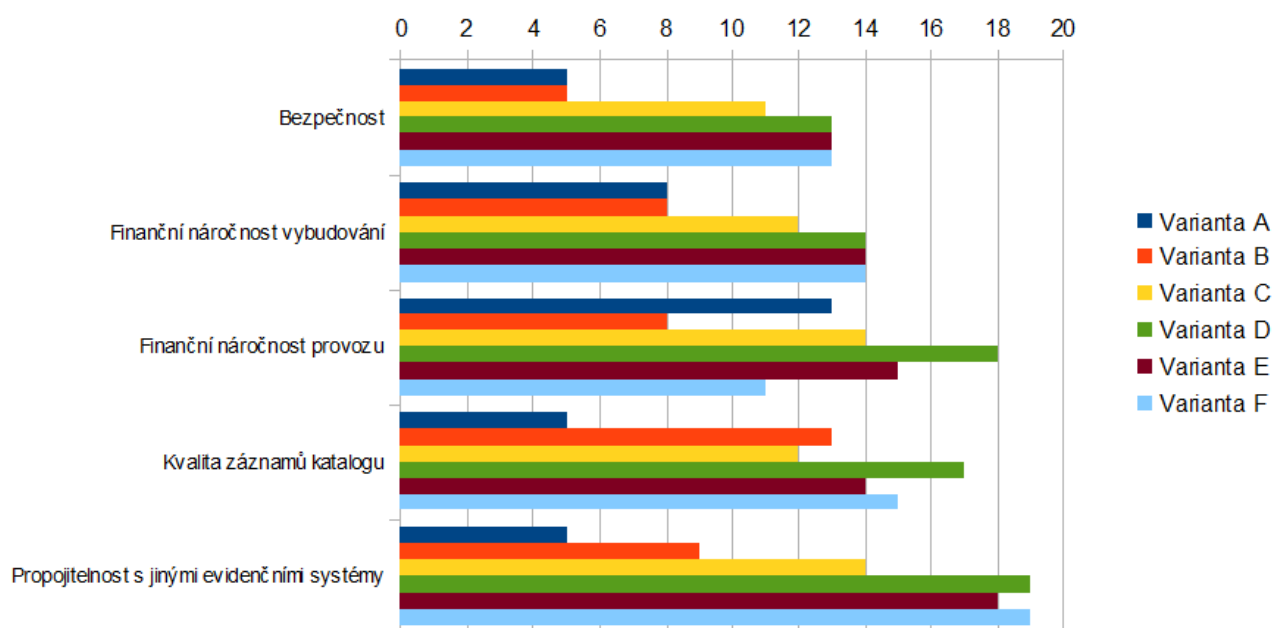
2.4.3.7 Výsledky srovnání jednotlivých variant

Následující tabulka uvádí souhrnné hodnocení jednotlivých výše představených variant.

Tabulka 21: Porovnání hodnocení jednotlivých variant

	Bezpečnost	Finanční náročnost vybudování	Finanční náročnost provozu	Kvalita záznamů katalogu	Propojitelnost s jinými evidenčními systémy
Varianta A	5	8	13	5	5
Varianta B	5	8	8	13	9
Varianta C	11	12	14	12	14
Varianta D	13	14	18	17	19
Varianta E	13	14	15	14	18
Varianta F	13	14	11	15	19

Následující graf pak přináší grafické porovnání hodnocení výše popsaných variant v rámci jednotlivých sledovaných hledisek.



Obrázek 3: Porovnání hodnocení jednotlivých variant

Z výše uvedená tabulky 21 a grafu na obrázku 3 vyplývá, že z hlediska zajištění bezpečnosti Datového katalogu byly nejlépe hodnoceny varianty A a B, tyto varianty mají také shodné hodnocení finanční náročnosti vybudování, což je dáno tím, že se tyto varianty liší pouze rozsahem redakce, která bude mít dopad spíše na financování provozu, než na finanční náročnost vybudování. Varianta B je nejlépe hodnocenou variantou z hlediska odhadované finanční náročnosti provozu. Protože finanční nároky zajištění úplné redakce centrálního datového katalogu jsou odhadovány vyšší než nároky provozu více datových katalogů, do kterých mohou přispívat pouze pověřené subjekty, a u kterého je prováděna pouze částečná redakce, je Varianta F hodnocena z hlediska finanční náročnosti provozu lépe než Varianta A.

Centrální katalog s úplnou redakcí má nejlepší předpoklady pro zajištění vysoké kvality záznamů datového katalogu, protože zde redaktori mohou nejsnáze zajistit dodržování jednotného způsobu vytváření katalogizačních záznamů a všechny záznamy musejí být před zveřejněním redigovány. Potenciál zajištění vysoké kvality katalogizačních záznamů je hodnocen o něco lépe v případě Varianty C, než v případě Varianty B. Varianta C je opět variantou s úplnou redakcí, i když v tomto případě se jedná o úplnou redakci na úrovni jednotlivých dílčích katalogů, neboť se zároveň jedná o variantu s oddělenými datovými katalogy. I když lze předpokládat určité rozdíly v přístupu jednotlivých redaktorů, očekáváme, že z celkového hlediska by kvalita katalogizačních záznamů měla být lepší, než v případě pouhé částečné redakce u Varianty B. Důvodem je fakt, že v případě Varianty B mohou být v Datovém katalogu obsaženy i málo kvalitní katalogizační záznamy, které neprošly redakcí. Nejhorší kvalita katalogizačních záznamů je očekávána v případě varianty D, kdy katalogizační záznamy může vytvářet kdokoli. I zde je aplikována úplná redakce, která ale bude patrně značně náročná vzhledem k tomu, že kvalita vytvářených záznamů bude pravděpodobně velmi různorodá. Redaktori se zde budou muset zaměřit nejen na dodržování metodických pokynů pro tvorbu katalogizačních záznamů, ale také na záznamy, které odkazují na nevyhovující obsah. Vzhledem k faktu, že katalogizační záznamy může v případě Varianty D vytvářet i široká veřejnost, lze záznamy s nevhodným obsahem očekávat. Ve výsledku by Datový katalog neměl obsahovat záznamy, které neprošly redakcí, ale díky náročnosti redakce může ve výsledku

Datový katalog obsahovat pouze menší množství záznamů nebo může docházet k významné časové prodlevě mezi odesláním záznamu do Datového katalogu a jeho zveřejněním.

Propojitelnost s jinými systémy je ovlivněna do značné míry tím, zda je prostředí datového katalogu homogenní či nikoli. Nejlépe jsou tak hodnoceny varianty A a B, které mají homogenní prostředí Datového katalogu. Navíc se jedná o varianty s centrálním datovým katalogem. Na příslušné evidenční systémy tak bude třeba vytvořit, na rozdíl od variant s oddělenými katalogy, pouze jedno rozhraní. Zajištění redakce datového katalogu z jednoho místa také umožní lepší dosažení stavu, kdy je obsah katalogizačních záznamů v souladu s požadavky na integraci s evidenčními systémy po obsahové stránce. Tato obsahová integrace může mít např. podobu konzistentního používání relevantních termínů v katalogizačních záznamech.

2.4.3.8 Doporučení pro budování Datového katalogu

Na základě expertního posouzení je k realizaci doporučena Varianta A. Varianta A je cílovou variantou, ve které lze díky centrálnímu datovému katalogu lépe zajistit soulad s požadavky v oblasti bezpečnosti, než je tomu v případě ostatních variant. Oproti ostatním variantám je zde také jednodušší zajištění vysoké kvality záznamů datového katalogu. Kvalita obsahu datového katalogu je považována za velmi důležitou, neboť nekvalitní katalogizační záznamy mohou odradit případné uživatele Datového katalogu a tím může dojít ke znehodnocení finančních prostředků vynaložených na jeho vybudování. Tato varianta se také jeví být v porovnání s ostatními variantami výhodnější z hlediska nákladů na vybudování Datového katalogu. Zejména v případě variant s více oddělenými datovými katalogy bude třeba na více místech zajistit pořízení katalogizačního nástroje, jeho úpravy a nasazení na jednotlivých vybraných orgánech veřejné správy. U jednoho centrálního datového katalogu by také mělo být jednodušší zajištění integrace s IS o ISVS a s ISDP. V případě variant s více oddělenými datovými katalogy by bylo třeba provést tuto integraci s každým z oddělených datových katalogů, což může v případě heterogenního softwarového prostředí vyžadovat vybudování několika různých rozhraní mezi výše uvedenými systémy a jednotlivými katalogizačními nástroji.

3 Koncepce katalogizace otevřených dat VS ČR

3.1 Úvod od Koncepce

Zatímco předcházející část dokumentu se zabývala popisem aktuálního stavu v oblasti publikování dat VS ČR a katalogizace dat veřejné správy ve světě a vymezením základních variant, které lze zvažovat při vymezení přístupu České republiky ke katalogizaci otevřených dat, tato část dokumentu obsahuje již samotnou Koncepti. Koncepte je navrhována s časovým výhledem realizace v délce 3 až 5 let, navazující část Plán realizace Koncepce potom rozpracovává jednotlivé kroky realizace a harmonogram realizace.

Koncepte katalogizace otevřených dat VS ČR představuje ucelenou sadu doporučení, která souhrnně odpovídají na otázku, kdo a jak by měl katalogizovat otevřená data VS ČR. Protože se ale za touto otázkou skrývá celá řada dílčích problémů a otázek, je Koncepte rozdělena do kapitol, které jsou vždy zaměřeny některou z nich. Jednotlivé otázky jsou uvedeny v tabulce 22. Tabulka zároveň u každé otázky uvádí i příslušné kapitoly, které se zabývají jejím řešením.

Tabulka 22: Otázky řešené v Koncepti a související kapitoly

Otázky řešené v koncepti	Místo řešení (kapitola)
Jaké jsou přínosy datového katalogu VS ČR?	kap. 3.2
Jaká data katalogizovat?	kap. 3.3
Kdo by měl data katalogizovat?	kap. 3.4
Jak by měl při katalogizaci postupovat?	kap. 3.5
Kdo bude spravovat a provozovat katalog dat VS ČR?	kap. 3.6
Jak bude zajištěna kvalita obsahu datového katalogu?	kap. 3.7
Jak bude zajištěna bezpečnost provozu datového katalogu?	kap. 3.8
Jaké jsou požadavky na nástroj pro datový katalog?	kap. 3.10
Kdo bude moci datový katalog využívat a za jakých podmínek?	kap. 3.11
Budou moci do datového katalogu přispívat občané ČR z řad veřejnosti?	kap. 3.11
Jak by měl být nástroj pro datový katalog pořízen?	kap. 3.10
Kdo a jak bude zajišťovat vzdělávání v oblasti katalogizace otevřených dat VS ČR?	kap. 3.12
Jaké budou potřebné či vhodné legislativní změny pro fungování datového katalogu?	kap. 3.13
Jak bude zajištěna vazba na připravovaný portál Publicdata.eu?	kap. 3.14

Otázky řešené v koncepci	Místo řešení (kapitola)
Jaká by měla být sw architektura nástroje pro datový katalog?	kap. 3. 9

Kromě hledání odpovědí na uvedené otázky je tato část dokumentu zakončena popisem scénářů užití katalogu otevřených dat VS ČR. Tyto scénáře by měly uceleně popsat zamýšlené použití Datového katalogu, které bude vycházet z doporučení k řešení jednotlivých otázek spojených s katalogizací otevřených dat VS ČR.

Scénáře užití i řešení jednotlivých otázek vycházejí ze situace, kdy je realizována doporučená varianta budování Datového katalogu uvedená v části Analýza současného stavu a varianty řešení. Zde bylo doporučeno realizovat Variantu A, jejíž atributy jsou uvedeny v tabulce 23.

Tabulka 23: Varianta budování Datového katalogu navržená k realizaci

Varianta/aspekt	Distribuvanost	Heterogenost	Vkládání	Redakce
A (cílová varianta)	Centrální katalog	Homogenní prostředí	Více pověřených subjektů	Úplná redakce

3.2 Přínosy katalogu otevřených dat VS ČR

Dříve, než bude pozornost věnována jednotlivým aspektům katalogizace otevřených dat veřejné správy České republiky, je třeba identifikovat přínosy, které tato aktivita může přinést. Řada přínosů identifikovaná v této kapitole má spíše celospolečenský dopad, který je obtížně měřitelný v peněžních jednotkách. Nicméně, již samotná identifikace těchto potenciálních přínosů by měla ilustrovat opodstatněnost katalogizace 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 subjekty²³ 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).

²³ Zde je použita terminologie citovaného Sdělení, informacemi shromažďovanými veřejnoprávními subjekty lze nicméně rozumět i data orgánů veřejné správy.

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.

3.2.1 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 zdoluhavé) 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.

3.2.2 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].

3.2.3 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 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).

Využití technologií propojitelných dat v rámci katalogu otevřených dat VS ČR má velký potenciál. Umožnily by vzájemné propojování souvisejících katalogizačních záznamů pomocí různých typů vazeb specifikujících např., že data katalogizovaná záznamem A doplňují/jsou ekvivalentní s/jsou podmnožinou/neshodují se s daty katalogizovanými jiným záznamem B. Tyto

souvislosti by značně usnadnily prohledávání katalogu. Uživatelé by měli lepší možnosti vyhledávání souvisejících, doplňujících či alternativních záznamů. Informační hodnota katalogu by tak značně narostla. Podobně by bylo možné pomocí technologií propojitelných dat vytvářet propojení mezi katalogizačními záznamy napříč různými katalogy, např. národními katalogy různých zemí.

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.

3.2.4 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.

3.3 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.

3.3.1 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 osobami²⁴. **Povinnost publikace a katalogizace otevřených dat bude ukládat jednotlivým orgánům veřejné správy příslušná legislativa** (viz kapitola 3.13).

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 jsou²⁵:

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.

²⁴ 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.

²⁵ Upraveno dle [54]

2. primární (původní) - data, která jsou zveřejněna původcem dat v podobě, v jaké byla původcem jako primární (původní) vytvořena²⁶. Za primární data se považují i
 - a. referenční údaje ze základních registrů,
 - b. data z registrů²⁷ a rejstříků VS,
 - c. agregovaná data (např. výsledky voleb) pokud není možné zveřejnit data, z nichž byla provedena agregace,
 - 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. zpřístupněná 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 uživatelům; poskytovatel dat může jednorázově vyžádat i úhradu za mimořádně náročné pořízení dat, pokud si uživatel zpřístupnění těchto dat vyžádá²⁸.

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

²⁶ Za původce dat je považován orgán veřejné správy, který údaje vytvořil.

²⁷ Pokud taková datová data ukládá publikovat jako veřejná data nějaká právní norma ČR.

²⁸ Upraveno podle § 17 odst. 1 zák. č. 106/1999 Sb. a podle návrhu úpravy směrnice 2003/98/ES [18].

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.

3.3.2 Příklady aplikace definice otevřených dat na existující data

Jak již bylo uvedeno v části Analýza současného stavu a varianty řešení, orgány veřejné správy již dnes publikují velké množství dat. I když je formální definice otevřených dat veřejné správy formulována až v rámci této Koncepce, existují již dnes data, která by bylo možné za otevřená data označit, nebo se splnění podmínek otevřených dat blíží. Příkladem takovýchto dat jsou data ze Základního registru územní identifikace, adres a nemovitostí zveřejňovaná podle § 47 a § 62 zák. č. 111/2009 Sb. (uvedené paragrafy se vztahují pouze k RÚIAN, obecně referenční údaje ze základních registrů za otevřená data považovat nelze). Vyhodnocení splnění podmínek otevřených dat těmito daty je uvedeno v tabulce 24. Pro srovnání jsou v této tabulce zhodnocena také data o volbách publikovaná Českým statistickým úřadem²⁹.

Tabulka 24: Splnění podmínek otevřených dat u vybraných existujících dat

Podmínka otevřených dat	Data z RÚIAN	Data o volbách
1. Úplnost	Splněno	Splněno
2. Primárnost	Splněno - data ze základních registrů jsou považována za primární	Splněno
3. Včasnost zveřejnění	Splněno	Splněno
4. Snadná dostupnost	Splněno	Splněno
5. Strojová čitelnost	Splněno - formát VFR	Splněno - formáty DBF, XML, MS Excel
6. Přístup bez omezení	Splněno	Splněno částečně - k některým datům a formátům je možné přistupovat pouze po registraci v průběhu volebních dnů
7. Otevřené standardy	Splněno - specifikace používaného formátu je bezplatně dostupná	Splněno - formáty DBF a MS Excel jsou rozšířené a zpracovatelné v řadě i bezplatných nástrojů
8. Jasné podmínky užití	Splněno - podmínky využití dat jsou definovány v zák. č. 111/2009 Sb. a také na webových stránkách ČÚZK	Nesplněno - podmínky užití nejsou uvedeny přímo na webových stránkách

²⁹ <http://www.volby.cz>

Podmínka otevřených dat	Data z RÚIAN	Data o volbách
9. Stálá dostupnost	Splněno	Splněno
10. Dostupnost s minimem nákladů	Splněno - data lze využívat bezplatně ³⁰	Splněno - data lze využívat bezplatně

Jak je z tabulky 24 patrné, veřejně zpřístupněná data z RÚIAN splňují všechny podmínky otevřených dat. Data o volbách Český statistický úřad (ČSÚ) publikuje v podobě HTML stránek na portálu www.volby.cz. Strojově čitelná data jsou online k dispozici ve formátech DBF a XML v průběhu voleb po registraci (služba je orientovaná zejména na subjekty z řad médií pro potřeby volebního zpravodajství). Sestavy a tabulky ve formátech MS Excel jsou s určitým časovým odstupem k dispozici v rámci publikací k minulým volbám. Tabulky s daty o volbách jsou k dispozici také v rámci veřejně přístupné databáze ČSÚ, kde je možné tyto tabulky exportovat do formátů MS Excel a XML.

3.4 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) 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.

V této koncepci navržený způsob katalogizace dat nebude³¹ 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 uživatel³².

3.4.1 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,

³⁰ Bezplatný přístup prostřednictvím internetu, data na datových nosičích jsou k dispozici za úplaty

³¹ Automatizovaně budou probíhat specializované procedury, které budou přispívat ke zvýšení kvality záznamů v katalogu dat (např. kontroly existence odkazů na data, příp. značkování dat).

³² Také označovaný jako průzkumník - např. v ISDP.

- 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.

3.4.2 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.

3.4.3 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,
- určování podmínek, za jakých jsou data zveřejněna ve formě otevřených dat, a to zejména právních (licence)³³, ekonomických, časových a technických,
- pověření kurátora správou dat.

3.4.4 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.

³³ V mezích platné legislativy.

3.4.5 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:
 - o vyplnění povinných atributů,
 - o existenci odkazu na zdroj dat,
 - o shodu popisu a obsahu dat,
 - o zařazení datové množiny do klasifikačních struktur³⁴ (např. EUROVOC),
 - o 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.

3.4.6 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.

3.5 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.

3.5.1 Postup katalogizace

Vložení záznamu do katalogu dat:

- Určení dat ke katalogizaci³⁵
- Založení katalogizačního záznamu³⁶ - 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

- Ověření korektnosti a klasifikace záznamu:
 - o vyplnění povinných atributů
 - o existenci odkazu na zdroj dat
 - o shodu popisu a obsahu dat

³⁴ Klasifikační struktury budou součástí návrhu Metodiky katalogizace otevřených dat VS ČR.

³⁵ Lze předpokládat, že v budoucnu budou k dispozici poloautomatizované nástroje, které umožní vyhledávat otevřená data a upozornit na jejich existenci.

³⁶ V Datovém katalogu budou uloženy pouze katalogizační záznamy a nikoli vlastní katalogizovaná data.

- o zařazení datové množiny do klasifikačních struktur (EUROVOC)
- o 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

3.5.2 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.

3.5.3 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 dále 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*) - jazyk³⁷, ve kterém je uveden popis záznamu a jeho další atributy
- **K: Platnost záznamu** (*P*) - obecné označení, zda je katalogizační záznam platný či nikoli³⁸
- **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

³⁷ Katalogizační nástroj musí připouštět možnost vícejazyčného popisu katalogizačního záznamu.

Výchozí hodnotou bude „CZ“.

³⁸ Výchozí hodnota bude nastavena na „Platný“. Ručně nutno změnit na „Neplatný“.

- **K³⁹: Poskytovatel dat (P)** - identifikace poskytovatele dat, skládá se z následujících atributů
 - o **K: IČO (P)** - identifikační číslo poskytovatele dat
 - o **Název poskytovatele dat (P)** - platný úplný název poskytovatele dat⁴⁰
 - o **Web poskytovatele dat (N)** - odkaz na domovskou stránku webu poskytovatele dat⁴¹
 - o **Kontakt (P)** - kontaktní email poskytovatele, na kterém poskytovatel poskytuje informace široké veřejnosti⁴²
- **K: Podmínky užití dat (P)** - informace o právních podmínkách využití dat uvedených v katalogizačním záznamu, skládá se z následujících atributů
 - o **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
 - o **K: Podmínky užití / Licence (N)** - označení podmínek / licence
 - o **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
 - o **K: Název ISVS (N)** - úplný název příslušného ISVS
 - o **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
 - o **K: Název datového prvku (N)** - úplný název datového prvku
 - o **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
 - o **K: Název záznamu (N)** - název odkazovaného katalogizačního záznamu
 - o **K: Typ vazby (N)** - označení typu vazby mezi záznamy (např. „doplňuje“, „nahrazuje“, atd. Konkrétní typy vazeb budou definovány v rámci metodiky)
 - o **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
 - o **K: Koncepty klasifikačních struktur (P)** - výčet konceptů klasifikačních struktur používaných pro klasifikaci dat uvedených v katalogizačním záznamu

³⁹ Až bude v plném provozu JIP / KAAS (Jednotný identitní prostor / Katalog autorizačních a autentizačních služeb) bude automatizovaně vkládáno.

⁴⁰ Bude automaticky doplněn z registru osob (ROS) na základě zadaného IČO poskytovatele dat (<http://www.szrcr.cz/registr-osob>).

⁴¹ Bude automaticky doplněn z Portálu veřejné správy (<http://portal.gov.cz>).

⁴² Bude automaticky doplněn z Portálu veřejné správy (<http://portal.gov.cz>).

- **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
 - **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, webové 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/OWL⁴³ v případě formátu RDF). V případě, že je schéma zapsáno ve více souborech, je uvedeno URL pro každý jednotlivý soubor.

⁴³ OWL - Ontology Web Language [59]

- o **MIME type (P)** - standardizované označení formátu dat⁴⁴
- o **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)
- o **K: Velikost souboru (N)** - velikost souboru v bytech v případě, že datovým zdrojem je datový soubor
- o **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 automaticky⁴⁵.

3.5.4 Doporučení ohledně klasifikačních struktur (slovníků a taxonomií)

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

- EUROVOC⁴⁶ – víceoborová taxonomie, existuje ve všech úředních jazycích EU, může být využita pro označení oblastí/oborů činnosti, kterých se data týkají,
- CZ-NACE⁴⁷ – primárně popisuje ekonomické činnosti a může být např. využita pro označení dat příslušnými ekonomickými činnostmi, kterých se týkají.

3.6 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.

⁴⁴ 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.

⁴⁵ V době implementace rozšířené funkcionality Datového katalogu budou využity služby poskytované geoportálem INSPIRE nebo základním registrem RÚIAN.

⁴⁶ <http://eurovoc.europa.eu/drupal/?q=cs>

⁴⁷ <http://apl.czso.cz/iSMS/en/klasstru.jsp?kodcis=80004&cisjaz=203>

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.

3.7 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.

3.7.1 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 25. Jak je uvedeno v tabulce, těchto kvalitativních atributů je možné pomocí současných technických dosáhnout jen manuální kontrolou záznamů.

Tabulka 25: 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.
Úplnost katalogu	V katalogu jsou katalogizovány všechny záznamy, jejichž přítomnost v katalogu vyplývá z kapitoly 3.3	Přímá manuální kontrola správcem katalogu, příp. redaktory.

Dále je nutné pro každý záznam v Datovém katalogu zajistit kvalitativní atributy uvedené v tabulce 26. 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 26: 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, 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.

3.7.2 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 uživatelé identifikované nedostatky dostaly k osobám odpovědným za jejich nápravu.

3.7.3 Využití metod strojového učení

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

3.8 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.

3.8.1 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 27: 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

ID	Hrozba
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.

3.8.2 Bezpečnostní požadavky a opatření pro jejich naplnění

Jednotlivé bezpečnostní požadavky, problémové oblasti, které řeší a příklady možných způsobů naplnění těchto požadavků jsou uvedeny v následující tabulce 28. 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 28: Bezpečnostní požadavky na Datový katalog

ID	Požadavek	Související hrozby	Zajištění	Poznámky
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	Jednoznačná identifikace a autentizace uživatele není vyžadována v případě koncových uživatelů (viz kap. 3.4.5)

ID	Požadavek	Související hrozby	Zajištění	Poznámky
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	
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	Neplatí pro plánované odstávky systému, které mohou být podle potřeby delší

ID	Požadavek	Související hrozby	Zajištění	Poznámky
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	Pořizování 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 odesíláním 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	Např. využití technologie CAPTHA48 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	Nepůsobí přímo proti vzniku bezpečnostních incidentů, ale pomáhá je řešit a pomáhá identifikovat jejich příčiny

3.8.3 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

⁴⁸ CAPTHA - 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]

řešení bezpečnostních incidentů. V rámci školení a vzdělávání uživatelů Datového katalogu by měly být uživatelé seznámeni i s obsahem této bezpečnostní politiky.

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 DoS⁴⁹ a DDoS⁵⁰.

3.8.4 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 katalogu⁵¹. 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. 3.5.3). Kurátor

⁴⁹ DoS - Denial of Service - útok s cílem způsobit nedostupnost cíle útoku (např. webové stránky nebo služby) pro jeho legitimní uživatele [61]

⁵⁰ DDoS - Forma útoku DoS, ve kterém se více systémů snaží zaplavit svými požadavky cíl útoku a způsobit tak jeho nedostupnost pro legitimní uživatele [61]

⁵¹ 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.

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 dat⁵², 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.

3.9 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:

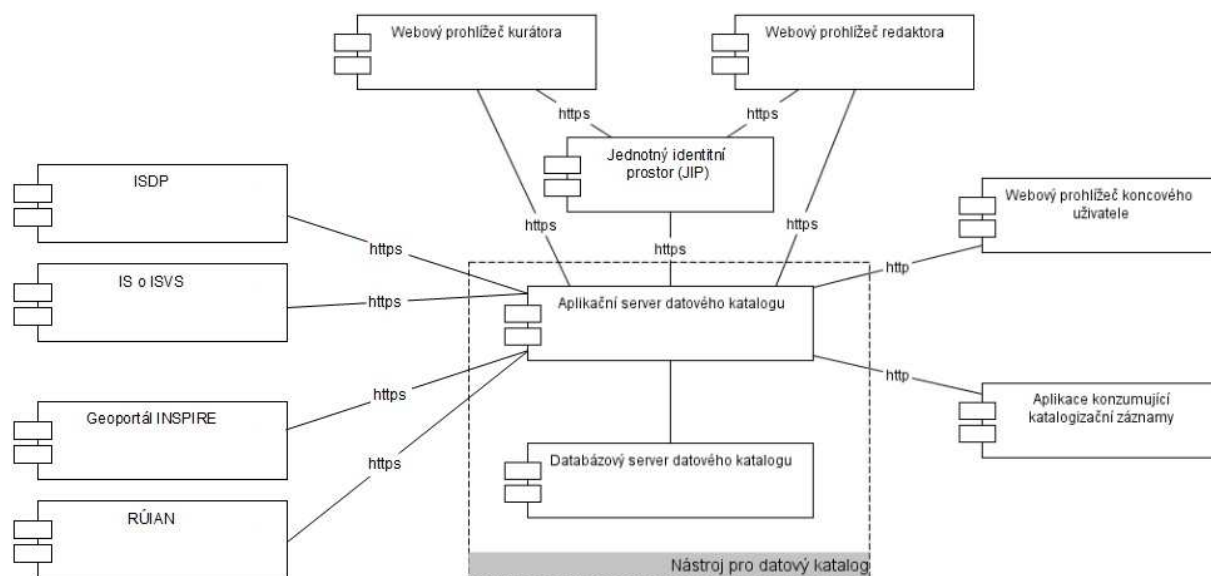
- databázový server datového katalogu - umožňuje ukládání, modifikaci a dotazování nad zadanými katalogizačními záznamy
- aplikační server datového katalogu - přijímá požadavky od uživatelů prostřednictvím protokolů HTTP či HTTPS, odesílá uživatelům výsledek požadavku v podobě HTML stránky a zajišťuje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

Znázornění architektury na obrázku 4 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 (viz kap. 2.1.1.2). 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 geookordináty pro označení území, ke kterému se vztahují data popsaná v katalogizačních záznamech.

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

⁵² 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.



Obrázek 4: Znárodnění základní softwarové architektury nástroje pro katalogizaci dat

3.10 Požadavky na nástroj pro datový katalog a způsob pořízení nástroje

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. 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 váhy⁵³ jednotlivých skupin 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

⁵³ Pokud se ve skupině výběrových kritérií objevují obligatorní kritéria, váha se týká pouze fakultativních kritérií.

- b. Vkládání, editace a zneplatňování záznamů, které mají strukturu vymezenou v kapitole 3.5.3
 - 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 3.8.2
5. Výkonnost a škálovatelnost z hlediska objemu zpracovávaných záznamů (20 %)

3.11 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 3.4. 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.

3.11.1 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.

3.12 Vzdelávání 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.

Vzdelávání 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.

3.13 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 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 zabývá podrobněji kapitola 3.3, 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.

3.13.1 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 celku⁵⁴, 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 vyplývajících 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“).

⁵⁴ 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. 3.3)

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.

3.13.2 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 3.3. 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.

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 odůvodnění⁵⁵. 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.

3.13.3 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 formu 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

⁵⁵ Do českého překladu pracovní verze novely Směrnice [18] byl pojem otevřená data (open data) přeložen jako zpřístupnění údajů.

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.

3.14 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.

3.14.1 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.

3.14.2 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. 3.5.4.

3.15 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 3.4 popisující jednotlivé role) využívat Datový katalog.

V následujících sekcích jsou popsány tyto 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. Vypovídání zpětné vazby a podnětů od koncových uživatelů

3.15.1 Pověření redaktora a přidělení příslušné role v Datovém katalogu**Tabulka 29: Pověření redaktora a přidělení příslušné role v Datovém katalogu**

Činnost	Role	Komentář
Pověření osoby výkonem role redaktor	Správce Datového katalogu	Akce proběhne postupem běžným pro subjekt správce Datového katalogu.
Proškolení pověřené osoby	Správce Datového katalogu, redaktor	Správce Datového katalogu zajistí proškolení pověřeného redaktora(-ů)
Přiřazení role redaktor pověřené osobě	Správce Datového katalogu	

3.15.2 Zrušení pověření redaktora**Tabulka 30: Zrušení pověření redaktora**

Činnost	Role	Komentář
Zrušení pověření osoby k výkonu role redaktora	Správce Datového katalogu	Akce proběhne postupem běžným pro subjekt správce Datového katalogu.
Odebrání role redaktor osobě s odebraným pověřením	Správce Datového katalogu	

3.15.3 Založení záznamu v Datovém katalogu**Tabulka 31: Založení záznamu v Datovém katalogu**

Činnost	Role	Komentář
Přihlášení do Datového katalogu	Kurátor dat	Kurátor dat využívá pro přístup identifikačních údajů pro provedení jednoznačné identifikace a autentizace uživatele. Pozn.: Bude využíván Jednotný identitní prostor, případně kvalifikované zaměstnanecké certifikáty
Založení katalogizačního záznamu	Kurátor dat	Vyplnění všech povinných atributů katalogizačního záznamu, které musí ručně kurátor vyplnit (viz kap. 3.5.3)

Činnost	Role	Komentář
Zařazení dat do klasifikačních struktur	Kurátor dat	Přiřazení odpovídajících konceptů ze slovníku EUROVOC, zařídění k odpovídajícímu orgánu veřejné správy, přiřazení příslušných ekonomických činností z klasifikace CZ-NACE, pokud o nich data vypovídají, přiřazení konceptů z dalších povinně určených taxonomií
Provázání se záznamy v systémech ISDP a IS o ISVS	Kurátor dat	Přiřazení příslušných identifikátorů datových prvků v systému ISDP v případě, že jsou v katalogizovaných datech použity. Přiřazení identifikátorů příslušných ISVS v IS o ISVS v případě, že z něj data pocházejí.
Odeslání vytvořeného katalogizačního záznamu k redakci	Kurátor dat	
Přihlášení do Datového katalogu	Redaktor	Redaktor využívá pro přístup přidělených identifikačních údajů pro provedení jednoznačné identifikace a autentizace uživatele.
Ověření korektnosti obsahu katalogizačního záznamu a správnosti jeho klasifikace	Redaktor	Zahrnuje ověření: vyplnění povinných atributů existence odkazu na zdroj dat shody obsahu dat s uvedeným popisem zařazení dat do klasifikačních struktur provázání se systémy ISDP a IS o ISVS. V případě, že záznam vykazuje nedostatky, je předán poskytovateli dat k opravě. Redaktor předá zprávu s označením nedostatků.
Zveřejnění katalogizačního záznamu v Datovém katalogu	Redaktor	Zpřístupnění katalogizačního záznamu koncovým uživatelům

3.15.4 Úprava záznamu v Datovém katalogu

Tabulka 32: Úprava záznamu v Datovém katalogu

Činnost	Role	Komentář
Přihlášení do Datového katalogu	Kurátor dat	
Vyhledání katalogizačního záznamu k úpravě	Kurátor dat	

Činnost	Role	Komentář
Provedení změn v obsahu katalogizačního záznamu	Kurátor dat	Provedení změn v jakémkoli atributu katalogizačního záznamu nebo i ve více attributech. Změna zařazení do klasifikačních struktur, změna provázání se systémy ISDP a IS o ISVS. Provedením změny se katalogizační záznam vrací zpět do stavu „redakce neprovedena“.
Odeslání vytvořeného katalogizačního k redakci	Kurátor dat	
Přihlášení do Datového katalogu	Redaktor	
Ověření korektnosti obsahu katalogizačního záznamu a správnosti jeho klasifikace	Redaktor	Zahrnuje ověření: vyplnění povinných atributů existence odkazu na zdroj dat shody obsahu dat s uvedeným popisem zařazení dat do klasifikačních struktur provázání se systémy ISDP a IS o ISVS. V případě, že záznam vykazuje nedostatky, je předán zpět poskytovateli dat k opravě. Redaktor předá zprávu s označením nedostatků.
Zveřejnění katalogizačního záznamu v Datovém katalogu	Redaktor	Zpřístupnění katalogizačního záznamu koncovým uživatelům

3.15.5 Zneplatnění záznamu v Datovém katalogu

Tabulka 33: Zneplatnění záznamu v Datovém katalogu

Činnost	Role	Komentář
Přihlášení do Datového katalogu	Kurátor dat	
Vyhledání katalogizačního záznamu, který má být zneplatněn	Kurátor dat	
Provedení zneplatnění katalogizačního záznamu	Kurátor dat	Provedení označení datového záznamu, jako neplatného. Kurátor uvede důvod zneplatnění záznamu. Zneplatnění katalogizačního záznamu je postoupeno redaktorovi k potvrzení.
Přihlášení do Datového katalogu	Redaktor	

Činnost	Role	Komentář
Potvrzení zneplatnění katalogizačního záznamu	Redaktor	Na základě uvedeného důvodu zneplatnění záznamu posoudí redaktor jeho oprávněnost. V případě oprávněného zneplatnění jej potvrdí a příslušný katalogizační záznam je označen jako neplatný. V případě, že zdůvodnění neplatnosti katalogizačního záznamu není dostatečné, redaktor zneplatnění nepotvrdí a vrátí zneplatnění k doplnění příslušnému poskytovateli dat.

3.15.6 Obnovení platnosti záznamu v Datovém katalogu

Tabulka 34: Obnovení platnosti záznamu v Datovém katalogu

Činnost	Role	Komentář
Přihlášení do Datového katalogu	Kurátor dat	
Vyhledání neplatného katalogizačního záznamu, u kterého má být obnovena platnost	Kurátor dat	
Provedení obnovení platnosti katalogizačního záznamu	Kurátor dat	Provedení označení datového záznamu, jako platného. Kurátor uvede důvod obnovení platnosti záznamu. Obnovení platnosti katalogizačního záznamu je postoupeno redaktorovi k potvrzení. Pokud je třeba po obnovení platnosti provést změny v obsahu katalogizačního záznamu, postupuje se podle scénáře „Úprava záznamu v Datovém katalogu“.
Přihlášení do Datového katalogu	Redaktor	
Potvrzení obnovení platnosti katalogizačního záznamu	Redaktor	Na základě uvedeného důvodu obnovení platnosti záznamu posoudí redaktor jeho oprávněnost. V případě oprávněného obnovení platnosti jej potvrdí a příslušný katalogizační záznam je označen jako platný. V případě, že zdůvodnění obnovení platnosti katalogizačního záznamu není dostatečné, redaktor jej nepotvrdí a vrátí obnovení platnosti k doplnění příslušnému poskytovateli dat.

3.15.7 Procházení Datového katalogu a poskytnutí zpětné vazby

Tabulka 35: Procházení Datového katalogu a poskytnutí zpětné vazby

Činnost	Role	Komentář
Vyhledání katalogizačního záznamu	Koncový uživatel	Vyhledání záznamu pomocí vyhledávacích funkcí, procházením skupin katalogizačních záznamů nebo procházením seznamu záznamů označených dle konceptů klasifikačních struktur dle zájmu koncového uživatele
Stažení dat	Koncový uživatel	Pokud katalogizační záznam obsahuje odkaz na stažitelný soubor s daty, může je uživatel stáhnout využitím uvedeného odkazu
Prohlížení obsahu dat	Koncový uživatel	Pokud to pro daný typ dat Datový katalog umožňuje, může uživatel využít funkcí pro zobrazení obsahu dat, např. v podobě tabulky nebo grafu, případně vizualizace dat se vztahem ke geografickému území na mapovém podkladě
Odeslání zpětné vazby	Koncový uživatel	Odeslání zpětné vazby týkající se katalogizačního záznamu, např. upozornění na nesprávně uvedené údaje, na neplatné odkazy na data nebo na nepřesné, zavádějící či nesrozumitelné popisy dat. Odeslání obecné zpětné vazby týkající se Datového katalogu, tj. komentáře, který se neváže specificky k některému z katalogizačních záznamů. Odeslání upozornění na existující data, která dosud nebyla katalogizována včetně označení místa, kde jsou data zveřejněna a orgánu veřejné správy, který data zveřejnil.

3.15.8 Vypořádání zpětné vazby a podnětů od koncových uživatelů

Tabulka 36: Vypořádání zpětné vazby a podnětů od koncových uživatelů

Činnost	Role	Komentář
Přihlášení do Datového katalogu	Správce Datového katalogu	Správce Datového katalogu využívá pro přístup přidělených identifikačních údajů pro provedení jednoznačné identifikace a autentizace uživatele.
Posouzení oprávněnosti podnětu	Správce Datového katalogu	Zjevně neopodstatněnými podněty, vágními podněty, neúplnými podněty a dalšími podněty bez řádných náležitostí se správce Datového katalogu nebude zabývat. Podněty shledané oprávněnými a opodstatněnými správce Datového katalogu vypořádá.

Činnost	Role	Komentář
Předání oznámení o existujících nekatalogizovaných datech označeným poskytovatelům dat	Správce Datového katalogu	Oznámení o existujících nekatalogizovaných datech předá správce Datového katalogu orgánům veřejné správy označeným v oznámení za poskytovatele dat ke zpracování.
Předání oznámení o nedostacích v kvalitě katalogizačních záznamů příslušným poskytovatelům dat	Správce Datového katalogu	Oznámení o nedostacích v kvalitě katalogizačních záznamů předá správce Datového katalogu příslušným poskytovatelům dat, kteří za jejich katalogizaci zodpovídají.
Vyhodnocení obecné zpětné vazby	Správce Datového katalogu	Obecná zpětná vazba poskytnutá koncovými uživateli Datového katalogu může mít různý obsah. Správce Datového katalogu tuto obecnou zpětnou vazbu adekvátním způsobem zpracuje. Poskytnutá zpětná vazba může být pro správce Datového katalogu podnětem k vytvoření změnového požadavku na Datový katalog, jehož vypořádání může zahrnovat i součinnost provozovatele Datového katalogu.
Vyrozumění koncového uživatele, který poskytl zpětnou vazbu	Správce Datového katalogu	Pokud koncový uživatel při poskytnutí zpětné vazby uvedl kontaktní údaje, vyrozumí správce Datového katalogu koncového uživatele o tom, jak byl jeho podnět vypořádán. Správce datového katalogu vyrozumí koncového uživatele, který uvedl kontaktní údaje i v případě, že neshledal jeho podnět oprávněným či opodstatněným.

4 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.

4.1 Etapy a hlavní skupiny činností

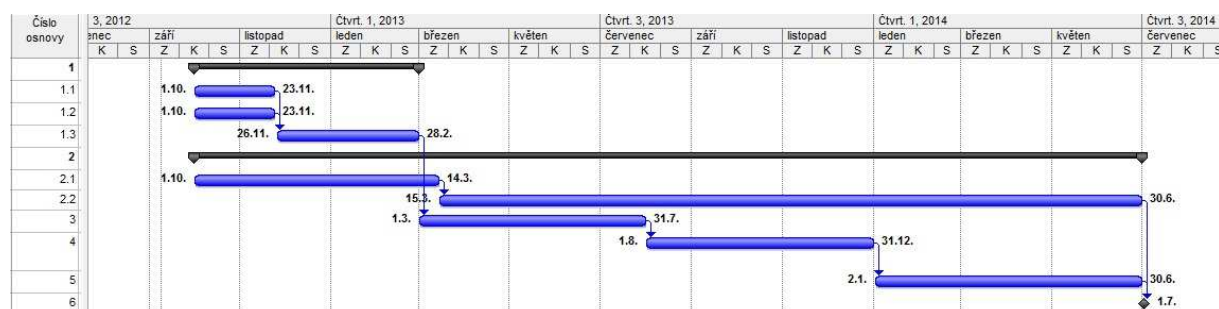
V této kapitole jsou vymezeny hlavní skupiny činností, které je třeba provést, aby byla realizována navržená Koncepce. Etapy mohou být realizovány formou samostatných projektů.

Tabulka 37: 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 data⁵⁶	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 ⁵⁷	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

⁵⁶ Předpokládá se, že příprava legislativních podmínek bude realizována interně pracovníky příslušných orgánů veřejné správy. Není tak uváděna pracnost a náklady souvisejících skupin činností.

⁵⁷ Účinnost 1.7.2014



Obrázek 5: Harmonogram etap a hlavních skupin činností

4.2 Popis etap realizace koncepce

V následujícím textu je uveden popis jednotlivých etap realizace Koncepce.

4.2.1 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.

4.2.1.1 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 (viz kap. 2.3.3.1), 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.

4.2.1.2 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 redaktori 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.

4.2.1.3 *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.

4.2.2 **Příprava legislativních podmínek pro publikaci otevřených dat**

Tato etapa je zaměřena na vytvoření legislativy v souladu s doporučeními uvedenými v kapitole 3.13. Má-li být katalogizace otevřených dat povinná pro všechny orgány veřejné správy, bude třeba, aby tuto povinnost uložil orgánům veřejné správy zákon. V kapitole 3.13 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.⁵⁸, 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.

⁵⁸ 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.

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.

4.2.3 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 3.9. 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.

4.2.4 Implementace vybraného nástroje

Obsahem etapy je:

- implementace vybraného řešení,
- vytvoření podmínek na MV ČR,
- příp. úprava Metodiky katalogizace otevřených dat dle plnohodnotného nastavení Datového katalogu,
- školení redaktorů.

Etapa se zaměřuje na implementaci vybraného katalogizačního nástroje, tj. dovývoj, 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.

4.2.5 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.

4.2.6 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.

4.3 Seznam výstupů

Tabulka 38: 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
V3	Katalogizační záznamy pro data z následujících datových zdrojů (pokud dojde k jejich otevření, viz [2]): • 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

ID	Výstup	ID etapy
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

4.4 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),

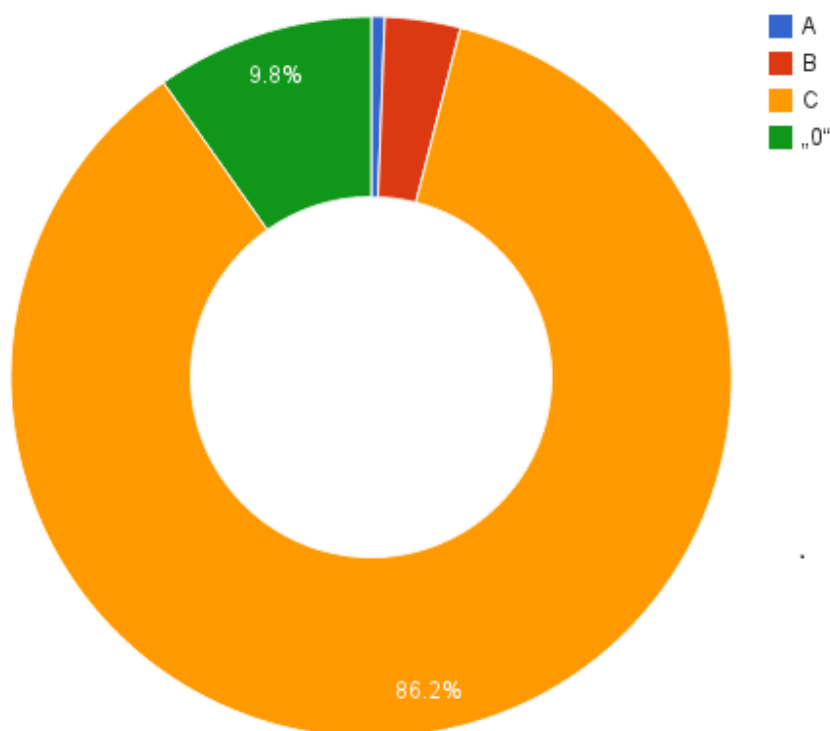
- třída C - subjekty publikující relativně malý počet dat a mající nejmenší počet katalogizačních záznamů (řádově jednotky záznamů),
- třída 0 - zvláštní třída, do které byly zařazeny ty subjekty, které nemají vlastní webové stránky, na kterých by publikovaly svá data (viz níže).

Při analýze orgánů veřejné správy bylo zjištěno, že ne všechny orgány mají vlastní webové stránky, na kterých by mohly publikovat svá data. Pro tyto orgány byla vytvořena výše uvedená zvláštní třída 0. I když to neplatí pro všechny orgány zařazené do třídy 0, tyto orgány zpravidla publikují data na portálu nadřazeného orgánu. Nicméně u dat na tomto portálu často chybí údaj o tom, který z konkrétních orgánů je poskytovatelem dat a podřazené orgány tak bylo obtížné zařadit do některé ze skupin A – C. Pro další výpočty odhadovaného počtu záznamů v Datovém katalogu bylo uvažováno, že orgány třídy 0 nebudou přispívat do celkového počtu žádnými záznamy, nicméně nadřazený orgán byl zařazen do třídy A nebo B podle očekávaného celkového počtu záznamů poskytovaného tímto orgánem a orgány jemu podřízenými. Ve skutečnosti ale i orgány ve třídě 0 mohou mít záznamy v Datovém katalogu, protože jimi pověření kurátoři dat mohou daný podřazený orgán přímo určit jako poskytovatele určitých dat při vytváření katalogizačního záznamu (dojde tak k doplnění informace, která dnes u publikovaných dat na souhrnných portálech není k dispozici).

Jako seznam analyzovaných subjektů byl využit seznam orgánů územní samosprávy, orgánů státní správy a ostatních orgánů veřejné moci, který je uveden v rámci seznamu držitelů datových schránek na Portálu veřejné správy (<http://seznam.gov.cz/ovm/welcome.do>). Podle § 6 zák. č. 300/2008 Sb. je datové schránka zřízena orgánu veřejné moci bezodkladně po jeho vzniku, případně na žádost orgánu. I když je pojem orgán veřejné moci používán ve výše uvedeném zákoně širší, než pojem orgán veřejné správy používán v tomto dokumentu a analyzovaný seznam držitelů datových schránek tak např. zahrnuje i soudy nebo složky Hasičského záchranného sboru, pro zjednodušení nebyl seznam upravován. Celkově tak bylo do tříd A – C a 0 zařazeno 7301 orgánů veřejné moci, rozdělení do tříd je uvedeno v následující tabulce 39.

Tabulka 39: 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



Obrázek 6: Orgány veřejné moci - podíl tříd podle očekávaného počtu katalogizačních záznamů

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 40 jsou uvedeny průměrné odhadované počty katalogizačních záznamů připadajících na jeden orgán veřejné správy příslušné třídy v každé z výše uvedených variantách.

Tabulka 40: Průměrné odhadované počty katalogizačních záznamů na jeden orgán veřejné správy

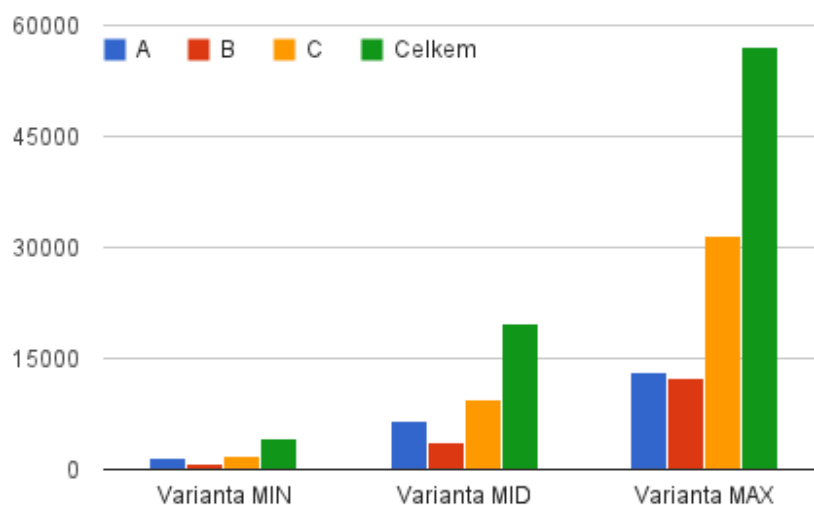
Třída	Varianta MIN	Varianta MID	Varianta MAX
A	33	150	300
B	3	15	55

Třída	Varianta MIN	Varianta MID	Varianta MAX
C	0,3	1,5	5
0	0	0	0

V následující tabulce 41 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 41: 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



Obrázek 7: Počty katalogizačních záznamů podle variant

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 Británie⁵⁹ obsahoval v červenci roku 2012 necelých devět tisíc záznamů.

Tabulka 42: Průměrný počet katalogizačních záznamů v Datovém katalogu za jeden rok, měsíc a pracovní den (zaokrouhleno 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 ⁶⁰	5	25	72

Průměrná doba, kterou bude potřebovat redaktor na redakci jednoho katalogizačního záznamu včetně redakce jeho případných oprav, je odhadována na 30 minut⁶¹. 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.

4.4.1 Pracnost pro redaktory obsahu Datového katalogu

Tabulka 43: 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 ⁶² , 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 43 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.

⁵⁹ <http://data.gov.uk/data>

⁶⁰ Počítáno s 22 pracovními dny v měsíci

⁶¹ V uváděném čase je zahrnuta i případná aktualizace záznamu v souvislosti se změnou periodicky se měnících datových zdrojů.

⁶² FTE - Full Time Equivalent

4.4.2 Pracnost pro kurátory dat

Tabulka 44: 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 45: 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 40) 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.

4.5 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). Text průvodního dopisu a zaslaný dotazník jsou uvedeny v příloze B. Žadostí byli osloveni výrobci katalogizačních nástrojů nebo dodavatelé informačních systémů pro veřejnou správu. V termínu pro doručení odpovědí byly obdrženy tři vyplněné dotazníky.

4.5.1 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 Koncepce 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.

4.5.2 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 porovnání cen srovnatelné úrovně podpory. Tým řešitelů Koncepce 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.

4.5.3 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 46 uvádí celkové roční náklady spojené s vytvářením a redakcí katalogizačních záznamů.

Tabulka 46: Průměrné celkové roční náklady spojené s vytvářením a redakcí katalogizačních záznamů (Kč, zaokrouhleno 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 47 uvedeny i průměrné odhadované náklady připadající na subjekty v jednotlivých třídách.

Tabulka 47: 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.

4.6 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.

4.7 Doporučení pro výběr dodavatele a provozovatele katalogizačního nástroje

Přestože odhadované ceny v odpovědích na žádost o poskytnutí informací (viz kap. 4.5) 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.

4.8 Přehled odhadovaných nákladů jednotlivých etap realizace Koncepce

V následující tabulce 48 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 48: 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 provoní 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)

5 Slovník základních pojmů

Pojem	Vysvětlení	Zdroj
Application Programming Interface (API)	Specifikované rozhraní pro komunikaci mezi softwarovými komponentami, může obsahovat např. specifikací procedur, datových struktur, tříd objektů nebo proměnných	[60]
Comma Separated Values (CSV)	Formát pro výměnu dat, kde každý záznam představuje jednu řádku souboru, a jednotlivé položky záznamu jsou odděleny čárkami; soubor může obsahovat záhlaví jako první řádek a jedná se o soubor prostého textu	[49]
Completely Automated Public Turing Test To Tell Computers and Humans Apart (CAPTCHA)	Program pro generování testů, ve kterých mohou lidé snadno uspět, ale počítače nikoli. Takovýto test může být založen např. na rozpoznávání obrázků nebo rozmazaného textu. Tyto testy slouží k např. rozlišování, zda se určitou webovou službu snaží použít člověk nebo jiný počítačový program a může být použit např. k obraně proti automatickému přidávání nerelevantních příspěvků (spamu) na diskusní fóra.	[22]
číselník	Seznam přípustných hodnot datového prvku obvykle ve formě dvojic, to znamená kódovaného údaje a hodnoty jeho kódu	Zák. č. 365/2000 Sb.
data	Jakékoli fyzicky zaznamenané znalosti (vědomosti), poznatky, zkušenosti nebo výsledky pozorování procesů, projevů, činností a prvků popisovaného světa.	
databáze (ve smyslu kombinace obsahu a struktury) - zvláštní právo pořizovatele databáze	Databází je pro účely tohoto zákona soubor nezávislých děl, údajů nebo jiných prvků, systematicky nebo metodicky uspořádaných a individuálně přístupných elektronickými nebo jinými prostředky, bez ohledu na formu jejich vyjádření.	Zák. č. 121/2000 Sb. (§88)
databáze (ve smyslu struktury) - autorskoprávní ochrana	Databáze, která je způsobem výběru nebo uspořádáním obsahu autorovým vlastním duševním tvorem a jejíž součástí jsou systematicky nebo metodicky uspořádány a jednotlivě zpřístupněny elektronicky či jiným způsobem	Zák. č. 121/2000 Sb. (§2, odst. 2)
databáze jako úřední dílo	Databáze, která je součástí právního předpisu.	Zák. č. 121/2000 Sb. (§ 94, odst. 2 s odkazem na § 3 písm. a)
data veřejné správy	Data, jejichž poskytovatelem je orgán veřejné správy	
datový prvek	Jednotka dat, která je v daném kontextu dále považována za nedělitelnou a je jednoznačně definován.	Zák. č. 365/2000 Sb.
Denial of Service (DoS)	Útok, jehož cílem je způsobit nedostupnost počítače nebo síťového zdroje pro jejich legitimní uživatele.	[61]
Distributed Denial of Service (DDoS)	Forma útoku DoS, ve kterém se více systémů snaží zaplavit svými požadavky cíl útoku a způsobit tak jeho nedostupnost pro legitimní uživatele.	[61]

Pojem	Vysvětlení	Zdroj
dokument	Obsah na jakémkoli nosiči (psaný či tištěný na papíře či uložený v elektronické formě nebo jako zvuková, vizuální nebo audiovizuální nahrávka); jakákoli část takového obsahu	Směrnice 2003/98/ES
DWG	Formát nástrojů společnosti Autodesk, Inc. Využitelný i pro geografická data.	[4]
eXtensible Markup Language	Značkovací jazyk pro popis datových objektů označovaných jako XML dokumenty, značky jsou využity pro vyjádření logické struktury dat v XML dokumentu	[10]
Geography Markup Language (GML)	Značkovací jazyk založený na XML pro popis geografických informací. GML je vyvíjen organizací Open Geospatial Consortium a je také vydán jako ISO standard ISO 19136:2007.	[37]
Hypertext Markup Language (HTML)	Značkovací jazyk pro publikování obsahu na Webu (World Wide Web, WWW)	[46]
informační systémy veřejné správy (ISVS)	Informační systémy veřejné správy jsou souborem informačních systémů, které slouží pro výkon veřejné správy. Jsou jimi i informační systémy zajišťující činnosti podle zvláštních zákonů.	Zák. č. 365/2000 Sb.
JavaScript Object Notation (JSON)	Jednoduchý formát pro výměnu dat, který je založen na podmnožině jazyka JavaScript	[25]
Jednotný identitní prostor (JIP)	Zabezpečená adresářová služba obsahující údaje pro autentizaci a autorizaci uživatelů	[50]
JSON-with-padding (JSONP)	Technika pro volání funkcí ze vzdáleného serveru, která je založená na předávání dat ve formátu JSON jako parametrů volané funkce	[43]
KML	Značkovací jazyk založený na XML určený pro vizualizaci geografických informací. Je vhodný pro zobrazování dat na mapových podkladech. Původně vyvíjen společností Google, nyní vyvíjen v rámci Open Geospatial Consortium.	[38]
licence	viz licenční smlouva	
licenční smlouva	Licenční smlouvou autor poskytuje nabyvateli oprávnění k výkonu práva dílo užít (licenci) k jednotlivým způsobům nebo ke všem způsobům užití, v rozsahu omezeném nebo neomezeném, a nabyvatel se zavazuje, není-li podle § 49 odst. 2 písm. b) sjednáno jinak, poskytnout autorovi odměnu.	Zák. č. 121/2000 Sb.
MIME type	standardizované označení pro typ dat	[20]
možnost dálkového přístupu	Přístup k informaci neomezeného okruhu žadatelů pomocí sítě nebo služby elektronických komunikací	Zák. č. 106/1999 Sb.
možnost dálkového přístupu	Přístup do informačního systému prostřednictvím sítě nebo služby elektronických komunikací (například s využitím internetu)	Zák. č. 365/2000 Sb.
Open Data Protocol (OData)	Protokol, který umožňuje standardizovaným způsobem webovým klientům číst a upravovat data na webových serverech. Kombinuje jednodušší protokol AtomPub s formátem JSON. Využívá standardního prostředku pro popis struktury vyměřovaných dat a částečně i jejich sémantiky. Umožňuje také propojovat související data, avšak pouze v rámci jednoho datového zdroje.	[29]

Pojem	Vysvětlení	Zdroj
orgán veřejné správy (orgán VS)	Ministerstvo, jiný správní úřad nebo územní samosprávný celek	Zák. č. 365/2000 Sb. (podle §3)
otevřená data veřejné správy	Data veřejné správy jsou považována za otevřená, pokud splňují alespoň vlastnosti čísel 1, 4, 5, 7, 8 a 10 z následujících vlastností: - 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. - primární (původní) - data, která jsou zveřejněna původcem dat v podobě, v jaké byla původcem jako primární (původní) vytvořena. Za primární data se považují i - referenční údaje ze základních registrů, - data z registrů a rejstříků VS, - agregovaná data (např. výsledky voleb) pokud není možné zveřejnit data, z nichž byla provedena agregace, - agregovaná data - (např. statistiky nad jinými otevřenými daty) pokud je uveden způsob agregace a odkaz na zveřejněná primární data, z nichž byla agregace provedena. - 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 nepřiměřené dlouhé prodlevy od okamžiku vzniku dat, - snadno dostupná - data jsou dostupná a dohledatelná běžnými ICT nástroji a prostředky, - strojově čitelná - data ve formátu, který je strukturovaný takovým způsobem, že pomocí programové aplikace lze z dat získat žádané (vybrané) údaje - neomezuující přístup - data dostupná způsobem, který nediskriminuje jednotlivce nebo skupinu osob, - 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, - zpřístupněna za jasně definovaných podmínek užití dat (licence) s minimem omezení - podmínky musí být jasně a zřetelně definovány a zveřejněny, - stále dostupná - data jsou dostupná on-line po dobu uvedenou jejich poskytovatelem, - 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 vyžádá.	upraveno dle [54]
otevřená propojitelná data	Otevřená data publikovaná podle zásad propojitelných dat (viz propojitelná data)	
otevřený standard	Standard, jehož specifikace je volně dostupná	

Pojem	Vysvětlení	Zdroj
Portable Document Format (PDF)	Formát souboru pro reprezentaci obsahu dokumentů nezávislý na aplikačním softwaru, hardwaru a operačním systému; od roku 2008 otevřený standard ISO ISO 32000-1:2008	[62]
propojitelná data	Data splňující sadu zásad pro publikování a propojování souvisejících dat na webu otevřeným a na standardech založeným způsobem. Konkrétními zásadami jsou: - pojmenování objektů na webu pomocí URI - použití HTTP URI, které umožňují je vyhledat v prostředí dnešního webu - při vyhledání URI jsou uživateli poskytnuta data o objektu reprezentovaná v modelu RDF - objekty jsou provázány pomocí odkazů mezi HTTP URI, takže je možné objevovat související objekty	upraveno dle [6]
Representational State Transfer (REST)	Sada architektonických principů pro návrh webových služeb, kde jednotlivé funkce těchto služeb jsou přímo přístupné pomocí protokolu HTTP	[48]
Resource Description Framework (RDF)	Model pro reprezentaci a výměnu dat prostřednictvím webu. Má vlastnosti, které umožňují spojování různorodých dat s různými schématy. Jedná se o jednoduchý aplikačně i platformově nezávislý model reprezentující data ve formě grafu.	[27]
Rich Text Format (RTF)	Formát pro výměnu textu a grafiky použitelný na různých výstupních zařízeních a operačních prostředích a systémech. Formát je vytvářen společností Microsoft Corporation a jeho specifikace je volně dostupná	[31]
Shapefile (SHP)	Vektorový formát pro geografická data vyvíjený společností Environmental Systems Research Institute, Inc. (ESRI)	[16]
Software as a Service (SaaS)	Model dodávky ICT služeb, ve „ <i>kterém specializovaný subjekt (Application Service Provider, ASP) nabízí funkcionalitu aplikace, infrastrukturu potřebnou pro její provoz a další služby potřebné pro poskytování ICT služby zákazníkovi na bázi předplatného</i> “ [58]	[58]
Software jako služba	viz Software as a Service	
SPARQL	Dotazovací jazyk a protokol pro RDF data	[45], [11]
správce číselníku	Právní subjekt odpovědný za tvorbu a distribuci číselníku	Zák. č. 365/2000 Sb.
správce datového prvku	Právní subjekt, který nové datové prvky předkládá, navrhuje jejich změnu nebo zrušení	Zák. č. 365/2000 Sb.
strojově čitelná data	Data ve formátu, který je strukturovaný takovým způsobem, že pomocí programové aplikace lze z dat získat žádané (vybrané) údaje.	upraveno dle [18]
subjekt veřejného sektoru	„ <i>Subjektem veřejného sektoru</i> “ (se) rozumí státní, regionální nebo místní orgány, veřejnoprávní subjekty a sdružení vytvořená jedním nebo několika takovými orgány nebo jedním nebo několika takovými veřejnoprávními subjekty	Směrnice 2003/98/ES
subjekty veřejné správy	Orgány státní správy (ministerstva) a nestátní veřejnoprávní subjekty, které se člení na veřejnoprávní korporace (územní samospráva, profesní samospráva) a ostatní veřejnoprávní subjekty (např. fondy, nadace).	[24]

Pojem	Vysvětlení	Zdroj
Uniform Resource Identifier (URI)	Uniform Resource Identifier (URI) je posloupnost znaků identifikující abstraktní nebo fyzický zdroj	[7]
Uniform Resource Locator (URL)	Podmnožina identifikátorů URI, které kromě identifikace zdroje obsahují i určení i způsobu přístupu k tomuto zdroji	[7]
Úřední dílo	Právní předpis, rozhodnutí, opatření obecné povahy, veřejná listina, veřejně přístupný rejstřík a sbírka jeho listin, jakož i úřední návrh úředního díla a jiná přípravná úřední dokumentace, včetně úředního překladu takového díla, sněmovní a senátní publikace, pamětní knihy obecní (obecní kroniky), státní symbol a symbol jednotky územní samosprávy a jiná taková díla, u nichž je veřejný zájem na vyloučení z ochrany	Zák. č. 121/2000 Sb.
veřejná správa	Správa veřejných záležitostí vykonávaných subjekty veřejné správy a jejich orgány. Veřejnou správu lze rozdělit na státní správu, veřejnou samosprávu a ostatní veřejnou správu.	[23]
veřejnoprávní subjekt	„Veřejnoprávním subjektem“ (se) rozumí jakýkoliv subjekt: zřízený za zvláštním účelem uspokojování potřeb veřejného zájmu, který nemá průmyslovou nebo obchodní povahu a který má právní subjektivitu a je financován převážně státem, regionálními nebo místními orgány nebo jinými veřejnoprávními subjekty nebo je těmito subjekty řízen, nebo v jeho správním, řídicím nebo dozorčím orgánu je více než polovina členů jmenována státem, regionálními nebo místními orgány nebo jinými veřejnoprávními subjekty	Směrnice 2003/98/ES
VFR	Výměnný formát RÚIAN/ISÚI	[15]

6 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

Zkratka	Význam
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

7 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/1Yw7H9qblPFOD02pgYIC_M492N8WI2c-r4BxiZH1m9Q/export?format=pdf&id=1Yw7H9qblPFOD02pgYIC_M492N8WI2c-r4BxiZH1m9Q
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. Mouzakitis, 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. OGDl 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

7.1 Odkazované předpisy

7.1.1 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

7.1.2 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)

Příloha A – Varianty budování datového katalogu

$V \setminus k$	B	FB	FP	Q	Prop
A	1,1,1,1,1 (5)	1,2,2,1,2 (8)	2,3,2,3 (13)	1,1,1,1,1 (5)	1,1,1,1,1 (5)
B	1,1,1,1,1 (5)	1,2,2,1,2 (8)	1,2,2,1,2 (8)	3,3,3,2,2 (13)	4,1,1,4,2 (9)
C	2,2,2,2,3 (11)	2,3,3,3,2 (12)	2,3,3,3,3 (14)	3,3,3,2,2 (12)	3,1,3,2,3 (14)
D	2,2,3,3,3 (13)	3,3,3,2,3 (14)	4,3,4,4,3 (18)	3,4,3,4,3 (17)	4,3,4,4,4 (19)
E	2,2,3,3,3 (13)	3,3,3,2,3 (14)	3,3,3,3,3 (15)	2,3,3,3,3 (14)	4,3,3,4,4 (16)
F	2,2,3,3,3 (13)	3,3,3,2,3 (14)	2,2,2,2,3 (11)	3,3,3,3,3 (15)	4,3,4,4,4 (19)

Obrázek 8: Hodnocení variant budování datového katalogu

Příloha B – Žádost o poskytnutí informací

Otevřená výzva na poskytnutí informací - nástroj pro datový katalog

Věc: Žádost o poskytnutí informací pro potřeby Ministerstva vnitra ČR

Vážený pane / paní,

Ministerstvo vnitra České republiky se podílí na plnění Akčního plánu České republiky „Partnerství pro otevřené vládnutí“. V současné době je zpracovávána Koncepce katalogizace otevřených dat veřejné správy ČR. V této souvislosti je řešena otázka pořízení a provozu softwarového nástroje pro Katalog otevřených dat veřejné správy ČR.

Pro zjištění potřebných indikativních informací pro vybudování takového řešení a jeho následného provozu se Ministerstvo vnitra rozhodlo Vás oslovit s žádostí o spolupráci a poskytnutí informací - vyplnění dotazníku uvedeného příloze 1. Bližší specifikace požadavků na katalogizační softwarový nástroj je uvedena v příloze 2.

Jsme si vědomi toho, že požadavek na poskytnutí informací je pouze nezávazným průzkumem trhu v oblasti ICT. Záleží tedy jen a pouze na Vás, zda se rozhodnete tohoto průzkumu zúčastnit, a pokud ano, jaké informace nám budete ochotni poskytnout.

Odpovědi zasílejte elektronicky nejpozději do 16.8.2012 do 12:00 hod.

Děkuji za Váš čas, který jste věnovali tomuto průvodnímu dopisu, a případným odpovědím na naše dotazy. V případě nejasností a doplňujících otázek, se můžete obracet na p. (doplněno později).

S pozdravem

(doplněno později)

Příloha 1: Žádost o poskytnutí informace - Dotazník

A) Kontakt na zpracovatele odpovědí na otázky

Vypracoval (jméno, příjmení):	Funkční zařazení:
Kontakt (telefon, e-mail):	Datum vypracování:
Název společnosti:	Sídlo společnosti:

B) SW nástroj pro Katalogizaci dat

Název sw nástroje:	
Dodávané verze:	
Výrobce nástroje, resp. vztah k firmě poskytující nástroj:	
Stručná charakteristika nástroje:	
Reference na nasazení sw nástroje (v ČR i v zahraničí):	
Operační systémy (využitelné pro provoz nástroje)	
Databázová prostředí (využitelná pro provoz prostředí)	
Vývojové prostředí:	

C) Licence, pracnost a ceny

Pozn.:

- Požadavky na nástroj jsou uvedeny v příloze 2.
- Všechny ceny uveďte v rozlišení cena s DPH a cena bez DPH.

Licenční politika:	
Požizovací cena licencí:	
Odhadovaná pracnost dovoje nástroje v člověkodnech dle požadavků uvedených v Příloze 2:	
Odhadovaná cena implementace nástroje včetně nákladů spojených s úpravami nástroje dle požadavků uvedených v Příloze 2:	
Platební podmínky:	

Odhadovaná cena provozu a údržby nástroje po dobu 5 let od ukončení implementace nástroje. Uveďte i způsob výpočtu nákladů na jeden rok provozu a údržby nástroje.	
Pokud je Vaše řešení možné pořídit v modelu „software jako služba (SaaS)“, jaké tarify služby nabízíte a jaká je roční výše nákladů využívání služby	
Další nabízené služby spojené s poimplementační podporou sw nástroje.	

Příloha 2: Specifikace požadavků na katalogizační nástroj

1. Základní funkční požadavky

a. Struktura záznamu v Datovém katalogu (katalogizační záznam).

Prvky struktury, které jsou povinné jsou označeny písmenem *P*, nepovinné prvky jsou pak označeny písmenem *N*. Položky označené „K:“ vyplňuje kurátor při katalogizaci dat. Položky označené „R:“ vyplňuje redaktor při redakci záznamu Datového katalogu. Ostatní povinné položky jsou generovány automaticky. Cílem je maximálně zjednodušit ruční práci při zakládání katalogizačního záznamu.

Základní prvky katalogizačního záznamu jsou:

- o **K: Název záznamu/název dat** (*P*) - výstižné pojmenování katalogizačního záznamu, respektive dat, která katalogizační záznam popisuje
- o **URL katalogizačního záznamu** (*P*) - jednoznačný neměnný identifikátor záznamu používaný v rámci sítě internet, bude vytvořen automaticky
- o **K: Jazyk záznamu** (*P*) - jazyk⁶³, ve kterém je uveden popis záznamu a jeho další atributy
- o **K: Platnost záznamu** (*P*) - obecné označení, zda je katalogizační záznam platný či nikoli⁶⁴
- o **R: Stav redakce záznamu** (*P*) - označení, zda katalogizační záznam prošel redakcí či nikoli
- o **K: Popis dat** (*P*) - text, který jasně uvádí, jaká data katalogizační záznam popisuje a o čem tato data vypovídají
- o **K: 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 dat⁶⁵
 - **Web poskytovatele dat** (*N*) - odkaz na domovskou stránku webu poskytovatele dat⁶⁶
 - **Kontakt** (*P*) - kontaktní email poskytovatele, na kterém poskytovatel poskytuje informace široké veřejnosti⁶⁷
- o **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

⁶³ Katalogizační nástroj musí připouštět možnost vícejazyčného popisu katalogizačního záznamu. Výchozí hodnotou bude „CZ“.

⁶⁴ Výchozí hodnota bude nastavena na „Platný“. Ručně nutno změnit na „Neplatný“.

⁶⁵ Bude automaticky doplněn z registru osob (ROS) na základě zadaného IČO poskytovatele dat (<http://www.szrcr.cz/registr-osob>).

⁶⁶ Bude automaticky doplněn z Portálu veřejné správy (<http://portal.gov.cz>).

⁶⁷ Bude automaticky doplněn z Portálu veřejné správy (<http://portal.gov.cz>).

- **K: Odkaz na popis podmínek užití / znění licence (N)** - nejlépe pomocí uvedení odkazu na webovou stránku se zněním podmínek / licenční smlouvy
- o **K: Vazba na ISVS (N)** - identifikace informačního systému veřejné správy v systému IS o ISVS⁶⁸, 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
- o **K: Vazba na ISDP (N)**⁶⁹ - 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
- o **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
- o **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
 - o **Název konceptu (P)** - úplný název konceptu v jazyce záznamu
 - o **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
 - o **Název konceptu (N)** - úplný název konceptu v češtině (v angličtině, pokud je jazyk záznamu angličtina)
 - o **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
- o **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
 - **K: Název/označení území (N)** - název či označení geografického území, hodnota je určena výběrem ze seznamu
- o **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

⁶⁸ Informační systém o informačních systémech veřejné správy (<https://www.sluzby-isvs.cz/ISoISVS/Applets/DefaultSSL.aspx>)

⁶⁹ Informační systém o datových prvcích (<https://www.sluzby-isvs.cz/ISDP/DefaultSSL.aspx>)

- **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** - 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** - 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/OWL 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 dat⁷⁰
- **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)⁷¹
- **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 automaticky⁷².
- b. Jazyková verze v CZ
 - c. Publikace pouze vybraných (redaktorem schválených) záznamů
 - d. Vyhledávání a procházení obsahu Datového katalogu,
 - e. Publikace obsahu datového katalogu ve formě otevřených a propojených dat (s využitím technologií Linked Data)
 - f. Možnost určit povinné a volitelné atributy záznamu v Datovém katalogu a vyžadovat vyplnění povinných atributů

⁷⁰ Doplněno automaticky na základě zadaného formátu zdroje.

⁷¹ Doplněno automaticky na základě zadaného formátu zdroje.

⁷² Automatické doplnění pomocí služeb geoportálu INSPIRE nebo základního registru RÚIAN.

- 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 (Jednotný identitní prostor / Katalog autorizačních a autentizačních služeb)

2. Nadstavbové funkční požadavky

- a. Možnost vkládat jako volitelné atributy - geokoordináty pro určení příslušnosti dat odkazovaných v záznamu v Datovém katalogu k územní oblasti (vazba na RUIAN 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

3. Bezpečnostní požadavky

- a. Obsah Datového katalogu (katalogizační záznamy) mohou vkládat, měnit a odstraňovat pouze oprávněné osoby
- b. Redakci katalogizačních záznamů mohou provádět pouze oprávněné osoby
- c. Záznamy, které musí projít redakcí, nesmí být zveřejněny před provedením redakce
- d. Činnosti spojené se zajištěním běžného chodu a údržby Datového katalogu mohou provádět pouze oprávněné osoby
- e. 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)
- f. 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
- g. V případě události vedoucí ke ztrátě obsahu Datového katalogu musí být možné určit, které katalogizační záznamy byly ztraceny
- h. Ochrana všech veřejně přístupných kanálů, kterými mohou uživatelé zasílat zpětnou vazbu, zprávy či žádosti do Datového katalogu před zneužíváním (spamem)
- i. 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

4. Možnost škálovat řešení Datového katalogu z hlediska objemu zpracovávaných záznamů

Request for Information

Dear sir / madame,

The Ministry of the Interior of the Czech Republic is participating in the “Open Government Partnership” Czech Republic action plan activities. Czech open government data cataloguing strategy is currently being prepared. As a part of this activity acquisition and maintenance of a software tool for the Czech open government data catalogue is being discussed.

In order to get required information about such solution, its development and subsequent maintenance the Ministry of the Interior of the Czech Republic has decided to publish open request for information in a form of a questionnaire which is provided in the Appendix 1. Detailed specification of requirements for the data cataloguing software is provided in the Appendix 2.

We are aware that this request for information is just a tentative ICT market survey. It is up to you to decide whether you want to participate in this survey or not and what amount of information you can provide.

All responses should be delivered electronically to the Ministry of the Interior of the Czech Republic before 16th August 2012, 12:00 pm.

Thank you for the time you spent reading this letter and eventually answering our questions. In case you have any questions, please, ask (tbd).

Yours faithfully

(tbd)

Appendix 1: Request for information - Questionnaire

A) Contact information

Prepared by: (name, surname):	Position:
Contact (telephone, e-mail):	Prepared on:
Company name:	Company address:

B) Data cataloguing software tool

Name of the software tool:	
Available versions:	
Software tool vendor and its relationship to the distributing company:	
Short description of the tool:	
Reference implementations of the tool (in the Czech Republic and other countries):	
Operating systems (that can be used to run the tool):	
Database systems (that can be used as a storage backend of the tool):	
Development environment:	

C) Licensing, required effort and price

Notes:

1. Specification of the requirements is provided in the appendix 2.
2. Provide all costs with and without VAT.

Licensing policy:	
Upfront licence costs:	
Estimated development effort required to meet the requirements specified in the appendix 2 (in mandays):	
Estimated price of the implementation including costs related to the development and customization according to the requirements specified in the appendix 2:	
Terms of payment:	

Estimated maintenance costs covering period of five years after the implementation is finished. Please, provide calculation formulas and all information necessary to calculate the annual maintenance costs:	
If you provide your solution as a service (SaaS) please, provide information about the offered tariffs and their prices per year:	
Available additional services related to the maintenance of the software tool:	

Appendix 2: Data cataloguing tool requirements specification

1. Basic functional requirements

a. Structure of the record of the Data catalogue (catalogue record).

Record attributes marked with the letter R are required attributes that must be filled in during the record creation. Attributes marked with the letter O are optional. These attributes must be included in the catalogue record but they can be left empty during the record creation. Attributes marked with "C:" are filled in by the curator of the data, attributes marked with "E:" are filled in by the editor of the Data catalogue. The rest of the attributes is filled in automatically by the cataloguing tool. The goal is to make creation of the catalogue record as easy as possible.

Basic catalogue record attributes:

- **C: Name of the record/data (R)** - name of the catalogue record or the data described by the catalogue record
- **URL of the catalogue record (R)** - unique permanent identifier of the record used on the Internet, it is generated automatically
- **C: Language of the catalog record (R)** - language⁷³ of the description and other attributes of the record
- **C: Record validity (R)** - distinguishes valid and invalid records⁷⁴
- **E: Approval state (R)** - distinguishes approved and rejected records and records waiting for approval
- **C: Description of the data (R)** - text describing the data in the catalogue record and what this data is about
- **C: Publisher of the data (R)** - identification of the publisher of the data which consists of the following attributes
 - o **C: IČO (R)** - identification number of the publisher of the data
 - o **Publisher's name (R)** - complete valid name of the publisher of the data⁷⁵
 - o **Publisher's web site (O)** - link to the homepage of the publisher's web site⁷⁶
 - o **Contact (R)** - email of the publisher which is available to the public⁷⁷
- **C: Terms of use of the data (R)** - information about the terms and conditions under which the data in the catalogue record can be used, it consists of the following attributes

⁷³ Data cataloguing tool must support multilingual catalogue records - description and values of other attributes in multiple languages

⁷⁴ Default value is "Valid". It can be manually changed to "Invalid".

⁷⁵ Automatically loaded from the Registry of persons (<http://www.szrcr.cz/registry-of-persons>) based on the provided identification number.

⁷⁶ Automatically loaded from the Portal of the Public Administration (<http://portal.gov.cz>).

⁷⁷ Automatically loaded from the Portal of the Public Administration (<http://portal.gov.cz>).

- o **C: Overview of the data openness criteria (R)** - overview of the criteria for the open government data and how the data in the catalog record meets this criteria
 - o **C: Terms of use / Licence (O)** - name or title of the terms of use / licence
 - o **C: Link to the terms of use / licence (O)** - link to the web site with the full text of the terms of use of the data or the full text of the licence under which the data is published
- **C: Relation to the PSIS⁷⁸ (O)** - identification of the public sector information system in the information system about the public sector information systems (IS about the PSIS)⁷⁹ in case that the data in the catalogue record is produced by such an information system, the identification of the system consists of the following attributes that must not be left empty in case that the identification is provided
 - o **C: Name of the PSIS (O)** - full name of the relevant PSIS
 - o **C: Link to the IS about the PSIS (O)** - identifier of the PSIS in the IS about the PSIS
- **C: Relations to the DEIS⁸⁰ (O)** - identification of the data elements that are used in the data of the catalogue record, every data element is identified by the following attributes that must not be left empty if the identification of the data element is provided
 - o **C: Name of the data element (O)** - full name of the data element
 - o **C: Link to the DEIS (O)** - identifier of the data element in the data elements information system
- **C: Related catalogue records (O)** - identification of the catalogue records that are related to the given catalogue record, this identification consists of the following attributes that must not be left empty if the identification is provided
 - o **C: Name of the record (O)** - name of the related catalogue record
 - o **C: Type of the relation (O)** - title of the relationship between the records (for example "complements", "substitutes" etc., cataloguing methodology will specify the types of relationship)
 - o **C: URL of the catalogue record (O)** - unique permanent identifier of the record used on the Internet
- **C: Classification of the record (R)** - catalogue record is classified using concepts of the classification taxonomies and by freely created tags if necessary
 - o **C: Concepts of the classification taxonomies (R)** - list of the concepts of the taxonomies that the data is labeled with
 - **C: EUROVOC (R)** - list of the EUROVOC concepts, every concept has the following attributes
 - **Name of the concept (R)** - full name of the concept in the language of the catalogue record
 - **Link to the concept (R)** - link to the web page of the concept from the portal of this taxonomy
 - **C: CZ-NACE (O)** - list of the relevant CZ-NACE concepts, every concept has the following attributes that must not be left empty if the concept is used
 - **Concept name (O)** - full name of the concept in Czech (in English if the language of the record is English)
 - **Link to the concept (R)** - link to the web page of the concept from the portal of this taxonomy
 - **C: Freely created tags (O)** - list of freely created tags used to classify the data in this catalog record, a tag is a freely created text string
- **C: Related geographic area (O)** - identification of the geographic area that the data in the catalogue record is related to, this identification consists of the following attributes that must not be left empty if the identification of the area is provided

⁷⁸ Public sector information system

⁷⁹ Information system about the public system information systems (<https://www.sluzby-isvs.cz/ISolSVS/Applets/DefaultSSL.aspx>)

⁸⁰ Data elements information system (<https://www.sluzby-isvs.cz/ISDP/DefaultSSL.aspx>)

- o **C: Name / title of the area (O)** - name or title of the geographic area, value of this attribute is selected from a list
- **C: Data sources (R)** - identification of the data sources that provide the data of this catalogue record, every data source has the following attributes
 - o **C: Data source name (R)** - full name of the data source
 - o **C: Data source description (R)** - text describing the data source
 - o **C: Data source type (R)** - type of the data source (e.g. file, web service, etc.)
 - o **C: Data source URL (R)** - URL for access to the data (e.g. URL of a data file, URL of a web service endpoint etc.)
 - o **C: Data source language (O)** - language of the data source if the language is relevant
 - o **C: Data source format (R)** - format of the data provided by the data source (e.g. concrete format of the data file or title of the data type in general) in a form that humans can easily understand
 - **C: Data source format title (R)** - text string representing the data source format, for example data file suffix if it is commonly used as the data format title (XLS, CSV etc.) or other commonly used text string (e.g. XML)
 - **C: Data source format documentation URL (R)** - URL of a document describing the data source format (e.g. documentation of the XML schema if the data source format is XML). In case that the documentation is provided in multiple documents, URLs of each of these documents are provided.
 - **C: Data source schema URL (R)** - URL of a file containing the data schema specifying the data source format in a machine readable language in a way that allows automatic validation of the data provided by the data source in case that such a language exists (e.g. XML Schema language for XML format, RDFS/OWL for RDF format). In case that the schema is described in multiple documents, URLs of each of these documents are provided.
 - o **MIME type (R)** - standardized title of the data format⁸¹
 - o **MIME type inner (O)** - standardized title of the data format that is wrapped in other data format (e.g. format of a data that is compressed by some compression algorithm)⁸²
 - o **C: File size (O)** - size of a file in bytes in case that the data source is a data file
 - o **C: Date of the last modification (O)** - date when the data source was last modified
 - o **Date of the last modification to the record (R)** - date when the catalogue record was last modified, it is filled in automatically by the cataloguing tool
 - o **Author of the last modification to the record (R)** - identification of the person who made the last modification to the catalogue record, it is filled in automatically by the cataloguing tool
 - o **C: Author of the last modification contact (R)** - contact information about the person who made the last modification to the catalogue record

Advanced catalogue record attributes:

- **C: Geo-coordinates of the area (O)** - set of geo-coordinates which is necessary to display the area on a map, in order to avoid the manual entry of the geo-coordinates it is recommended that curator of the data only selects the area (from the list or on a map) and the cataloguing tools fills in the set of geo-coordinates automatically⁸³.

- b. Tool available in Czech language
- c. Publication of selected records (records approved by the editor)
- d. Searching and browsing the data catalogue

⁸¹ Filled in automatically based on the selected data source type.

⁸² Filled in automatically based on the selected data source type.

⁸³ Automatic filling of geo-coordinates using services provided by the INSPIRE geoportal or the Basic register of area identification, addresses and properties.

- e. Publication of the data data catalogue contents in the form of Linked Open Data (using linked data technologies)
- f. Ability to set required and optional attributes of the catalogue record and force filling of the required ones
- g. Ability to use external classification taxonomies (e.g. EUROVOC)
- h. Ability to integrate automatic tools (e.g. content quality management tools)
- i. Existence of a defined application programming interface (API)
- j. Ability to submit proposals to improve quality of the data catalogue contents by the end users
- k. Approval process for the catalogue records including ability to mark catalogue record attributes that need to be corrected and ability to notify the publisher of the data
- l. Statistics and analysis of the records in the data catalogue (numbers of catalogue records per concepts of the classification taxonomies, development over time and number of records per publisher)
- m. Unified identification and authentication of users using the system Unified identity space

2. Extending functional requirements

- a. Ability to add geo-coordinates to mark an area that the data is related to (relationship to the Basic register of area identification, addresses and properties)
- b. Visualization of the data (e.g. displaying data on a map)
- c. Ability to use external classification taxonomies (e.g. EUROVOC) without the need to copy the taxonomies into the cataloguing tool
- d. Search by geographic locations
- e. Multilingual data cataloguing tools - ability to create and store translations of the catalogue records into different languages

3. Security requirements

- a. Contents of the data catalogue can be created, changed and removed only by the authorized persons
- b. Catalogue records can be approved only by the authorized persons
- c. Catalogue records that must be approved before they can be published must not be published without the approval
- d. Maintenance of the data catalogue can be performed only by the authorized persons
- e. In case of event that causes inaccessibility or unavailability of the data catalogue and its contents it must be possible to make it fully available again within 24 hours since the inaccessibility of the data catalogue is reported (it does not apply to the planned maintenance of the data catalogue)
- f. In case of event leading to loss of the data catalogue contents it must be possible to restore the contents at least to the state it was one hour before this event
- g. In case of event leading to loss of the data catalogue contents it must be possible to determine which catalogue records were lost
- h. All channels that users can use to send feedback or messages to the data catalogue must be protected from abuse (spam)
- i. Ability to determine who made changes to the contents of the data catalogue, who performed actions related to the approval of the catalogue records and actions related to the maintenance of the data catalogue

4. Scalability of the Data catalog solution according to the amount of processed cataloguing records