

Univerzita Pardubice

Fakulta ekonomicko-správní

Otevřená data a jejich praktické využití

Kateřina Komárková

**Bakalářská práce
2017**

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2016/2017

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Kateřina Komárková**
Osobní číslo: **E13935**
Studijní program: **B6202 Hospodářská politika a správa**
Studijní obor: **Veřejná ekonomika a správa: Veřejná ekonomika**
Název tématu: **Otevřená data a jejich praktické využití**
Zadávací katedra: **Ústav ekonomických věd**

Z á s a d y p r o v y p r a c o v á n í :

Cílem práce je charakteristika otevřených dat a jejich legislativního pojetí v České republice. V práci budou popsány možnosti využití těchto dat s uvedením konkrétních příkladů. Cílem práce je analytické zpracování těchto dat pomocí zvoleného nástroje.

Osnova:

- Charakteristika otevřených dat a jejich legislativního pojetí v České republice.
- Metodika publikace otevřených dat veřejné správy.
- Konkrétní příklady dobře a špatně zveřejněných dat.
- Možnosti využití těchto dat v soukromé sféře.
- Analýza dat z oblasti veřejné správy za použití zvoleného nástroje.

Rozsah grafických prací: -

Rozsah pracovní zprávy: **cca 35 stran**

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam odborné literatury:


BERG, M., BOČEK, J. Otevřená data ve státní správě - nová éra rozhodování. Praha: Nadace Open Society Fund Praha, 2012. ISBN 978-80-87110-24-9

LIDINSKÝ, V. et al. eGovernment bezpečně. Praha: Grada, 2008. 145 s. ISBN 978-80-247-2462-1

MRÁČEK, J., BOČEK, J., ČEPICKÝ, J. Jak otevírat data?. Praha: Fond Otakara Motejla, 2014. ISBN 978-80-87725-16-0

PEKOVÁ, J. Veřejné finance: teorie a praxe v ČR. Vyd. 1. Praha: Wolters Kluwer Česká republika, 2011. 642 s. ISBN 978-80-7357-698-1

Vedoucí bakalářské práce:


Ing. Simona Pichová

Ústav ekonomických věd

Datum zadání bakalářské práce: **4. září 2016**

Termín odevzdání bakalářské práce: **28. dubna 2017**


doc. Ing. Romana Provazníková, Ph.D.

děkanka

L.S.


doc. Ing. Jolana Volejníková, Ph.D.

vedoucí ústavu

V Pardubicích dne 4. září 2016

PROHLÁŠENÍ

Prohlašuji, že jsem tuto práci vypracoval/a samostatně. Veškeré literární prameny a informace, které jsem v práci využil/a, jsou uvedeny v seznamu použité literatury.

Byl/a jsem seznámen/a s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako Školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnicí Univerzity Pardubice č. 9/2012, bude práce zveřejněna v Univerzitní knihovně a prostřednictvím Digitální knihovny Univerzity Pardubice.

PODĚKOVÁNÍ

Tímto bych ráda poděkovala svému vedoucímu práce Ing. Simoně Pichové za její odbornou pomoc, konzultace a cenné rady, které mi pomohly při zpracování bakalářské práce.

Rovněž bych také chtěla poděkovat panu Bc. Martinu Sakalovi ze společnosti Bits2b, s.r.o., za vstřícnost a pomoc při zpracovávání potřebných informací a podkladů.

ANOTACE

Cílem bakalářské práce je charakteristika otevřených dat a jejich legislativního pojetí. Dále popis metodiky publikace otevřených dat a uvedení konkrétních příkladů jejich zveřejnění a následného využití. V závěru práce pak analýza otevřených dat z oblasti veřejné správy za použití zvoleného nástroje.

KLÍČOVÁ SLOVA

otevřená data, propojená data, datový katalog, kvalita dat

TITLE

Open Data and their practical use.

ANNOTATION

The aim of this Bachelor's dissertation is to define "open data" term and its legal frame. The methodology of open data publication and its subsequent usage including specific cases is also part of this study. Analysis of the open data coming from public administration by using elected tool.

KEYWORDS

open data, linked data, data catalogue, data quality

OBSAH

ÚVOD	10
1 CHARAKTERISTIKA OTEVŘENÝCH DAT A JEJICH LEGISLATIVNÍHO POJETÍ	11
1.1 POJEM OTEVŘENÁ DATA.....	11
1.2 PROPOJITELNÁ OTEVŘENÁ DATA	12
1.3 VZNIK OTEVŘENÝCH DAT.....	13
1.4 OTEVŘENÁ DATA V ČESKÉ REPUBLICE	14
1.5 PRÁVNÍ ÚPRAVA V ČR	15
1.6 PRÁVNÍ ÚPRAVA NA ÚROVNI EVROPSKÉ UNIE	16
2 ODŮVODNĚNÍ PUBLIKACE OTEVŘENÝCH DAT	17
2.1 MOŽNOSTI VYUŽITÍ DAT V SOUKROMÉ SFÉŘE.....	18
2.1.1 <i>Mapa kriminality.cz</i>	18
2.1.2 <i>Váš majetek.cz</i>	19
2.1.3 <i>Mapa exekucí.cz</i>	19
2.1.4 <i>Návrh na další možné využití otevřených dat</i>	20
3 METODIKA PUBLIKACE OTEVŘENÝCH DAT VEŘEJNÉ SPRÁVY	21
3.1 DATOVÁ SADA.....	21
3.2 DATOVÝ KATALOG.....	21
3.3 STUPNĚ OTEVŘENOSTI DAT	22
3.4 PROCES PUBLIKACE OTEVŘENÝCH DAT	24
4 KONKRÉTNÍ PŘÍKLADY DOBŘE A ŠPATNĚ ZVEŘEJNĚNÝCH DAT	26
4.1 ZVEŘEJNĚNÍ JÍZDNÍCH ŘÁDŮ VEŘEJNÉ DOPRAVY.....	26
4.2 ZVEŘEJŇOVÁNÍ DAT DLE ZÁKONA O REGISTRU SMLUV	27
5 ANALÝZA DAT Z OBLASTI VEŘEJNÉ SPRÁVY ZA POUŽITÍ ZVOLENÉHO NÁSTROJE	30
5.1 OTEVŘENÁ DATA Z REGISTRU SMLUV	30
5.2 PŘÍPRAVA DAT.....	31
5.3 VIZUALIZACE DAT	31
5.4 CHYBY V REGISTRU SMLUV	33
5.4.1 <i>Identifikace smluvních stran</i>	33
5.4.2 <i>Vymezení předmětu smlouvy</i>	35
5.4.3 <i>Chybějící nebo špatně zadaná cena</i>	35
5.4.4 <i>Identifikační číslo smluvní strany</i>	37
5.4.5 <i>Chybné označení plátce a příjemce</i>	37
5.4.6 <i>Duplikace smluv</i>	38
5.4.7 <i>Nedodržení lhůty od uzavření smlouvy k jejímu zveřejnění</i>	38
5.5 SHRNUTÍ KVALITY DAT Z REGISTRU SMLUV	39
ZÁVĚR	40
POUŽITÁ LITERATURA	41
SEZNAM PŘÍLOH	43

SEZNAM TABULEK

Tabulka 1: Příklad správně sestavené tabulky pro datové sady	21
Tabulka 2: Shrnutí všech požadavků, které musí splňovat data v jednotlivých stupních otevřenosti.....	24

SEZNAM ILUSTRACÍ

Obrázek 1: Stupně otevřenosti dat podle „5-star deployment scheme for Open Data“	22
Obrázek 2: Proces publikace otevřených dat ve veřejné správě.....	25
Obrázek 3: Údaje z adresáře o jednom dumb souboru.....	30
Obrázek 4: Výstup z analytického softwaru Qlik Sense s daty z Registru smluv.....	32
Obrázek 5: Počet zveřejněných smluv za období 1.7.2016-31.3.2017.....	33
Obrázek 6: Graf špatně zveřejněných smluvních stran	34
Obrázek 7: Graf znázorňující procento záznamů zveřejněných bez ceny a alespoň s jednou cenou.....	35
Obrázek 8: Graf zobrazující chyby ve zveřejňování ceny v metadatech	36
Obrázek 9: Procentuální rozdělení záznamů podle toho, zda je u nich uveden údaj o plátcí či příjemci peněz	37

SEZNAM ZKRATEK A ZNAČEK

CSV	Formát dat, kde jsou hodnoty oddělené čárkami.
ČR	Česká republika
ČSÚ	Český statistický úřad
DPH	Daň z přidané hodnoty
EU	Evropská unie
GIF	Grafický pohyblivý formát.
HTML	Značkovací jazyk určený k tvorbě webových stránek.
ICT	Informační a komunikační technologie.
IČO	Identifikační číslo organizace.
ID	Identifikace
IT	Informační technologie.
JPG	Grafický rastrový formát.
JSON	Otevřený textový formát pro přenos dat mezi serverem a webovou aplikací.
KČ	Koruna česká.
MS	Microsoft
OECD	Organizace pro hospodářskou spolupráci a rozvoj.
OSN	Organizace spojených národů.
PDF	Přenosný formát dokumentů.
PNG	Přenosná síťová grafika.
Sb.	Sbírka zákonů
TED	Technologie, Zábava, Design.
TXT	Formát textového souboru.
URL	Jednotná adresa zdroje.
USA	Spojené státy americké.
ÚOHS	Úřad ochrany hospodářské soutěže.
VS	Veřejná správa.
XML	Formát obecného rozšiřitelného značkovacího jazyka.
XLS	Formát dat souboru, specifický pro Microsoft Excel.
XLSX	Formát dat souboru, specifický pro Microsoft Excel.

ÚVOD

Tato práce se zabývá problematikou otevřených dat. Termín „otevřená data“ je v oblasti informačních technologií a businessu již poměrně známý a čím dál tím více narůstá potřeba se těmito daty zabývat, shromažďovat je a analyzovat ve prospěch vlastního podnikání a dosažení většího zisku. Kde jsou však z mého pohledu největší rezervy ve využití otevřených dat, je veřejná správa. V této oblasti dochází ke sběru a ukládání velkého množství velmi zajímavých dat, které v sobě skrývají obrovský potenciál a při jejich využití mohou vést k celkovému zlepšení efektivity veřejné správy, k úsporám veřejných prostředků pro fungování státní správy a samosprávy, jednodušší komunikaci mezi jednotlivými úřady, ale také k usnadnění komunikace občanů s úřady a vyšší transparentnosti státních institucí.

V prvních kapitole se práce zabývá samotnou charakteristikou otevřených dat, jejich historií a také legislativní úpravou. Druhá kapitola je zaměřena na nejdůležitější přínosy otevírání dat pro veřejnou správu, stát, ale také pro samotné občany. Dále jsou zde popsány některé již existující možnosti využití otevřených dat z oblasti soukromé podnikatelské sféry. Ve třetí kapitole je objasněno několik základních pojmů, které souvisí s otevíráním dat a dále se práce věnuje metodice publikace otevřených dat, ve které jsou popsány jednotlivé stupně otevřenosti dat, které určují jejich výslednou kvalitu a také proces publikace otevřených dat krok po kroku. Čtvrtá kapitola je zaměřena na praktické využití dat a jsou v ní uvedeny dva konkrétní příklady dobře a špatně zveřejněných dat v České republice. V závěrečné páté kapitole jsou pak správně otevřená data z kapitoly 4 analyzována pomocí zvoleného nástroje. Tímto vybraným nástrojem pro analýzu dat je software Qlik Sense Desktop. Tento nástroj je v současné době jedním z nejlepších pro práci s daty a jejich vizualizaci, a také je velmi dostupný pro běžné uživatele, protože je možné jej v základní verzi stáhnout a bezplatně využívat.

Cílem práce je charakterizování pojmu otevřená data, jejich legislativní úpravy na území České republiky a popis možností využití těchto dat s uvedením konkrétních příkladů. Dalším cílem je analytické zpracování konkrétních dat z oblasti veřejné správy. Analyzována budou data z Registru smluv, který je veřejný a data z něj volně dostupná na internetu ke stažení. Práce bude obsahovat jejich souhrnné statistické zpracování a dále analýzy těchto výsledků.

1 CHARAKTERISTIKA OTEVŘENÝCH DAT A JEJICH LEGISLATIVNÍHO POJETÍ

1.1 Pojem otevřená data

Označení otevřená data je poměrně mladý pojem označující formu, nikoliv obsah dat. Může se jednat o data, která produkuje každý z nás, například o našem zdravotním stavu či sportovních výkonech, může se ale také jednat o data, která sbírají univerzity při svých výzkumech nebo konkrétní firmy o svých zákaznících či zaměstnancích. Tato práce se však bude zabývat především daty veřejné správy, tedy těmi, které sbírá nebo vytváří státní správa a samospráva. Protože je pojem otevřená data tak čerstvý, není prozatím spoután určitou jednoznačnou definicí. Velmi obecnou, a tedy i dobře přijatelnou pro otevřená data, je definice podle Open Knowledge¹ – Open Definition („otevřených znalostí“). Otevřená, podle této definice znamená, že jsou data dostupná online na internetu, kde je může kdokoliv využívat, upravovat či sdílet bez jakýchkoliv omezení. Využívat a sdílet je může kdokoliv za podmínky, že bude vždy uveden autor těchto dat a stejné podmínky budou platit i pro další uživatele [10].

Podle Open Knowledge je dále potřeba, aby data byla otevřená jak technicky, tak právně. Technicky otevřená data jsou taková, která jsou úplná, tedy je přístupný celý obsah databáze, ve strojově čitelném formátu a otevřená za minimálních možných nákladů. Co se týče právní otevřenosti, musí být data zveřejněna pod otevřenou licenci, tedy veřejně dostupnou pro všechny uživatele.

Organizace Sunlight Foundation² v roce 2010 definovala vlastnosti, které by měla mít otevřená data veřejné správy. Tato data by měla podle [18] být:

1. **Úplná** – tedy zveřejněná v maximálním možném rozsahu. Tento rozsah může být definován usnesením vlády, právním předpisem či přímo poskytovatelem dat.
2. **Primární** – nebo jinak řečeno původní, jsou data zveřejněná původcem dat v takové formě, v jaké byla vytvořena.
3. **Zveřejněná bez zbytečného odkladu** – to znamená, že zveřejnění dat není zpožděno kvůli činnostem nesouvisejícím s jejich přípravou. Činnosti, které jsou nezbytné pro publikaci dat, musí být provedeny bez dlouhé prodlevy od okamžiku vzniku dat.

¹ Open Knowledge - Mezinárodní nezisková organizace působící v oblasti otevřených dat.

² Sunlight Foundation – Nezisková organizace zabývající se otevřenými daty.

4. **Snadno dostupná elektronicky či fyzicky** – jsou data dostupná na internetu a snadno dohledatelná běžnými ICT nástroji a prostředky.
5. **Strojově čitelná a strukturovaná** – tedy ve vhodném formátu připraveném pro tyto účely k okamžitému využití. Může se jednat o grafický formát (fotografie, naskenovaný dokument), geodata či textové dokumenty. Nejčastěji se však jedná o strukturovaná číselná data, tedy ve formátu tabulky.
6. **Neomezující přístup** – jsou data dostupná způsobem nediskriminujícím jednotlivce či skupinu uživatelů.
7. **Používající otevřené standardy** – tato data musí být ve formátu, který je bezplatně dostupný pro jakékoliv využití, nebo jsou bezplatně do tohoto formátu převoditelná dostupnými aplikacemi.
8. **Dostupná k dalšímu využití bez omezení (uvolněná jako součást „public domain“³)** – podmínky využití dat musí být zřetelně a jasně definovány a zveřejněny.
9. **Stále dostupná** – jsou dostupná on-line po dobu uvedenou jejich poskytovatelem
10. **Dostupná uživatelům při vynaložení minima možných nákladů na jejich získání** – poskytovatelé dat jsou oprávněni žádat úhradu v maximální výši nepřevyšující výši nákladů spojené s jejich zpřístupněním uživateli.

Těchto deset podmínek pro zveřejňování dat bylo na základě *Koncepce katalogizace otevřených dat VS ČR* s úpravami implementováno do prostředí české veřejné správy. Úpravami je myšleno zejména to, že data veřejné správy ČR jsou považována za otevřená už ve chvíli, kdy splňují alespoň šest podmínek a těmi jsou podmínky č. 1, 4, 5, 7, 8 a 10. Není tedy striktně požadováno, aby data byla primárně, zveřejněna bez zbytečného odkladu, neomezující přístup a stále dostupná, protože tyto podmínky není vždy snadné zajistit. Samozřejmě pokud některá data splňují všech deset podmínek, jsou potom mnohem lépe využitelná pro potenciální uživatele a je proto vhodné se o jejich naplnění alespoň pokusit [4].

1.2 Propojitelná otevřená data

Přestože je tato práce zaměřena především na problematiku otevřených dat, je tento pojem vhodné rozšířit ještě o tzv. propojitelná otevřená data. Ta vznikají ve chvíli, když se strukturovaná data začnou navzájem propojovat na základě vzájemných souvislostí.

³ Public domain – Autor díla se rozhodl poskytnout dané dílo volně k využití bez nároku na jeho další ochranu.

K propojení se využívá technologie propojitelných dat, které umožňují vzájemně propojovat související data pocházející z různých zdrojů a také vyjádřit, o čem data a propojení mezi nimi vypovídají a zařadit je do kontextu. Jsou propojena pomocí odkazů, stejně jako je to u webových stránek s tím rozdílem, že u webových stránek jejich propojení nenese tvrzení o jejich vzájemném vztahu. Propojováním jednotlivých datových sad roste informační hodnota všech těchto dat a v jejich narůstajícím množství je to nezbytnou součástí práce s daty. Samotná veřejná správa disponuje velkým množstvím dat, které spolu navzájem souvisí a pro jejich lepší použitelnost je nutné tyto souvislosti zachytit v podobě jejich propojení. Mezi pojmy propojitelná a propojená data je velký rozdíl. Oba dva využívají stejné principy, ale liší se tím, zda se jedná o data, která jsou pro propojení zatím pouze připravena z pohledu použitých technologií (propojitelná data), nebo zda jsou tato data již skutečně propojena s jinými souvisejícími daty (propojená data) [1].

1.3 Vznik otevřených dat

O otevřených datech, se vůbec poprvé začíná mluvit ve Velké Británii v roce 2009 a to na popud několika britských autorit, které po tamním Ministerstvu dopravy požadovalo zveřejnění dat o jízdních řádech. Tato data byla do té doby komerční a chráněna velmi drahou licencí. O otevření těchto dat se zasloužil i tehdejší britský premiér Gordon Brown, který podpořil nátlak na ministerstvo, a to nakonec požadovaná data zdarma zveřejnilo. Data měla především sloužit komukoliv s nápadem na jejich využití a inovace. Zpočátku vzniklo mnoho nových mobilních aplikací, které využívaly tato data, a také byl velký zájem o zveřejnění dat z jiných oblastí veřejné správy. Tak tedy vznikl pojem otevřená data [10]. S prvním systematickým zveřejňováním dat v otevřené podobě přišla ve stejném roce vláda USA v čele s Barackem Obamou, který v rámci své administrativy vyhlásil iniciativu otevřeného vládnutí (angl. Open Government). Tato iniciativa a její koncept je založen na zpřístupnění informací široké veřejnosti s důrazem na jejich technickou a právní otevřenost. Postupem času se rozrostla do takových rozměrů, že dnes je tento pojem známý a využíván již po celém světě. Největšími lídry v této oblasti jsou státy jako již zmiňované USA v čele s Barackem Obamou, Velká Británie, kde je velkým stoupencem otevřených dat nynější britský premiér David Cameron společně se zakladatelem služby www, tedy moderního internetu, Tim Berners-Lee, kteří oba mimo jiné na toto téma vystupovali na prestižní kalifornské konferenci TED [1].

1.4 Otevřená data v České republice

V České Republice se o otevřených datech poprvé jednalo v roce 2011, kdy se na zasedání vlády 14. září schválilo přistoupení České republiky k Open Government Partnership (OGP). OGP je iniciativa, která zastřešuje všechny aktivity směřující k otevírání dat a Česká Republika se připojením k ní zavázala k tomu, že bude podporovat transparentní, otevřenou a efektivní státní správu, jež se od otevřeného vládnutí očekává. Velkou zásluhu na přistoupení ČR k otevřeným datům měl Fond Otakara Motejla⁴, který od začátku usilovně prosazoval otevírání dat pro transparentní veřejnou správu. Fond na počátku roku 2012 vydal první studii s názvem „Otevřená data ve státní správě: Nová éra rozhodování“ a v dubnu stejného roku také došlo ke schválení prvního Akčního plánu pro otevřené vládnutí vládou ČR. V lednu roku 2013 byl spuštěn web otevrenadata.cz, který si klade za cíl rozšířit využití otevřených dat v ČR a zároveň informovat širokou veřejnost o možnostech využití dat z veřejné správy. Díky tomuto fondu, se pojem „otevřená data“ dostal do vládních dokumentů a některé první instituce a města začaly svá data zveřejňovat. Patřily mezi ně například Český statistický úřad, Česká obchodní inspekce či město Děčín a další [10]. Fond Otakara Motejla od roku 2013 také každoročně vyhlašuje soutěž s názvem „Společně otevíráme data“ jejímž cílem je ocenit nejlepší aplikace, fungující nad otevřenými daty. Této soutěže se mohou zúčastnit studenti, neziskové organizace, firmy ale i veřejnost a soutěžit tak o zajímavé ceny. Velkým přínosem byl pak rok 2015, kdy se k otevírání svých dat připojila i některá ministerstva. Nejlépe se tohoto úkolu zhostilo Ministerstvo financí v čele s tehdejším náměstkem Lukášem Wagenknechtem. Po vzoru ministerstev poté následoval důležitý krok ze strany pražského magistrátu a jím zřízeného Institutu pro plánování a rozvoj, který vytvořil vlastní katalog otevřených dat, který je přístupný i ostatním městským firmám a úřadům v Praze. O tom, že se však otevírání dat netýká jen velkým měst a institucí svědčí fakt, že svá data začala sbírat a zveřejňovat i menší města jako například Bohumín, nebo nově také obec Huntířov, které má necelých tisíc obyvatel.

V březnu roku 2016 předložila vláda poslanecké sněmovně návrh novely zákona č.106/1999 Sb. o svobodném přístupu k informacím, která poprvé zavádí termín „otevřená data“ do českého právního prostředí. O tom už ale více v následujícím odstavci [13].

⁴ Fond Otakara Motejla – Fond založený v roce 2012, který se snaží prostřednictvím aktivní společnosti a otevřených dat docílit efektivní a transparentní veřejné správy v ČR.

1.5 Právní úprava v ČR

Mezi hlavní zákony, které se týkají otevřených dat, jejich zpřístupnění ze strany státní správy a samosprávy v ČR, patří zákony č.106/1999 Sb. o svobodném přístupu k informacím a zákon č. 123/1998 Sb. o právu na informace o životním prostředí. První ze zmiňovaných zákonů, tedy zákon o svobodném přístupu k informacím, se týká subjektů státní správy a samosprávy, ale také organizací a firem jimi zřízených a vlastněných. Nařizuje jim, že musí zveřejňovat některé povinné informace, ale také další informace, o které si zažádají sami občané [10]. Jak již bylo zmíněno v předchozím odstavci, dne 6.zář 2016 byla prezidentem ČR schválena novela tohoto zákona, která historicky poprvé definuje pojem „otevřená data“ do českého legislativního prostředí. Zákon po novelizaci pojednává o otevřených datech jako o strojově čitelných informacích, které jsou zveřejněny způsobem umožňujícím dálkový přístup, jejich způsob a účel využití není nikterak omezen a jsou evidovány v katalogu otevřených dat. Toto je velká změna oproti původnímu znění zákona, protože pro práci s daty je důležité nejen jejich samotné zveřejnění, ale především to, v jaké formě jsou zveřejněny a jak lze s nimi potom dále pracovat. Novela tedy nestanovuje povinnost zveřejňovat nové informace, ale pouze ty, které již zveřejňované jsou, nově však ve formě otevřených dat. Novela vstoupila v platnost v lednu roku 2017 [13].

Oba zákony jsou společně také zakotveny na ústavní úrovni, a to konkrétně v Listině základních práv a svobod ve člancích 17 a 35. V článku 17 je popsáno právo na informace jako jedno ze základních politických práv občana, naproti tomu v článku 35 se toto právo nachází i v právu hospodářském, sociálním či kulturním a zaručuje všem právo na včasné a úplné informace o životním prostředí a přírodních zdrojích. Dále existují také zákony, které mohou jistým způsobem omezovat přístup k informacím. Jedním z nich je například zákon č.412/2005 Sb. o ochraně utajovaných informací a o bezpečnostní způsobilosti, který omezuje zveřejnění dat ze zcela pochopitelných důvodů. Naproti tomu jsou ale zákony, které jsou spíše zneužívány v souvislosti s těmito omezeními. Mezi ně patří utajování informací o obchodním tajemství, ochraně osobnosti nebo osobních údajích, které vycházejí zejména ze zákona č.101/2000 Sb., o ochraně osobních údajů a také z § 11–16 zákona č. 40/1964 Sb., občanského zákoníku. Některé instituce veřejné správy na základě těchto zákonů odmítají poskytnout dané informace, avšak toto jejich jednání je často protizákonné.

Mezi další častá omezení patří například zákaz diskriminace. Jedná se o to, že ke všem již zveřejněným informacím musí mít podle zákona všichni stejný přístup a podmínky poskytování informací musí být transparentní pro všechny žadatele [1].

1.6 Právní úprava na úrovni Evropské unie

V rámci Evropské unie zveřejňování dat upravuje směrnice EK č.2003/98/EC, která se označuje také jako PSI direktiva o informacích veřejného sektoru (Public Sector Information Directive). Tato směrnice z roku 2003 se 26. června 2013 vzhledem k okolnostem dočkala své aktualizace, a to na směrnici 2013/37/EU. Nejvýznamnějšími přínosy novelizace jsou například ve výslovném zavedení obecného pravidla, že jakékoliv dokumenty zpřístupněné orgány veřejného sektoru, lze opakovaně využít k jakémukoliv komerčnímu i nekomerčnímu účelu, pokud tedy nejsou chráněny autorským právem třetí strany. Dále rozšířila okruh institucí, na které se bude povinnost zveřejňování dat vztahovat například o muzea či univerzitní knihovny a také stanovila pravidla o poplatcích za využívání dat, a to především tak, že vládní data by měla být poskytnuta zdarma či za minimální cenu a to takovou, která nepřevyšuje náklady na pořízení kopií nebo datových nosičů. Výjimkou pro výše poplatků jsou pouze data, která pro správní orgán na základě licenční smlouvy sbírá někdo jiný. Potom mohou být poplatky i vyšší. Vlády zemí EU měli stanovenou tuto novelizaci implementovat do konce roku 2015 [15]. V souvislosti s tím, se v ČR připravila novela již zmiňovaného zákona č. 106/1999 Sb. o svobodném přístupu k informacím, do kterého byla směrnice EU transponována a tím zakotvena v českém právním systému. Tento postup napojení, do již existujícího zákona není v kontextu EU zcela typickým [1].

2 ODŮVODNĚNÍ PUBLIKACE OTEVŘENÝCH DAT

Jedním z hlavních důvodů publikace otevřených dat, je možnost zvýšení konkurenceschopnosti ČR na světovém trhu. Neefektivita veřejné správy, vysoká korupce a nedůvěra ke státním institucím jsou hlavními problémy, které jsou řešitelné prostřednictvím otevřených dat. Obecně poskytování informací, je jednou z mnoha služeb veřejných institucí, na kterých se dá tímto ledacos zlepšit. Občan či podnikatel, který se rozhoduje na základě dostatečného množství správných a úplných informací může svou činností dále velkou měrou přispívat k mezinárodní konkurenceschopnosti a to například tím, že dostane možnost tato data dále sdílet a analyzovat, vytvářet nad nimi aplikace, které generují přidanou hodnotu a zisk. Data tedy mohou být zdrojem inovací v různých oblastech jako například ve zdravotnictví, dopravě, logistice či bankovníctví.

Dalším velmi důležitým přínosem otevírání dat je zvýšení ekonomického potenciálu. Otevřená data vytvářejí zcela nový trh, na kterém se z nově získané masy informací dají vytvářet nové webové a mobilní aplikace, které mají přínos jak pro jejich tvůrce v podobě zisku, tak také pro jejich uživatele. Díky těmto projektům jsou data obohacena o přidanou hodnotu a tou je zejména jejich zpracování do podoby, kdy nám poskytují již přehledné a jasné informace a dané oblasti či nám napomáhají v dalším rozhodování.

Jedním z hlavních přínosů, které by mělo přinést otevřené vládnutí, je však zvýšení transparentnosti veřejné správy a jejích institucí. Cílem je zejména podpořit rozpočtovou transparentnost, ale také zvýšit občanskou angažovanost a efektivně bojovat s korupcí. Právě korupce je spojena zejména s problematikou veřejných zakázek. U těch patří mezi nejvýznamnější změny například povinnost zadavatele veřejné zakázky zveřejňovat plné znění smlouvy, seznam subdodavatelů a jejich podíl na ceně a také konečnou cenu zakázky. Samotné jednací řízení a celý proces zadávání veřejné zakázky by měl podléhat uveřejnění dle zákona o svobodném přístupu k informacím [1].

Toto jsou tedy hlavní tedy hlavní přínosy otevřeného vládnutí a s ním spojeném otevírání dat. Pro shrnutí jsou to tedy:

- Zlepšení služeb veřejné správy.
- Podpora ekonomického růstu.
- Posílení transparentnosti úřadů.

2.1 Možnosti využití dat v soukromé sféře

Možností využití otevřených dat je celá řada. Některé, z již existujících řešení fungujících na bázi otevřených dat jsou popsány níže v práci.

2.1.1 Mapa kriminality.cz

V předchozím odstavci jsou uvedené některé hlavní přínosy otevřených dat, ke kterým již v současné době existují reálné služby a aplikace. Příkladem využívání otevřených dat ve prospěch zlepšení služeb veřejné správy je projekt s názvem Mapakriminality.cz, který vytvořil a zrealizoval Michal Tošovský z centra Pro Police Otevřené společnosti v roce 2011. V rámci tohoto projektu sesbíral data o trestné činnosti z celé ČR a vytvořil jednotnou databázi s jednoduchou mapovou aplikací na webu [10]. Cílem tohoto webu je veřejnosti usnadnit orientaci v datech o trestné činnosti poskytovaných Policí ČR. Návštěvník stránek může data procházet přímo prostřednictvím tabulkových a mapových náhledů, nebo si může data sám stáhnout v počítačově zpracovatelném formátu a analyzovat za použití vlastních nástrojů. Procházení dat na webu je velmi přehledné a jednoduché. Uživatel si vybere konkrétní oblast, může to být celá ČR, kraj, okres či jednotlivé městské obvody a dále časovou úsečku, pro kterou chce data vizualizovat. Po výběru těchto údajů se již zobrazí graf, ze kterého lze snadno vyčíst informace o množství spáchaných trestných činů, index kriminality⁵ pro dané území, složení konkrétních typů trestných činů jako například vraždy, znásilnění či krádeže a také procento objasněných trestných činů z celkového počtu pro daný výběr. Lidé tak mohou velmi rychle a snadno zjistit, jak bezpečno je v místě jejich bydliště nebo podle procenta objasněných trestných činů zhodnotit, jak je policie v dané lokalitě úspěšná při řešení a omezování trestné činnosti. Dále pak také mohou porovnávat jednotlivá území mezi sebou, a to pak může být jedním z kritérií při výběru lokality pro nové bydliště [7].

Michal Tošovský s tímto nápadem uspěl na prestižní soutěži On Think Tanks Data Visualisation Competition a mimo jiné začal tento projekt zajímat i vedoucí pracovníky policejního sboru [10]. Policistům, ale i úředníkům ministerstva vnitra se tato webová aplikace natolik zalíbila, že se rozhodli poskytovat sdružení Otevřené společnosti, o.p.s. podrobnější informace pravidelně každý měsíc a díky tomu je mapa kriminality velmi detailní a je možné v ní procházet data zpětně po jednotlivých měsících. Lze tedy po měsících sledovat, jak se kriminalita vyvíjela a reagovala například na otevření nové herny či obchodního centra v dané lokalitě [8].

⁵ Index kriminality – Počet trestných činů na 1000 obyvatel.

2.1.2 Váš majetek.cz

Jako dobrý příklad využívání otevřených dat pro podporu ekonomického růstu a zároveň pro zvýšení transparentnosti úřadů slouží český webový portál www.vasmajetek.cz. Stát, respektive státní instituce, každoročně hospodaří s majetkem ať už movitým, či nemovitým. Jedná se o budovy, pozemky, auta, počítače a další majetek ve vlastnictví úřadů a státních firem, který se tyto instituce snaží prodat, a to většinou za zlomek pořizovací ceny. K prodeji se běžně využívají různé aukce, přímé prodeje nebo veřejné výběrové řízení, o kterých se však běžní občané jen stěží včas dozvědí, a proto v důsledku nízké poptávky po tomto majetku zůstávají ceny velmi nízko. Státní instituce sice mají povinnost zveřejňovat informace o těchto prodejkách, avšak pro potenciální zájemce je nadměru složité a zdoluhavé procházet jednotlivé webové stránky konkrétních institucí a hledat na nich aktuální nabídku. Tomu by měla pomoci právě webová aplikace Váš majetek.cz, která všechna data o aukcích a prodejkách shromažďuje na jednom místě, kde zájemce může zadat kritéria pro vyhledání konkrétní věci a má zde dostupné všechny potřebné informace, včetně data a času konání akce, místa a také vyvolávací ceny. Díky tomu se na aukce dostaví více zájemců, cena se tak může vyšplhat mnohem výš, často i na úroveň tržní ceny, a tím se získá do veřejné pokladny více peněz [10].

2.1.3 Mapa exekucí.cz

Dalším velmi zajímavým počinem organizace Otevřená společnost, o.p.s. je projekt Mapa exekucí, který prostřednictvím webové aplikace zpřístupňuje data o exekucích v ČR. Cílem tohoto projektu není jen zpřístupnit informace o exekucích, ale také pomocí různých studií a výzkumů, které porovnávají způsoby oddlužení osob v jednotlivých státech EU, vyvrátit některé mylné představy a mýty a zadlužených lidech. Zejména pak představy o důvodech zadlužení.

Data zde využitá pochází s Centrální evidence exekucí, kterou má zákonnou povinnost spravovat Exekutorská komora ČR. Do Centrální evidence exekucí se podle §2 vyhlášky č. 329/2008 Sb. zapisují informace o vymáhané povinnosti fyzických a právnických osob včetně údajů o jejich místě bydliště. Díky tomu mohla vzniknout tato mapa, ze které lze snadno zjistit, v jakém kraji nebo konkrétní obci žije nejvíce obyvatel zatížených exekucí, dále například jaký je podíl obyvatel v exekuci k celkovému počtu obyvatel na daném území, nebo jaká je průměrná jistina na jednu exekuci v dané oblasti.

Problematika zadluženosti obyvatel ČR je zařazena do vládní Strategie boje proti sociálnímu vyloučení 2016 – 2020 a díky této interaktivní mapě, je možné zjistit aktuální stav zadluženosti

nejen v celorepublikovém měřítku, ale i na úrovni krajů a obcí. Díky tomu lze přesněji zacílit opatření na místní úrovni například v podobě různých dluhových poraden, či posílení terénní sociální práce v konkrétní oblasti a tím alespoň trochu zadlužování předcházet nebo zmírnit [19].

2.1.4 Návrh na další možné využití otevřených dat

Existuje mnoho možností, jak otevřená data využít ať už pro komerční nebo soukromé účely. Jednou z nich je například využití dat z rejstříku osob se zákazem plnění veřejných zakázek, propojit ho s daty z rejstříku nespolehlivých plátců a zároveň s Obchodním rejstříkem a Živnostenským rejstříkem a získat tak detailní data o potencionálních obchodních partnerech či klientech. Vše by mohlo být vizualizováno například ve webové aplikaci, kde by bylo možné poměrně snadno a rychle získat všechny dostupné informace a pomoci tak podnikatelům při rozhodování o budoucích zakázkách apod.

Velmi podobně funguje například webová aplikace Firmo.cz, která shromažďuje informace o firmách i fyzických osobách, jejich aktivitách a upozorňuje na nové obchodní příležitosti nebo naopak hrozby. Také monitoruje změny v obchodním rejstříku, upozorňuje na probíhající insolvence, dlužníky či na poptávku nebo nabídku zaměstnanců. Díky této aplikaci je tak možné sledovat změny u dodavatelů, konkurence či obchodních partnerů [6]

3 METODIKA PUBLIKACE OTEVŘENÝCH DAT VEŘEJNÉ SPRÁVY

Metodika publikace otevřených dat představuje ucelený soubor doporučení, kterými je vhodné se řídit při publikaci otevřených dat VS. V první části jsou definovány základní pojmy, týkající se otevřených dat a poté se práce věnuje formátům dat, které jsou dle své kvality členěny do jednotlivých stupňů otevřenosti. Poté již následuje samotný popis procesu publikace otevřených dat krok po kroku až po jejich zveřejnění [5].

3.1 Datová sada

Termín datová sada, neboli častěji používané označení dataset, je množina dat, které spolu určitým způsobem souvisí a jsou proto zveřejňovány jako jeden nezměněný celek. Jedná se o největší jednotku v datové paměti. Tvoří ji jednotky dat, popřípadě záznamy se shodnými nebo slučitelnými formáty. Nejčastěji dataset obsahuje jednu tabulku či statistickou datovou matici, ale může jich obsahovat i více. Řádky těchto tabulek či matic, reprezentují jednotlivé položky a určují tak celkový počet záznamů v souboru, zatímco sloupce označují hodnoty jedné proměnné. Datová sada tedy může mít libovolný počet řádků a sloupců, přičemž v řádcích by měly být jednotlivé položky a ve sloupcích atributy s nimi související. Dále by jednotlivé položky měly mít stejný význam. Jako příklad slouží následující tabulka:

Tabulka 1: Příklad správně sestavené tabulky pro datové sady.

ID	Jméno	Příjmení	Město
1001	Karel	Novák	Pardubice
1002	Jan	Procházka	Hradec Králové
1003	Martina	Šrámková	Brno

Zdroj: vlastní zpracování

Takto sestavená tabulka ve správně zvoleném formátu, ideálně XML nebo CSV, lze označit za kvalitní zdroj otevřených dat. Pokud je v datové sadě více tabulek, které spolu souvisí, je nutné, aby byly vždy propojeny totožným atributem. V případě této tabulky by nejlépe posloužil první sloupec s označením ID, který by případně další doplňující informace z jiných tabulek propojil se správnými položkami [9].

3.2 Datový katalog

Datový katalog otevřených dat je primárním zdrojem datových sad. V rámci katalogu však nejsou uložena přímo samotná data nebo datové sady, ale jsou zde ukládány informace o jednotlivých datových sadách, jako informace o jejich umístění, o správci datové sady a dále

například o jejich validitě či aktuálnosti a další související skutečnosti. Jedná se tedy o strukturovaný web, který plní funkci rozcestníku na stránky jednotlivých datasetů. Katalogizace datových sad tak slouží k snadnějšímu vyhledávání a větší přehlednosti v otevřených datech. Katalogů existuje celá řada od těch menších, které si spravují jednotlivé instituce pro svá data, až po ty centrální, které sdružují všechny datové sady.

Centrální datové katalogy má v dnešní době již mnoho zemí a za ty nejlepší jsou považovány katalogy ze zemí jako například Německo, Velké Británie nebo Švýcarsko. Katalogy provozují také globální instituce jako OECD, OSN nebo Světová banka. V ČR existuje Národní katalog otevřených dat, jehož správcem je Ministerstvo vnitra ČR. Tento katalog vznikl v rámci Akčního plánu Partnerství pro otevřené vládnutí (OGP) a sdružuje menší katalogy jednotlivých státních institucí. Nalezneme zde tedy katalogizovaná data z ministerstev, ČSÚ, krajských a městských úřadů atd. [10].

3.3 Stupně otevřenosti dat

Již zmiňovaný vynálezce webu Sir Tim Berners-Lee, definoval v roce 2006 model pětiúrovňové stupnice, ve které je každý stupeň označen hvězdami od jedné do pěti. Počet hvězd označuje stupeň otevřenosti dat, přičemž za otevřená, už se považují data i s jednou hvězdou. Podle tohoto modelu musí být ve všech stupních zajištěna právní otevřenost, dále se pak jednotlivé stupně liší podle formátu, ve kterém jsou data publikována. Pětistupňový model je graficky znázorněn na Obrázku č.1 a dále budou jednotlivé stupně otevřenosti popsány.



Obrázek 1: Stupně otevřenosti dat podle „5-star deployment scheme for Open Data“

Zdroj: zpracováno dle [2]

***První stupeň** otevřenosti dat je omezen pouze požadavkem na právní otevřenost. Jedná se o zveřejnění dat takovým způsobem, aby jejich podmínky užití (licence) nebyly v rozporu s Definicí otevřenosti (Open Knowledge, 2014) a uživatelé je tak mohli využívat pro libovolné účely. Zvolený formát je na tomto stupni libovolný, takže se může jednat například o naskenované dokumenty, či dokumenty ve formátu PDF nicméně by v každém případě měly být zveřejněny online. Výhodou zveřejňování dat v rámci prvního stupně otevřenosti je zejména jednoduchost a relativně nízká pracnost na zveřejnění pro poskytovatele dat, nicméně o to obtížnější může být jejich využití pro potencionální uživatele.

****Druhý stupeň** musí mimo právní otevřenost dat splňovat ještě podmínku strojové čitelnosti. Data musí být tedy čitelná nejen pro člověka, ale i pro počítač, který však neumí zpracovávat data ve formě obrázků a je třeba mu je poskytnout ve formě znaků. Na této úrovni je ještě zachována možnost využití proprietárního formátu, tedy nějakým způsobem uzavřeného či speciálního. Nevýhodou proprietárního formátu může být neexistence volně dostupného nástroje, který by ho dokázal zpracovat a uživatel je pak nucen si odpovídající software pořídit. Typickým příkladem může být tabulka z programu MS Excel ve formátu XLS. Stejně jako u prvního stupně platí také podmínka online dostupnosti dat, která bude provázet i všechny další úrovně.

*****Třetí stupeň** již využívá otevřené formáty dat, které jsou snadno zpracovatelné nezávisle na konkrétním použitém nástroji. Jednoduše řečeno jde o to, aby dataset mohl použít uživatel s běžně dostupným softwarem. V tomto případě se může jednat například o formát XLSX, přestože pochází ze stejného programu jako uzavřený formát XLS, je možno s ním pracovat pomocí alternativních programů jako jsou LibreOffice či OpenOffice a je možné ho tedy označit za otevřený. Dále se může jednat o formáty CSV, XML nebo JSON. Tříhvězdičkový stupeň považujeme za standard pro data veřejné správy/samosprávy ČR.

******Čtvrtý stupeň** je v českém prostředí považován spíše za nadstandard. Zásadní podmínkou je využití univerzálního identifikátoru (URI) nad jednotlivými sloupci dat. Jedná se v podstatě o přiřazení vlastnosti ke každému datovému objektu, díky nimž se stávají data propojitelná (viz. kapitola 1.2). Lze se na ně tedy odkazovat, obdobně jako na webové stránky, nicméně samotná data neobsahují odkazy na jiné datové zdroje, takže ještě stále nejsou propojená. Technologické řešení takto strojově srozumitelné struktury dat vyžaduje využití RDF systému (Resource Description Framework), neboli systém popisu zdrojů.

*******Pátý stupeň** obsahuje již propojená data, která nejenže využívají URI pro identifikaci objektů, ale také obsahují odkazy na jiná data, které s nimi souvisí a tím jsou zasazena do

kontextu. Propojení spočívá v tom, že je alespoň u dvou datasetů určeno, která data mají společná. To je velmi výhodné například pro jednotlivé orgány Veřejné správy, kterým stačí udržovat si své vlastní datové sady a na ostatní je možné se pouze odkazovat, čímž se snižuje potřeba duplikace jednotlivých dat na více místech. I zde se využívá systém RDF s propojením na další zdroje [10].

Tabulka 2: Shrnutí všech požadavků, které musí splňovat data v jednotlivých stupních otevřenosti.

Stupeň otevřenosti	Požadavky na otevřená data				
	Právní otevřenost	Strojová čitelnost	Otevřenost formátu	URI (univerzální identifikátory)	Propojení na jiná data
*	Ano	Ne	Ne	Ne	Ne
**	Ano	Ano	Ne	Ne	Ne
***	Ano	Ano	Ano	Ne	Ne
****	Ano	Ano	Ano	Ano	Ne
*****	Ano	Ano	Ano	Ano	Ano

Zdroj: zpracováno dle [2]

3.4 Proces publikace otevřených dat

Podle *Metodiky publikace otevřených dat Veřejné správy České republiky* je prvním krokem při publikaci dat **analýza a výběr dat k uveřejnění**. Analyzovat můžeme jakékoliv databáze, které nejsou v režimu ochrany dat a je tedy možné je volně publikovat. Vhodná data pro publikaci jsou například údaje o veřejných zakázkách, veřejných financích, územního plánování, podkladové dokumenty k rozhodnutím úřadů a samospráv, zajímavá jsou ale i geografická data, která lze vizualizovat na mapě, data o výsledcích voleb, sčítání lidu, stavu veřejných služeb atd.

Dalším krokem je **výběr vhodného formátu dat**. Jako zcela nevhodné formáty se považují některé softwarové formáty typu PDF, JPG, GIF či PNG, různé skenované dokumenty nebo fotky dokumentů. Tyto formáty jsou pro další zpracování zcela nevhodné a nevyužitelné. Protože by data měla být zveřejněna ve strojově čitelné podobě tedy v databázi, jsou nejvhodnější strukturované dokumenty ve formě tabulek CSV nebo pomocí vlastní definice organizace dat XML. Pro zveřejňování jednodušších numerických či kombinovaných souborů dat však postačí i tabulka z MS Excel ve formátu XLS a pro textová data je dobrým základem formát HTML.

Co se týče **určení způsobu přístupu k datům**, zásadní je rozhodnout, zda budou data zpřístupněna v podobě stažitelných souborů nebo pomocí webových služeb. Jestliže se rozhodneme zpřístupnit data pomocí stažitelných souborů, musíme určit, jakou zvolíme formu jejich aktualizace. Jestliže se budou aktualizovat celé soubory dat, tedy nejen nově vzniklá data

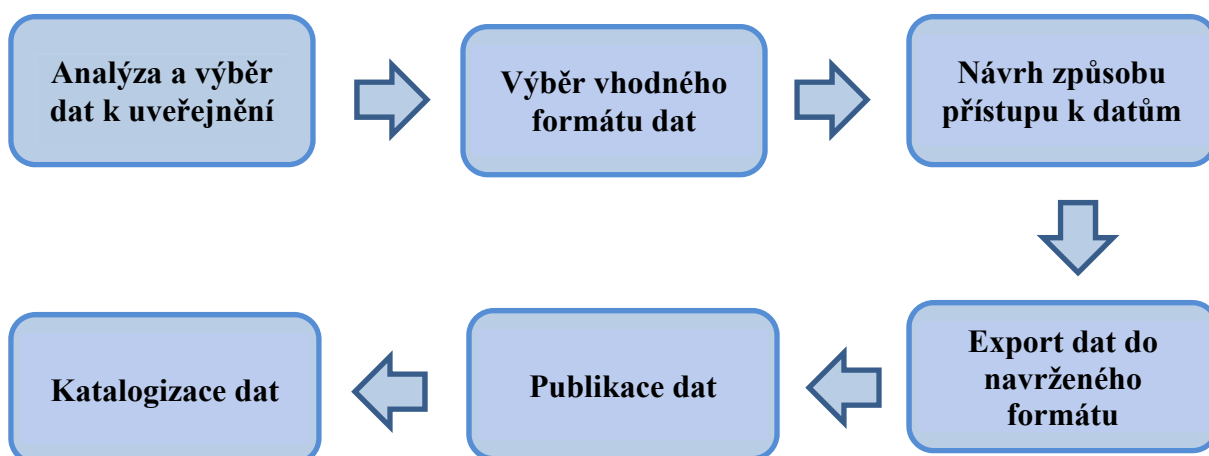
ale i ta stávající, je vhodné poskytnout uživatelům v určených časových intervalech nové soubory ke stažení namísto těch původních. To znamená, že uživatel vždy získá nejaktuálnější data z posledního exportu, ve kterých nejenže budou změněna data původní, ale i připsána data nová. Druhá situace, která může u souborů ke stažení nastat, je taková, že se budou přibývat pouze přírůstky dat, tedy že původní soubory zůstávají vždy stejné a není nutné je měnit. V této situaci je spíše vhodné v pravidelných časových intervalech zveřejňovat nová data a změny. V případě, že se data velmi dynamicky mění, je nutné k nim zajistit aktuální online přístup do databáze pomocí vhodného dotazovacího rozhraní ve formě webové služby.

Další krok je spojen především s technickým zajištěním daného **exportu do navrženého formátu**. Jestliže je formát zvolen, je nutné vytvořit skripty nebo nakonfigurovat nástroje pro transformaci dat, které do něj exportují data z relační databáze. V závislosti na typu formátu se tento postup mění.

V následujícím kroku dochází k samotné **publikaci dat**. V rámci toho je nutné zvolit vhodnou webovou prezentaci dat a také URL, na které budou data dostupná uživatelům. Ideálně by data měla být publikována na jedné samostatné stránce určené k publikaci otevřených dat, na kterou bude vést viditelný odkaz ze strany úvodní a také zde budou uvedeny podmínky jejich užití, nebo zvolená licence, jenž se těchto dat týká.

Poslední fází je **katalogizace dat**. V této konečné fázi se vytváří záznam o zveřejněných otevřených datech v datovém katalogu, aby byla pro potenciální uživatele lépe dohledatelná [5]. (viz. kapitola 3. odst.2)

Na následujícím obrázku jsou jednotlivé kroky graficky znázorněny.



Obrázek 2: Proces publikace otevřených dat ve veřejné správě.

Zdroj: zpracováno dle [5]

4 KONKRÉTNÍ PŘÍKLADY DOBŘE A ŠPATNĚ ZVEŘEJNĚNÝCH DAT

V současné době již mnoho veřejných institucí a úřadů zveřejňuje svá data. V některých případech jsou zveřejněna správně, ale mnohdy tomu tak není. V této práci je nejprve uveden příklad špatně zveřejňovaných dat firmou CHAPS spol. s.r.o., a dále pak příklad zveřejňování dat správným způsobem prostřednictvím Registru smluv.

4.1 Zveřejnění jízdních řádů veřejné dopravy

Dne 1. září 2014 nabyla účinnosti novela vyhlášky o jízdních řádech veřejné linkové dopravy (122/2014 Sb.), konkrétně se však zaměříme na aktualizaci § 9 odstavce 8, která pojednává o organizaci celostátního informačního systému o jízdních řádech. V tomto odstavci se píše:

*„Jízdní řády jsou neprodleně po jejich postoupení do celostátního informačního systému o jízdních řádech zveřejňovány způsobem umožňujícím dálkový přístup. Jízdní řády podle odstavců 4 a 5 jsou uveřejňovány v podobě podle § 6 odst. 3 **způsobem umožňujícím automatizované zpracování** a současně v podobě podle § 3 až 5. Jízdní řády podle odstavce 6 jsou uveřejňovány v podobě, ve které byly do celostátního informačního systému o jízdních řádech postoupeny ministerstvem.“* [12]

Tato novela přikazuje povinnost zveřejňovat data o jízdních řádech takovým způsobem, aby bylo možné jejich automatizované zpracování. Celostátní informační systém o jízdních řádech veřejné linkové osobní dopravy má na základě pověření Ministerstva dopravy České republiky na starost firma CHAPS spol. s.r.o. Tato firma se zabývá vývojem IT aplikací a systémů a také vytvářením komplexních softwarových řešení zejména pro oblast osobní dopravy, působí na trhu již od roku 1993, kdy pro společnost České dráhy navrhla a vyvinula první Elektronické jízdní řády nyní známé jako IDOS [17]. Podle ÚOHS však firma CHAPS dlouhodobě zneužívala své dominantní postavení na trhu, protože od roku 2008 až do srpna roku 2015 odmítala data o jízdních řádech zpřístupňovat ostatním uživatelům v potřebném formátu, v jakém byly vytvořeny. Z toho důvodu firma CHAPS léta vedla soudní spory s několika dalšími firmami, které chtěly data o jízdních řádech využít ve svém podnikání. Jedním z největších a nejznámějších sporů se vedl mezi firmou CHAPS s.r.o. a Seznam.cz obchod, s.r.o., který chtěl data použít ve svých webových mapách [3].

Zmiňovanou novelou měl monopol této firmy skončit, nicméně tomu tak nebylo a výsledkem je v podstatě horší řešení, než tomu bylo doposud, nicméně v souladu s platnými vyhláškami. Problém nastal totiž ve chvíli, kdy se ukázalo, že zmiňovaná vyhláška se týká jen

autobusových jízdních řádů nikoliv těch vlakových. Vlakovými se zabývá vyhláška 173/1995 Sb., do jejíž novely, která je také platná od 1. září 2016 se příslušné ustanovení nedostalo. Díky tomu může firma CHAPS zveřejňovat data o vlakových jízdních řádech v takovém formátu, v jakém mu je předává provozovatel dráhy, ten však nemusí být strojově čitelný. To tedy v praxi znamenalo, že u autobusových dat se vývojáři dočkali několika textových souborů ve formátu CSV, které se dají strojově zpracovat, ale co se týče jízdních řádů vlaků, ty firma poskytla ve formátu PDF, jehož zpracování je pracné, nákladné a také velmi nespolehlivé. Stejně tak jsou ve formátu PDF zveřejněny i mezinárodní autobusové linky, na které se bohužel novelizace také nevztahuje. Je tedy více než jasné, že firma CHAPS se snaží zveřejňovat data tak, aby se nedostala do sporu s legislativou, ale zároveň sama neučiní jediný vstřícný krok k jejich otevření, pokud k tomu nebude donucena jiným způsobem. Dokud se tedy neprovede novelizace vyhlášky o vlakových jízdních řádech, nedá se očekávat žádná pozitivní změna ze strany firmy CHAPS a důvod k tomu je velmi jednoduchý, totiž jediný, kdo získává všechna data o vlakových i autobusových linkách ve strojově čitelném formátu je vydavatelství Mafra, které provozuje již zmiňovaný portál IDOS a je tedy zřejmé, že jejich spolupráce je oboustranně výhodná. Tím, že firma CHAPS neposkytne data ve stejném formátu i ostatním firmám pak vzniká konkurenční výhoda pro vydavatelství Mafra. Toto je bohužel příklad toho, jak může soukromá firma vydělávat na státních zakázkách [17].

4.2 Zveřejňování dat dle zákona o registru smluv

Jako příklad dobře zveřejňovaných otevřených dat lze v ČR uvést zveřejňování soukromoprávních smluv podle zákona č.340/2015 Sb., ze dne 24. listopadu 2015. Tento zákon vstoupil v účinnost dne 1. července 2016 a udává povinnost subjektům uvedeným v zákonu zveřejňovat smlouvy, jejichž peněžní plnění převyšuje nebo se rovná částce 50 000 Kč bez DPH, v elektronické podobě ve veřejném centrálním registru, který spravuje Ministerstvo vnitra. Mezi subjekty, kterých se zákon týká, patří státní úřady, kraje, větší obce (výjimku dostaly obce 1. a 2. stupně bez úřadu s rozšířenou působností), dále státní příspěvkové organizace a fondy, veřejné výzkumné instituce, veřejné vysoké školy a také firmy, ve kterých má stát, kraj nebo obec většinový podíl. Výjimku zde tvoří pouze ty firmy, ve kterých sice vlastní většinový podíl stát, kraj nebo obec, nicméně jeho akcie jsou obchodovatelné na burze a z toho důvodu není vhodné jejich smlouvy zveřejňovat.

Dále se měl zákon týkat také všech státních a národních podniků, avšak už v době, kdy zákon vstupoval v platnost, měla poslanecká sněmovna přichystaný návrh na jeho změnu prostřednictvím novely. Novela se měla týkat pouze vyjmutí národního podniku Budějovický

Budvar z působnosti zákona, a to z toho důvodu, že tento podnik se pohybuje ve vysoce konkurenčním prostředí a zveřejňováním svých smluv by mohl ohrozit svou konkurenceschopnost a zhoršit své postavení na světovém trhu.

Původní novela zákona č. 340/2015 Sb., která se měla týkat pouze vyjmutí Budějovického Budvaru, se během legislativního procesu dočkala řady pozměňujících návrhů, které značně rozšířily výčet výjimek ze zákona. Podle novely by se nakonec zveřejňování smluv vyhnuly i státní podniky s většinovou účastí státu, krajů nebo obcí, které mají obchodní nebo průmyslovou povahu. Dále Český rozhlas a Česká televize, které by sice pod zákon spadali, ale nebyli by povinni poskytovat veřejně smlouvy týkající se výroby a vysílání sportovních pořadů včetně nákupu práv, a dohody o nákupu práv převzatých pořadů. Další výjimky se týkají i Veřejných vysokých škol, které by podle novely nebyly povinné poskytovat smlouvy uzavřené se soukromými podnikateli [11].

Za předpokladu, že však v současnosti subjekt splňuje podmínky pro vkládání smluv do registru, má zákonem jasně definovaná pravidla. Dokumenty musí být poskytnuty v otevřeném, strojově čitelném formátu, v předem stanoveném rozsahu a včetně metadat⁶. Dále jsou zákonem stanovené lhůty pro zveřejnění, a to bez zbytečného odkladu, avšak nejdéle do 30 dnů od uzavření smlouvy. Během prvního roku účinnosti zákona je možné tuto lhůtu běžně využívat, nicméně po prvním roce jeho účinnosti, tedy od 1. července 2017 bude platit podmínka, že smlouva, která má být podle zákona uveřejněna, bude účinná ode dne uveřejnění v registru smluv. To v praxi znamená, že smlouvy, které nebudou zveřejněny ve stanovené lhůtě, nebudou platné a pokud nedojde k jejich zveřejnění ani do 3 měsíců od jejich uzavření, budou považovány za zrušené. To samé platí i u všech dodatků a změn, které by souviseli s danou smlouvou. Při samotném vkládání smluv do registru mají povinné subjekty dle zákona ve stanovené lhůtě zaslat příslušný elektronický formulář (viz. Příloha A), společně s příloženou kopií smlouvy ve stanoveném formátu, datovou schránkou zřízenou správcem registru smluv. Ten ji pak bezodkladně zveřejní na portálu veřejné správy včetně všech metadat. V registru smluv si může subjekt zvolit z několika možností, jakou operaci chce provést. Nejčastěji se jedná o klasické zveřejnění záznamu, tedy smlouvy, která do té doby v registru ještě nebyla vložena. Dále pak může vkládat jednotlivé přílohy, k již zveřejněným smlouvám, nebo zvolit možnost modifikace záznamu a zveřejněnou smlouvu upravit. Poslední možností je zneprístupnění záznamu, kde stačí zadat pouze identifikátor smlouvy, a smlouva bude

⁶ Metadata – Jsou data, která poskytují informace o jiných datech.

znepřístupněna. Dle zákona musí být při zveřejňování smlouvy ve formuláři povinně vyplněny tato metadata:

- identifikace smluvních stran
- vymezení předmětu smlouvy
- cena, a pokud ji smlouva neobsahuje, hodnotu předmětu smlouvy, lze-li ji určit
- datum uzavření smlouvy

Tyto čtyři informace jsou nezbytné pro další využití dat. Díky nim má potenciální uživatel možnost data automatizovaně zpracovávat, třídít je a sdružovat, aniž by musel jednotlivé smlouvy procházet a údaje v nich dohledávat. V případě, že tato metadata nebudou vyplněna, považuje se takto zveřejněná smlouva od 1.7.2017 za nepublikovanou, a tedy podle zákona za neplatnou. Další nepovinné položky, které mohou subjekty zveřejnit, jsou například:

- identifikační číslo smluvní strany
- datová schránka smluvní strany
- adresa smluvní strany
- zdali je smluvní strana plátce nebo příjemce peněz
- číslo smlouvy
- podepisující osoba
- měna

Publikující strana má možnost některá z těchto dat z důvodu ochrany obchodního tajemství nezveřejnit. Jedná se konkrétně o identifikaci smluvních stran, vymezení předmětu smlouvy a uvedení ceny, popřípadě hodnoty smlouvy. Takto nezveřejněné informace však musí objektivně splňovat znaky obchodního tajemství podle Občanského zákoníku (§ 504) a dále musí být tyto informace začerněny i ve zveřejněné smlouvě [16].

Z pohledu otevřených dat, se dají tyto zákonné požadavky na lhůty, formáty a povinně zveřejňovaná metadata ke smlouvám považovat za dostačující pro automatizovanou práci s těmito daty a z toho důvodu se jimi bude práce dále zabývat v následující kapitole.

5 ANALÝZA DAT Z OBLASTI VEŘEJNÉ SPRÁVY ZA POUŽITÍ ZVOLENÉHO NÁSTROJE

V rámci bakalářské práce bude provedena analýza dat z Registru smluv. Do tohoto registru mají dle zákona č.340/2015 Sb. Zákon o registru smluv od 1. července 2016, povinnost vkládat zákonem stanovené subjekty veškeré soukromoprávní smlouvy, které byly uzavřeny od tohoto data. Veškerá problematika týkající se Registru smluv, byla popsána v kapitole 4.2.

5.1 Otevřená data z Registru smluv

Data z Registru smluv, lze získat na webovém portálu <https://smlouvy.gov.cz/> zřízeným Ministerstvem vnitra. Na stránce s názvem Otevřená data, se nachází odkaz <https://data.smlouvy.gov.cz/>, pod kterým jsou vypsané obsahy adresářů všech aktuálně vygenerovaných souborů. Jednotlivé měsíce jsou členěny do tzv. dump souborů, které mají vždy označení: `dump_<ROK>_<MĚSÍC>.xml`

Každý den se jednotlivé dump soubory aktualizují, aby se doplnily o nově přidaná data od začátku měsíce. Detailní údaje, o konkrétních souborech jsou obsaženy v indexovém souboru s názvem **index.xml**, který je umístěn v adresáři spolu s ostatními soubory. Na obrázku níže je příklad informací, které jsou dostupné o jednom konkrétním měsíci v adresáři, tedy o jednom dump souboru.

```
<dump>
  <mesic>2</mesic>
  <rok>2017</rok>
  <hashDumpu algoritmus="sha1">31ea7c9d769a745356126f0896f9f3f4d77ee23f</hashDumpu>
  <velikostDumpu>64906967</velikostDumpu>
  <casGenerovani>2017-04-25T00:18:59+02:00</casGenerovani>
  <dokoncenyMesic>1</dokoncenyMesic>
  <odkaz>https://data.smlouvy.gov.cz/dump_2017_02.xml</odkaz>
</dump>
```

Obrázek 3: Údaje z adresáře o jednom dump souboru.

Zdroj: [14]

Význam elementů z Obrázku 3 může být vysvětlen následovně:

- **mesic** - označuje číslo měsíce, ze kterého data pochází
- **rok** - označuje rok souboru, ve kterém byla data vytvořena
- **hashDumpu** – kontrolní algoritmus, který slouží k převodu vstupních dat
- **velikostDumpu** - znamená velikost daného souboru v bytech

- **casGenerovani** – datum a čas vygenerování daného souboru
- **dokoncenyMesic** – zde může být hodnota 1 nebo 0 (1= dokončený měsíc, ve kterém už nejsou očekávány další změny, 0= stále přibývají nová data za daný měsíc)
- **odkaz** - URL souboru ve formátu XML, tato URL pak také slouží k nahrání dat z Registru smluv do zvoleného nástroje

5.2 Příprava dat

Pro vizualizaci dat z veřejného Registru smluv, jsem zvolila analytický nástroj Qlik Sense. Jedná se o produkt firmy QlikTech, jehož licenci pro verzi Qlik Sense Desktop lze zdarma stáhnout na internetu jak pro osobní, tak pro komerční využití. Plná verze produktu nabízí možnost práce s daty nejrůznějších formátů jako například CSV, TXT, XML a další...

Data z Registru smluv jsou ve formátu XML a jejich příprava probíhá kompletně v Qlik Sense Desktop, protože data z XML se nedají analyticky zpracovat a je potřeba je transformovat do struktury analogické k relačním databázím. Příprava dat obnáší tyto kroky:

- příprava Qlik Sense asociativního datového modelu – mapování XML elementů do tabulek modelu (úprava názvů jednotlivých atributů pro lepší přehlednost) a vytvoření datového modelu⁷ (viz. Příloha B)
- krok načtení seznamu měsíčních extraktů z adresáře dump souboru
- vytvoření cyklu pro načítání jednotlivých extraktů
- příprava specifických dimenzí a metrik

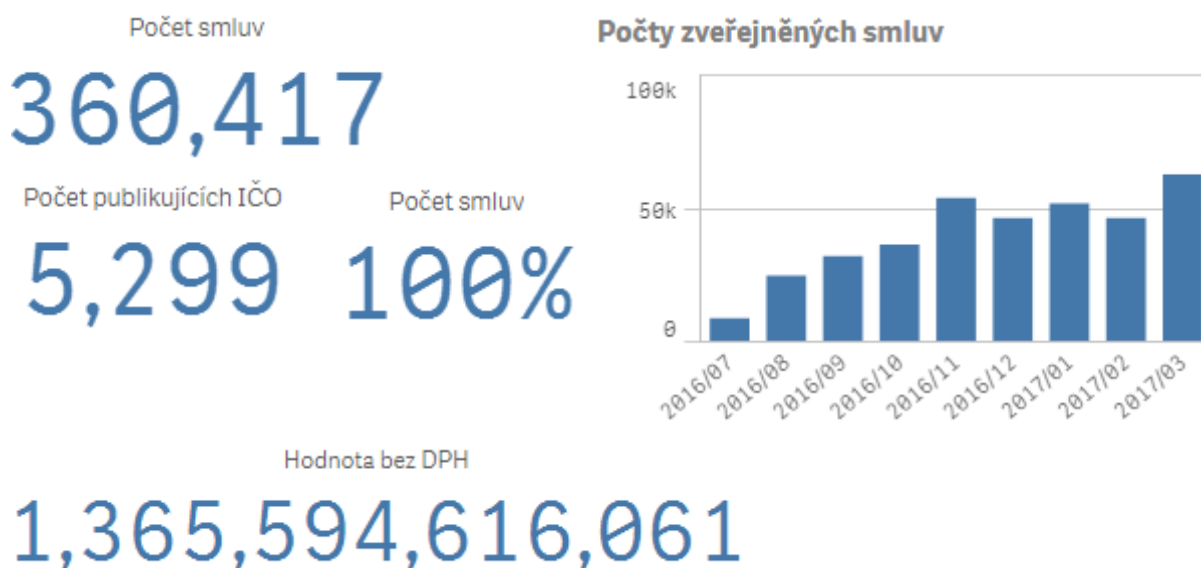
5.3 Vizualizace dat

Připravená data se poté dají vizualizovat v Qlik Sense Desktop, který rozpoznává jednotlivé atributy a položky, a umožňuje tak data vizualizovat pomocí širokého spektra grafů, tabulek a textových polí. Uživatel má možnost vytvářet logické a statistické filtry a pomocí matematických funkcí dosahovat přesných výpočtů.

Na následujícím Obrázku 4 je ukázka uživatelského prostředí zvoleného nástroje s nahranými daty z Registru smluv. V tomto konkrétním přehledu jsou vytvořeny některé základní metriky, znázorňující počet smluv v registru celkem, což také potvrzuje ukazatel počtu smluv v %, který tedy v tomto případě ukazuje 100 %, a dále počet publikujících subjektů

⁷ Datový model – Forma sdružení jednotlivých tabulek pomocí společných atributů.

a celkovou hodnotu všech smluv bez DPH v Kč. Dále se zde pak také nachází graf s počty zveřejněných smluv za jednotlivé měsíce.



Obrázek 4: Výstup z analytického software Qlik Sense Desktop s daty z Registru smluv.

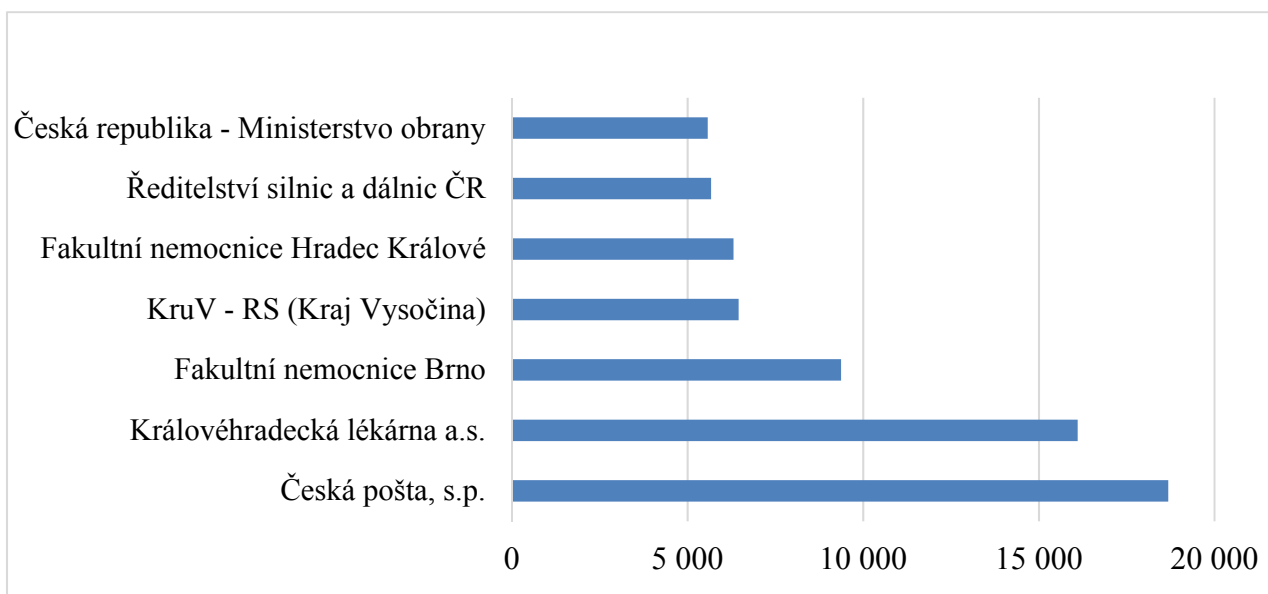
Zdroj: vlastní zpracování

Dle Obrázku 4 do Registru smluv za vybraných 9 měsíců, zveřejnilo svá data **5 299** subjektů, na které se tato povinnost vztahuje. Tato hodnota byla zjištěna pomocí součtu unikátních IČO, které se automaticky do dat zapisují prostřednictvím datové schránky při publikaci smluv.

Bylo zveřejněno celkem **360 417** platných smluv v celkové hodnotě necelých **1 366 mld. Kč**. V souboru dat je možné nalézt i mnoho záznamů (smluv), které mají ve sloupci „Platný záznam“ označení 0, což v těchto datech znamená, že je záznam neplatný. Jedná se o záznamy, které byly z nějakého důvodu modifikovány, avšak jejich původní verze automatiky zůstává nahrána v souboru dat. Všechna tyto neplatné záznamy byly v úvodu analýzy dat odfiltrovány, a práce se tedy bude vždy zabývat pouze těmi smlouvami, která jsou v registru zveřejněny jako platné.

Dále lze na grafu z Obrázku 4 vidět, jak se postupně po měsících zvyšoval počet zveřejněných smluv registru. Během července, tedy prvního měsíce platnosti zákona, bylo zveřejněno pouze **8 507** platných smluv. Tento počet se v dalších měsících postupně navyšoval a v březnu 2017 bylo zatím publikováno nejvíce smluv a to přesně **62 511**. Denně přibýlo za období červenec 2016 až březen 2017 v registru průměrně **1316** smluv za zhruba **5 mld. Kč**

Následující graf (Obrázek 5), znázorňuje sedm nejčastěji publikujících subjektů v Registru smluv. Je zřejmé, že nejvíce aktivním publikujícím subjektem je jednoznačně Česká Pošta, s.p., která za zvolené období zveřejnila celkem **18 683** smluv.



Obrázek 5: Výběr 7. nejčastěji publikujících subjektů za období 1. 7. 2016 – 31. 3. 2017

Zdroj: vlastní zpracování

Dále dle Obrázku 5 také velmi aktivně publikuje své smlouvy Královéhradecká lékárna a.s., která zveřejnila celkem **16 106** smluv. Na třetím místě ve zveřejňování smluv je potom Fakultní nemocnice Brno, která zveřejnila celkem **9 368** smluv.

Během prvotního průzkumu a při procházení jednotlivých záznamů dat bylo nalezeno mnoho nedostatků týkající se kvality dat. Tyto nedostatky byly natolik závažné a v takové míře, že by velmi pravděpodobně zapříčinily vznik nepravdivých výsledků statistik a informace z dat vyplývající by se nedaly označovat za důvěryhodné. Následující část práce se bude zabývat těmito nedostatky v datech z Registru smluv. V následující kapitole jsou jednotlivé chyby popsány a analyzovány jejich četnosti z celkového souboru dat.

5.4 Chyby v Registru smluv

5.4.1 Identifikace smluvních stran

Jak již bylo popsáno v kapitole 4.2, publikující má možnost nezveřejnit název smluvní strany, pokud se jedná o údaj, který lze dle zákona označit jako obchodní tajemství. Bohužel vkladatel nemá možnost zvolit ve formuláři informaci, že se jedná o smlouvu obsahující obchodní tajemství, a protože identifikace smluvních stran je ve formuláři nastavena jako povinná položka, zadavatelé pak vyplňují do kolonek formuláře různé znaky, či text vysvětlující

absenci dat. Tento jev lze v datech Registru smluv nalézt celkem u **19 694** záznamů. V **926**. případech publikující subjekty zvolily možnost napsat různé znaky jako pomlčky, tečky či znaky jako „XXX“. V dalších **18 768** případech je pak v názvu protistrany text zmiňující fakt, že se jedná o obchodní tajemství. Pokaždé je však tato informace zapsána jiným způsobem, a proto je velmi složité tyto záznamy odhalit. V horších případech pak nastával problém ve chvíli, kdy publikující nevěděl, co vyplnit do údajů o smluvní straně, a rozhodne se, že jako smluvní stranu vyplní údaje o sobě. Tím pak vznikají záznamy, kde je publikující i protistrana jeden a ten samý subjekt a tím se samozřejmě velice zkreslují výsledná data. Celkově je takto zveřejněno **13 563** smluv.

Na následujícím grafu (Obrázek 6) je znázorněna tato chybovost v procentech z celkového souboru dat.



Obrázek 6: Graf špatně zveřejněných smluvních stran.

Zdroj: vlastní zpracování

Celkem je tedy z pohledu identifikace smluvních stran **91 %** smluv zveřejněno správně. U **5 %** smluv není zveřejněna žádná smluvní strana nebo je zde uvedena informace o obchodním tajemství a u **4 %** je jako protistrana zapsán stejný subjekt jako publikující.

V některých případech je však vyplněn dvakrát stejný subjekt u jednoho záznamu, přestože se nejedná o obchodní tajemství. Například Státní podnik Povodí Ohře, takto zveřejnil celkem **630** smluv v celkové hodnotě více než **820 mil. Kč bez DPH**. Při namátkové kontrole

jednotlivých smluv pak bylo zjištěno, že se v tomto případě nejedná o obchodní tajemství a není tedy důvod název smluvní strany nezveřejnit.

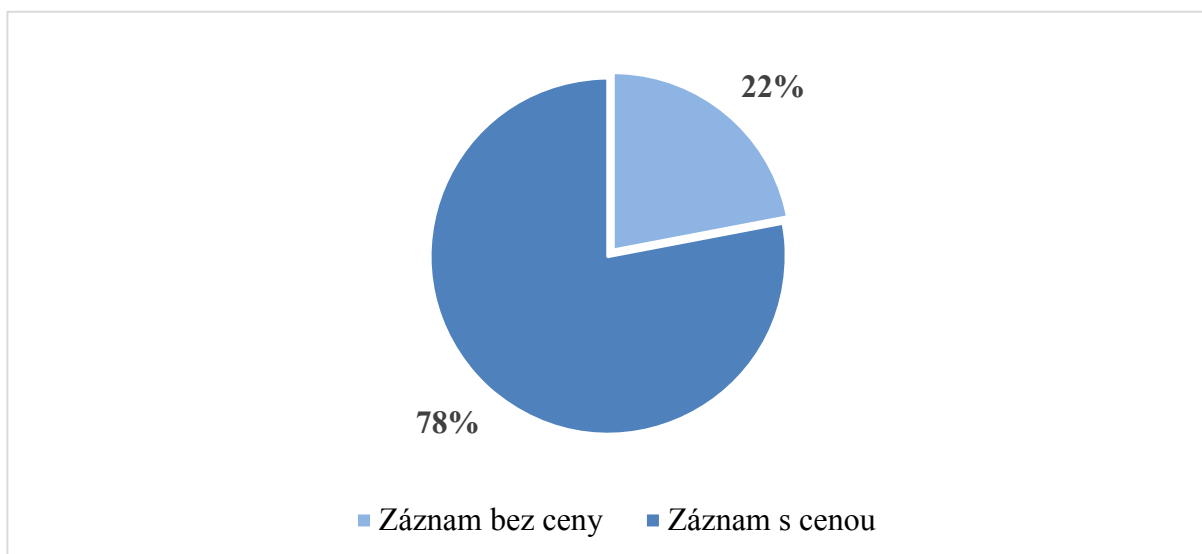
5.4.2 Vymezení předmětu smlouvy

Celkem u **307** smluv není uveden předmět smlouvy. Z tohoto počtu celkem **302** smluv bez předmětu smlouvy zveřejnil Dopravní podnik hl. města Prahy. Zvláštní je ale fakt, že tyto záznamy jinak obsahují informace o identifikaci smluvní strany včetně IČO a u **209** případů i cenu smlouvy bez DPH.

5.4.3 Chybějící nebo špatně zadaná cena

Jedním z velkých nedostatků Registru smluv je zadávání cen do formulářů. V zákoně je sice jasně definováno, že cena je jednou ze čtyř povinných položek, které musejí být vyplněny, avšak v reálných datech často tento údaj chybí. Hlavním důvodem by mohl být fakt, že organizace mají opět možnost cenu nezveřejnit a zároveň ji i začernit ve smlouvě, pokud se jedná o obchodní tajemství. Z toho důvodu je položka ceny ve formulářích při vyplňování nastavena jako nepovinná, a formulář se odešle i bez ní. Problém však nastává ve chvíli, kdy cenu nevyplní ani subjekt, který by ji ze zákona zveřejnit měl.

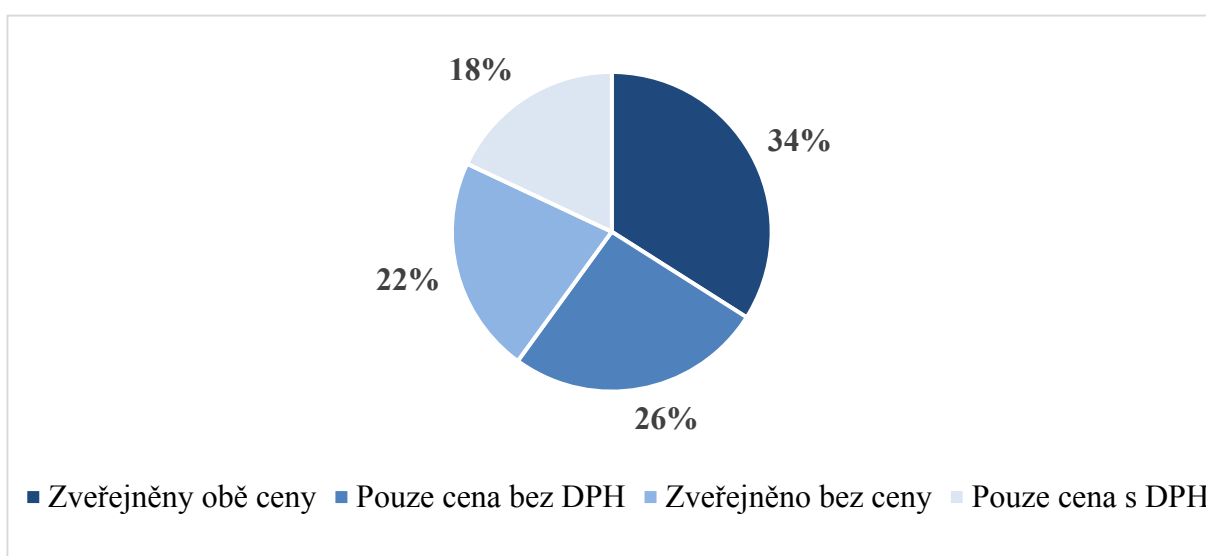
Na následujícím grafu (Obrázek 7) lze vidět, že bez ceny je zveřejněno celkem **78 843** smluv. Bohužel již není možné zjistit, kolik z nich je zveřejněno bez ceny z důvodu zachování obchodního tajemství, protože by to znamenalo otevřít všech **78 843** záznamů, a posoudit důvod nezveřejnění ceny individuálně nahlédnutím do konkrétní smlouvy.



Obrázek 7: Graf znázorňující procento záznamů zveřejněných bez ceny a alespoň s jednou cenou.

Zdroj: vlastní zpracování

Z celkového počtu **360 514** smluv jich je bez ceny zveřejněno celkem **78 843** což je celých **22 %**. Dalších 78 % smluv obsahuje v záznamu metadat alespoň jednu cenu. Nejvíce zveřejněných smluv bez uvedení částky má znovu Česká Pošta, s.p., která takto zveřejnila celkem **16 036** smluv. V některých případech je v metadatech sice vyplněná cena, ale ne vždy je vyplněna tak, jak by měla být. Někteří zadavatelé vkládají cenu v Kč chybně do políčka připraveného pro vložení ceny v cizí měně. Někteří zase nerozlišují, zdali píšou do formuláře cenu s DPH či bez DPH, jestli ji vkládají do správných polí a nezdědka je možné také nalézt chyby ve výpočtech cen ať už s DPH nebo bez a potom tedy částka v metadatech neodpovídá skutečnému stavu z příložené smlouvy. Tyto chyby jsou zaznamenány na následujícím grafu (Obrázek 8), kde jsou tyto četnosti uvedeny v procentech z celkového objemu dat.



Obrázek 8: Graf zobrazující chyby ve zveřejňování ceny v metadatech.

Zdroj: vlastní zpracování

Graf (Obrázek 8) ukazuje, že **34 %** smluv je zveřejněno s oběma cenami, **26 %** má zveřejněnou pouze cenu bez DPH, naopak pouze cena s DPH je uvedena u **18 %** a u **22 %** není cena zveřejněna vůbec.

Toto jsou chyby v registru smluv, týkající se povinných položek. Ve většině případů, by takto zveřejněná smlouva, byť obsahující jen jeden z uvedených nedostatků, byla považována podle zákona za špatně zveřejněnou, a tudíž za neplatnou. Dále se práce bude věnovat i ostatním položkám ve formuláři, které mají možnost publikující strany vyplnit. Zákon tyto informace nutně nevyžaduje zveřejňovat, nicméně z mého pohledu jsou některé z nich velmi důležité pro práci s celým souborem dat a skrývají v sobě obrovský potenciál a možnosti využití, pokud by byly zveřejněné.

5.4.4 Identifikační číslo smluvní strany

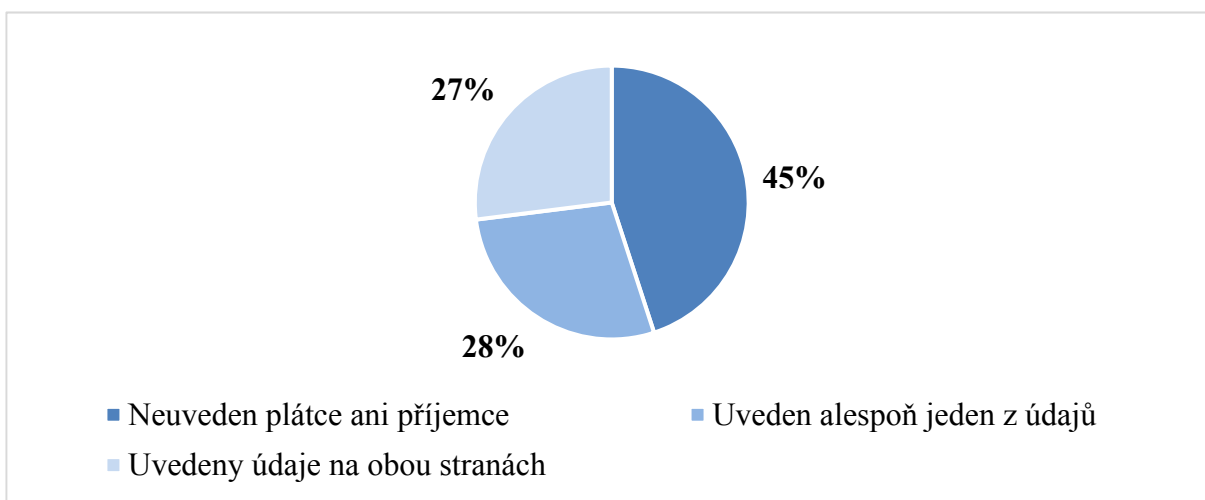
Podle mého názoru je jedním z nejdůležitějších údajů pro analýzu dat z Registru smluv identifikační číslo smluvní strany neboli IČO. Identifikační číslo publikující strany se do metadat automaticky propíše prostřednictvím datové schránky, ze které je formulář odeslán. Co se však týče vyplňování identifikačního čísla smluvní strany, ta je pro publikující stranu nepovinná a u každé 7. smlouvy byla skutečně tato informace vynechána to je celkem 15 % z celého objemu dat. Opět se v tomto případě nedá přesně určit, u kolika případů šlo o nezveřejnění z důvodu obchodního tajemství.

5.4.5 Chybné označení plátce a příjemce

Publikující subjekty mají možnost ve formulářích vybírat z rozevíracího seznamu možnost, zdali jsou oni plátcí nebo příjemci peněz. V ideálním případě zadavatel tento údaj vyplní jak na své straně, tak i za ostatní smluvní strany, aby bylo zřejmé, kdo komu peníze skutečně platil. V mnoha případech je u jednoho záznamu více smluvních stran, a pokud nejsou v metadatech jasně definovány role, vznikají pak nejasnosti ohledně této ceny. Ta je totiž do metadat zadávána vždy s plusovým znamínkem, přestože by publikující strana byla plátce těchto dat a měl by tedy na její straně vznikat závazek, a tedy záporná částka.

Z celkového počtu **360 417** zveřejněných záznamů je celkem **161 792** publikováno bez specifikace plátce a příjemce peněz. Celkem **103 457** záznamů má specifikovanou alespoň jednu ze stran, podle které pak lze logicky odvodit údaje o druhé straně, pokud se jedná o smlouvu pouze se dvěma smluvními stranami. Pouze u **95 641** záznamů jsou pak uvedeny údaje na obou stranách.

Na grafu (Obrázek 9) níže je popsán jev znázorněný v procentech.



Obrázek 9: Procentuální rozdělení záznamů podle toho, zda je u nich uveden údaj o plátcí či příjemci peněz.

Zdroj: vlastní zpracování

Z celkového počtu **360 471** smluv je celkem **27 %** zveřejněno správně s uvedenými údaji o směru platby peněz, u **28 %** je pak uveden alespoň jeden z těchto údajů a u celých **45 %** tyto informace chybí úplně. To bohužel v praxi znamená, že pro necelých **143 mld. Kč** není známo, kdo komu peníze platil.

5.4.6 Duplikace smluv

Problémy při analýza dat pak způsobují i duplicitní záznamy. Publikující subjekty mají povinnost, k již zveřejněným smlouvám přikládat i veškeré dodatky a přílohy, které se jich týkají. Pro tyto dokumenty totiž platí stejné podmínky pro zneplatnění jako u samotných smluv. Zadavatel tedy musí tyto související dokumenty bezodkladně zveřejnit v registru smluv tak, aby bylo zřejmé, k jaké smlouvě dokument patří. Zásadní chybu však zadavatelé dělají při zadávání ceny do formuláře přílohy, protože zde v mnoha případech píšou stejnou cenu, jaká je uvedena v původní smlouvě, a tím dochází k duplikaci jedné ceny do více dokumentů. Formálně jsou tyto dokumenty všechny platné a z pohledu analytického nástroje se ceny sčítají, ačkoliv jde stále o jednu a tu samou cenu.

5.4.7 Nedodržení lhůty od uzavření smlouvy k jejímu zveřejnění

Povinnost zveřejňovat smlouvy nejdéle do 30 dnů od jejich uzavření, je zákonem stanovena až od 1. července 2017. Do té doby je možné zveřejnit smlouvu i po vypršení této lhůty bez hrozících sankcí, které jsou v tomto případě zneplatnění smlouvy, popřípadě její úplné zrušení. Pokud by však zákon nebyl takto benevolentní a lhůta zveřejňování by byla kontrolována již během prvního roku účinnosti zákona, bylo by do konce března roku 2017 celkem **76 395** smluv zveřejněno pozdě. To je celých **21 %** z celkového objemu zveřejněných smluv.

5.5 Shrnutí kvality dat z Registru smluv

Ačkoliv se zpočátku zdálo, že by data měla být správně otevřená a vhodná pro analyzování, již po prvním bližším zkoumání jednotlivých záznamů bylo zřejmé, že jsou v datech jisté nedostatky a jejich špatná kvalita brání v dosažení důvěryhodných výsledků analýz.

Z celkového počtu **360 417** záznamů, obsahuje pouze **83 200** správnou identifikaci smluvních stran, zároveň také vyplněné cenu bez DPH i s DPH a určení kdo je plátce a kdo příjemce peněz. To znamená, že pouze **23 %** smluv je zveřejněno v takové kvalitě, která má skutečnou vypovídající hodnotu v rámci otevřených dat a má smysl je dále analyzovat. Pouze **8 %** smluv má pak zveřejněny všechny údaje včetně nepovinných.

Jedním z hlavních nedostatků Registru dat, je na základě výsledků analýzy těchto dat jednoznačně nedefinovaný postup při zveřejňování smluv obsahující obchodní tajemství. Tyto smlouvy jsou tedy publikovány stejně, jako všechny ostatní a nelze je pak od ostatních v celkovém souboru dat žádným způsobem odlišit. Někteří publikující píší informaci o tom, že je daný údaj neveřejný do předmětu smlouvy, jiní zase do názvu smluvní strany a někteří jednoduše nechávají pole prázdná úplně. Tyto záznamy je však ve velkém množství dat nereálné oddělit pomocí filtru dat, a proto musí uživatel pracovat i s těmito záznamy, ačkoliv neobsahují úplné informace a tím zkreslují celkové výsledky.

Dalším velkým problémem jsou neurčité vztahy mezi publikujícími a smluvními stranami. Ve chvíli, kdy nelze zjistit, kdo komu za danou transakci platí a kdo naopak peníze přijímá, ztrácí zveřejněná částka svůj smysl. Stejně je to pak i u špatné identifikace smluvních stran. Nelze-li určit, kdo je protistrana v daném vztahu, nebo je subjekt zapsán jako publikující a zároveň jako smluvní strana, opět zde dochází k znehodnocení výsledků datových analýz.

Zásadní je pak také absence ceny v záznamech. Ačkoliv u některých záznamů chybí z důvodu zachování obchodního tajemství, což je dle zákona v pořádku, není pak možné vytvářet statistiky například o celkových příjmech a výdajích za určité období u konkrétní instituce, protože tyto výsledky jsou nepřesné a neodpovídají skutečnosti.

ZÁVĚR

Cílem práce bylo charakterizování otevřených dat včetně jejich legislativního vymezení a také popis možností jejich využití na konkrétních příkladech. Dále se práce zabývala metodikou publikace otevřených dat z oblasti Veřejné správy a také samotným procesem jejich publikace. Dalším cílem pak bylo provedení analýzy otevřených dat, která byla aplikována na data z Registru smluv. Data, která lze získat z tohoto registru byla nejprve podrobně popsána a vysvětlena v kapitole 4.2 a dále potom následovala samotná analýza těchto dat pomocí zvoleného nástroje Qlik Sense Desktop.

Během prvotní analýzy bylo zjištěno, že za vybrané období 7/2016 – 3/2017 bylo v Registru smluv zveřejněno celkem 360 417 záznamů v celkové hodnotě necelých 1 366 mld. Kč. Dále pak byl na grafu znázorněn narůstající počet zveřejněných záznamů po jednotlivých měsících, kde jednoznačně nejvíce záznamů bylo uveřejněno v posledním měsíci vybraného období, tedy v březnu roku 2017. Tento narůstající počet zveřejněných záznamů během jednotlivých měsíců pravděpodobně souvisí s blížícím se termínem úplné účinnosti zákona, neboť ustanovení zákona o registru smluv, podle kterého bude účinnost smluv podmíněna právě jejich uveřejněním, nabývá účinnosti až po prvním roce platnosti zákona, tedy 1. července 2017.

V průběhu analýzy pak byly zjištěny jisté nedostatky týkající se kvality dat a tyto nedostatky byly poté podrobně popsány a analyzovány v další části práce (kapitola 5.4).

V práci bylo zjištěno, že hlavním nedostatkem dat uveřejněných prostřednictvím Registru smluv je jasně nedefinovaný postup při zveřejňování smluv obsahující obchodní tajemství a zároveň pak absence jejich kontroly ze strany správce smluv, tedy Ministerstva vnitra. Pokud by byla zavedena alespoň elementární kontrola základních metadat, zamezilo by se tím vzniku vysoké chybovosti v těchto datech. Chybně zveřejněná data totiž znemožňují kvalitní veřejnou kontrolu a stejně tak zamezují jejich využití potencionálním uživatelům.

Datová nekvalita, společně s již zmiňovanou plánovanou novelou zákona č. 340/2015 Sb., která by v případě schválení zapříčinila úbytek zveřejňovaných dat více než o polovinu, může snadno překazit původní záměr Registru smluv na zvýšení transparentnosti veřejné správy a možnost kontroly veřejných institucí občany ČR. Novelu zákona s pozměňovacími návrhy během dubna 2017 projednával senát, který ji však neschválil v plném rozsahu a vrátil zpět k projednání poslanecké sněmovně. Nové hlasování poslanců bude rozhodovat o konečné podobě Registru smluv.

POUŽITÁ LITERATURA

- [1] BERG, M. a BOČEK, J. *Otevřená data ve státní správě - nová éra rozhodování*. Praha: Nadace Open Society Fund Praha, 2012. ISBN 978-80-87110-24-9.
- [2] HAUSENBLAS, M.: 5 ★ *OPEN DATA* [online]. 2012 [cit. 2016-12-21]. Dostupné z: <http://5stardata.info/cs>
- [3] Chaps dostal pokutu 2,2 milionu. Dostatečně konkurenci nezpřístupňoval data o jízdách [online]. 7.1.2016. [cit. 2016-12-21]. Dostupné z: <http://byznys.ihned.cz/c1-65065830-chaps-dostal-pokutu-2-2-milionu-korun-nezpristupnil-nalezite-data-o-jizdnich-radech-konkurenci> [9]
- [4] CHLAPEK, D., KUČERA, J., NEČASKÝ, M.: *Koncepce katalogizace otevřených dat VS ČR (zkrácená verze)* [online]. Praha, září 2012. Citováno 10.11.2016. Dostupné z WWW: <http://www.korupce.cz/assets/partnerstvi-pro-otevrene-vladnuti/otevrena-data/Koncepce-katalogizace-otevrenych-dat-VS-CR---zkracena-verze.pdf>
- [5] CHLAPEK, D., KUČERA J., NEČASKÝ, M.: *Metodika publikace otevřených dat veřejné správy ČR*. [online]. Listopad 2012. Citováno 8.11.2016. Dostupné z WWW: http://www.korupce.cz/assets/partnerstvi-pro-otevrene-vladnuti/otevrena-data/Metodika_Publ_OpenData_verze_1_0.pdf
- [6] *Kompletní obchodní informace. FIRMO* [online]. Praha, 2016 [cit. 2017-04-25]. Dostupné z: <http://www.firmo.cz/cs/proc-firmo>
- [7] *Mapa kriminality. Mapa kriminality* [online]. Praha, 2016 [cit. 2017-04-08]. Dostupné z: <http://mapakriminality.cz/o-aplikaci/>
- [8] *Mapa kriminality ukáže, jak je nebezpečný policejní okrsek, kde žijete*. IDNES.cz [online]. Praha, 2013 [cit. 2017-04-08]. Dostupné z: http://zpravy.idnes.cz/otevrena-spolocnost-vydala-novou-mapu-kriminality-fhc-/krimi.aspx?c=A131126_111653_krimi_cen
- [9] MINIHOFFER, O., HAVLÍČEK, M. a STARÝ, J. *Anglicko-český slovník: zpracování dat, telekomunikace a kancelářské systémy = English-Czech Dictionary: data processing, telecommunications, office systems*. 1. vyd. Voznice: Leda, 1994. 621 s. ISBN 80-901664-1-5.
- [10] MRÁČEK, J., BOČEK, J. a ČEPICKÝ, J. *Jak otevírat data?*. Praha: Fond Otakara Motejla, 2014. ISBN 978-80-87725-16-0.

- [11] Novela zákona o registru smluv. Zprávy Alfa9 [online]. Praha, 2017 [cit. 2017-04-10]. Dostupné z: <http://zpravy.alfa9.cz/absolutenm/templates/zprava.aspx?a=47277>
- [12] O jízdnicích řádech veřejné linkové dopravy. *Portál veřejné správy: Informace pro občany ČR* [online]. 2015 [cit. 2017-04-25]. Dostupné z: <https://portal.gov.cz/app/zakony/zakonPar.jsp?idBiblio=82300&nr=122~2F2014&rpp=15#local-content>
- [13] Otevřená data: Jak Česko umí využít potenciál, který má internet. *Ekonom* [online]. 2015, 2016(48), 36 [cit. 2017-04-07]. Dostupné z: <http://www.otevrenadata.cz/res/data/003/003663.pdf>
- [14] Otevřená data. *Registr smluv* [online]. Praha, 2016 [cit. 2017-04-25]. Dostupné z: <https://data.smlouvy.gov.cz/>
- [15] PETERKA, J.: Digitální agenda pro Evropu [online]. Publikováno 31.5.2010. Citováno 11.12.2016. Dostupné z WWW: <http://www.lupa.cz/clanky/digitalni-agenda-pro-evropu/>
- [16] Registr smluv. *Registr smluv* [online]. Praha: Ministerstvo vnitra České republiky, 2016 [cit. 2017-04-08]. Dostupné z: <https://smlouvy.gov.cz/>
- [17] Správce jízdnicích řádů CHAPS dostal od ÚOHS pokutu dva miliony: *Ekonomika. IDNES.cz: Ekonomika* [online]. Praha, 2016 [cit. 2017-04-12]. Dostupné z: http://ekonomika.idnes.cz/chaps-dostal-pokutu-od-uohs-dp0-/ekonomika.aspx?c=A160107_161544_ekonomika_rts
- [18] Sunlight Foundation: Ten Principles for Opening Up Government Information [online]. 2010. Citováno 15.12.2016. Dostupné z WWW: <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>
- [19] Zlepšili jsme přístup k informacím o exekucích!. *Otevřená společnost* [online]. Praha, 2017 [cit. 2017-04-08]. Dostupné z: <http://www.otevrenaspolecnost.cz/uz-jste-to-slyseli/5364-zlepsili-jsme-pristup-k-informacim-o-exekucich>
- [20] Zveřejnění záznamu. *Portál veřejné správy* [online]. Praha: Ministerstvo vnitra, 2015 [cit. 2017-04-25]. Dostupné z: <https://portal.gov.cz/webfiller/formservice/filler.open?docid=1718132479>

SEZNAM PŘÍLOH

Příloha A: Formulář pro zveřejnění záznamu v Registru smluv.....	44
Příloha B: Datový model vytvořený z dat Registru smluv.....	45

Příloha A: Formulář pro zveřejnění záznamu v Registru smluv

Smluvní strany

Identifikace smluvní strany (případně více smluvních stran)

Doplnit údaje podle IČO nebo DS

Název subjektu Identifikační číslo Datová schránka

Adresa subjektu

Útvar / Odbor / Organizační složka Příjmení / Platce

Identifikační údaje publikující smluvní strany (název, adresa, IČO, identifikátor datové schránky) budou doplněny automatizovaně z údajů datové schránky odesílatele formuláře. Tyto údaje vyplňuje jen, pokud se od údajů datové schránky odesílatele liší, či v případě, že publikujete v zastoupení.

Údaje o smlouvě

Předmět smlouvy

Datum uzavření Číslo smlouvy Podepisující osoba

Hodnota v Kč bez DPH Hodnota v Kč včetně DPH Hodnota v cizí měně Měna

Navázaný záznam

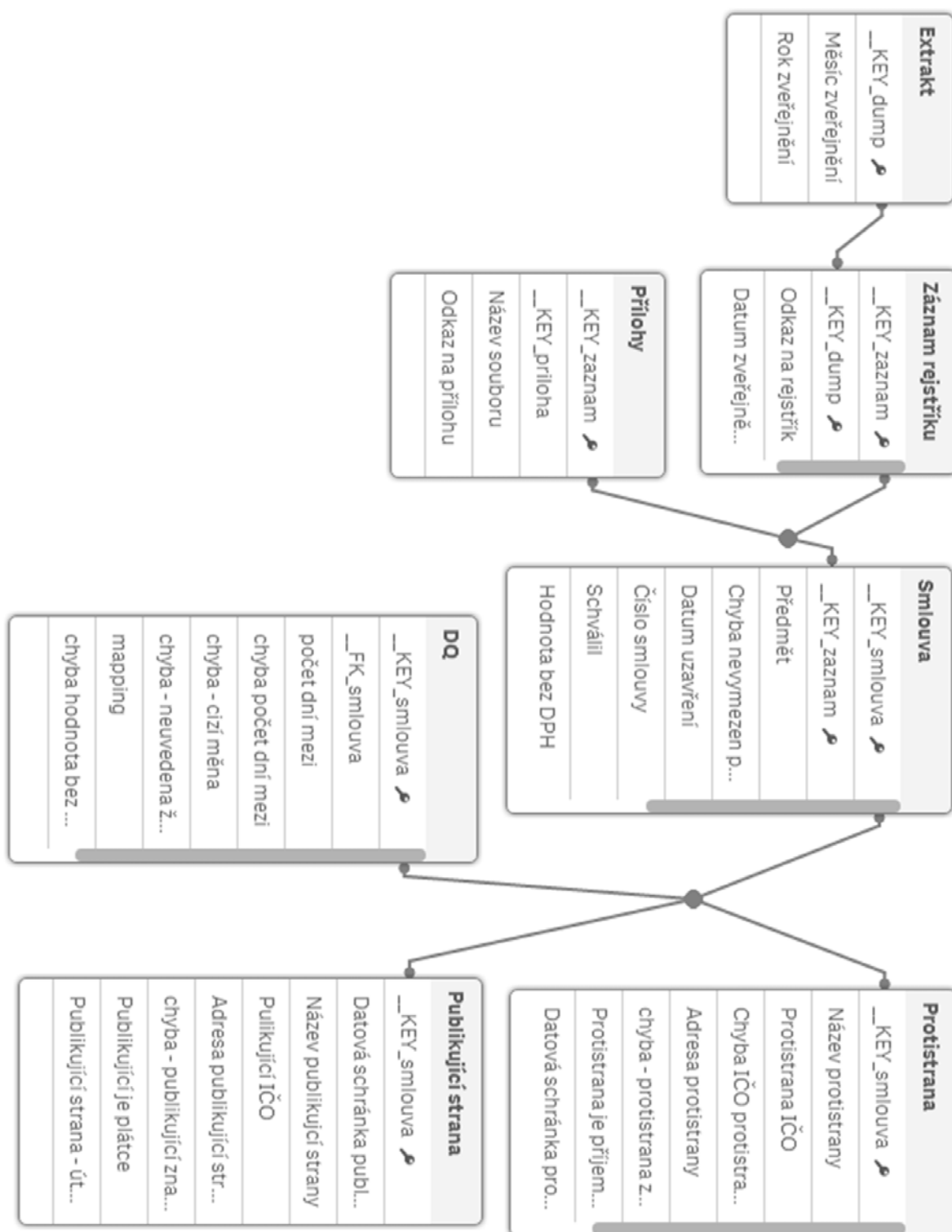
Přílohy

Vložit přílohu **Není vložena příloha**

- +

Zdroj: [20]

Příloha B: Datový model vytvořený z dat Registru smluv



Zdroj: vlastní zpracování