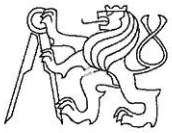
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ



ZADÁNÍ BAKALÁŘSKÉ PRÁCE

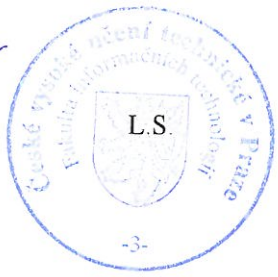
Název:	Otevřená data ve státní správě
Student:	Josef Smrž
Vedoucí:	Ing. Josef Gattermayer
Studijní program:	Informatika
Studijní obor:	Informační systémy a management
Katedra:	Katedra softwarového inženýrství
Platnost zadání:	do konce zimního semestru 2016/17

Pokyny pro vypracování

1) Nastudujte a formou rešerše zpracujte nejnovější trendy v oboru otevřených dat.
2) Analyzujte města (zahrňte Vancouver a Londýn), kde jsou open data nejvíce využívána a popište tamní "dobré praktiky".
3) Popsané "dobré praktiky" aplikujte v ČR a zhodnoťte situaci na MČ Praha 8.
4) Získané poznatky zobecněte a navrhněte vhodné datasety ke zveřejnění. Navrhněte vhodné procesy zveřejňování datasetů.

Seznam odborné literatury

Dodá vedoucí práce.



Ing. Michal Valenta, Ph.D.
vedoucí katedry

prof.Ing. Pavel Tvrdík, CSc.
děkan

V Praze dne 15. února 2015

České vysoké učení technické v Praze
Fakulta informačních technologií
Katedra softwarového inženýrství

Bakalářská práce

Otevřená data ve státní správě

Josef Smrž

Vedoucí práce: Ing. Josef Gattermayer

11. května 2015

Poděkování

Mé poděkování patří Ing. Josefu Gattermayerovi za podnětné rady a vedení této práce, stejně tak jako pracovníkům úřadu MČ Praha 8 za spolupráci při získávání informací o jeho chodu. Díky si zaslouží i má rodina a přátelé za podporu během studií, kterou nešetřili, když byla nejvíce potřebná.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V Praze dne 11. května 2015 .....

České vysoké učení technické v Praze
Fakulta informačních technologií
© 2015 Josef Smrž. Všechna práva vyhrazena.
Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí, je nezbytný souhlas autora.

Odkaz na tuto práci

Smrž, Josef. *Otevřená data ve státní správě*. Bakalářská práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2015.

Abstract

The subject of this bachelor thesis is the usage of open data in the state administration. Its aim is to valorize the situation regarding data publishing at the office of the municipal district of Prague 8 and to suggest the utilization of open data. This step would enhance the transparency of the office, improve public awareness, as well as increase the efficiency at the local self-government.

Within the thesis, one can also find data sets suitable for publication, as well as its proposed publishing processes. The proposal is based on the latest trends which have been identified by the analysis of cities, which already made a great progress in usage of open data. Acquired observations are subsequently used to improve the situation at the office of the municipal district of Prague 8.

The thesis is also dealing with the issue of open data in general and clarifies other terminology associated with the subject.

Keywords open data, local government, municipal district of Prague 8, National open data catalogue

Abstrakt

Tématem této bakalářské práce je využití *otevřených dat* ve státní správě. Cílem práce je zhodnotit situaci týkající se publikování dat na úřadě Městské části Praha 8 a navrhnout jeho přechod na *otevřená data*, což by umožnilo zvýšení transparentnosti úřadu, zlepšení informovanosti občanů a zefektivnění chodu samosprávy.

Práce rovněž zahrnuje návrh datových souborů, vhodných ke zveřejnění úřadem, a procesů jejich zveřejňování. Návrh je založen na nejnovějších trendech, které byly zjištěny analýzou praktik měst nejvíce využívajících *otevřená data*. Získané poznatky jsou následně využity ke zlepšení situace na úřadě.

Dále se práce zabývá také problematikou *otevřených dat* jako celku a objasňuje pojmy s ní spojené.

Klíčová slova otevřená data, státní správa, MČ Praha 8, Národní katalog otevřených dat

Obsah

Úvod **1**
Cíle práce . . . 2
Struktura práce . . . 2

1 Rešerše **3**
1.1 Úvod do problematiky otevřených dat . . . 3
1.1.1 Definice otevřených dat . . . 3
1.1.2 Dataset . . . 4
1.1.3 Katalog otevřených dat . . . 4
1.2 Smysl otevřených dat . . . 5
1.2.1 Motivace k otevření dat . . . 5
1.3 Licence . . . 6
1.4 Druhy publikovaných dat . . . 6
1.4.1 Strukturovaná data . . . 6
1.4.2 Geografická data . . . 7
1.4.3 Ostatní druhy dat . . . 8
1.5 Stupně otevřenosti dat . . . 8
1.6 Technologie pro otevřená data . . . 10
1.6.1 Forma dat . . . 10
1.6.2 Úložiště . . . 10
1.6.3 Katalogy . . . 11

2 Analýza **15**
2.1 Otevřená data v ČR . . . 15
2.1.1 Městská část Praha 8 . . . 15
2.1.2 Děčín . . . 19
2.1.3 Projekty a instituce . . . 20
2.2 Otevřená data v zahraničí . . . 20
2.2.1 Vancouver . . . 20

2.2.2 Londýn 23
2.3 Poznatky z analyzovaných institucí 24
2.3.1 Dopady otevření dat 24
2.3.2 Porovnání ČR se zbytkem světa 25

3 Návrh **27**
3.1 Technologie použité pro otevřená data 27
3.1.1 Způsob publikace dat 27
3.1.2 Forma datasetů 28
3.2 Datasety vhodné ke zveřejnění 29
3.2.1 Data k publikaci obecně 29
3.2.2 Datasety ke zveřejnění na úřadu MČ Praha 8 29
3.3 Návrh procesu zveřejňování datasetů 30

Závěr **33**
Přínosy práce 33
Poučení z praxe 34
Další vývoj 34

Literatura **35**

A Seznam použitých zkratek **39**

B Obsah přiloženého CD **41**

C Obrazová příloha **43**

Seznam obrázků

1.1 Pětihvězdičková *stupnice otevřenosti dat* podle Sira Tima Berners-Leeho [1] 9

2.1 Žebříček informační vyspělosti všech členských států EU [2] 25

C.1 Náhled obrazovky *katalogu* CKAN [3] 43
C.2 Proces přidání *datasetu* do *katalogu* CKAN [4] 44
C.3 Náhled obrazovky *katalogu* Junar používaného městem Sacramento [5] 45
C.4 Náhled obrazovky *katalogu* Socrata používaného městem Chicago [6] 46
C.5 Národní katalog otevřených dat [7] 47
C.6 *Katalog* DKAN Ministerstva financí ČR [8] 48
C.7 *Datasety* MČ Praha 8 v *katalogu* CKAN Czech Republic [9] 49
C.8 Obrazovka přidání příspěvku do systému Publix [10] 50
C.9 Příklad špatně strukturovaných informací v *datasetu* s harmonogramem blokového čištění MČ Praha 8 [11] 51
C.10 *Katalog otevřených dat* města Vancouver [12] 52
C.11 *Katalog otevřených dat* města Londýn [13] 53

Seznam tabulek

2.1 Zastoupení *datasetů* podle formátů v *katalogu otevřených dat* města Vancouver [14] 22
2.2 Zastoupení *datasetů* podle formátu v *katalogu* London Datastore [13] 23

Úvod

S příchodem informační společnosti začala světem hýbat nová komodita, kterou jsou data. Data nás dnes všechny obklopují a nesou v sobě cenné informace. Lze díky nim například zjistit, jaké bude zítra přibližně počasí, za jak dlouho asi lidstvu dojdou fosilní paliva či jak velké oblibě občanů se těší prezident republiky. Hodnota dat spočívá hlavně v tom, že díky nim dokážeme predikovat, co se asi v budoucnu stane a náležitě se na to připravit. Tedy, pokud by dle předpovědi mělo zítra pršet, raději si s sebou vezmeme deštník.

Hodnotu a sílu dat si uvědomují i všechny velké společnosti. Většinou platí, že čím více dat shromáždí a čím více informací z nich dokáží vytěžit, tím více se to promítne na jejich ziscích. Ty jsou pak tou hlavní motivací, aby do získávání a zpracovávání dat investovaly nemalé finanční prostředky.

Dobrým příkladem využití dat mohou být různé věrnostní programy obchodníků. Zákazník si založí klubovou kartu a při každém nákupu nevědomky poskytuje přehled o svých nákupních návycích. Po analýze dat všech zákazníků je pak možné upravit nabídku tak, aby toho zákazníci nakoupili co nejvíce.

Ve státní správě je ale situace dost odlišná. Úřady i jiné veřejné instituce mají dat rovněž mnoho, protože zde ale chybí motivační faktor v podobě zisku, není po zpracování dat často využit jejich plný potenciál. Důvod je ten, že data, která mají úřady k dispozici, nebývají příliš zveřejňována, a když už ano, tak většinou v dále nevyužitelné podobě. To není ideální stav, protože vzhledem k tomu, že jsou tyto instituce financovány z veřejných peněz, mají občané na jejich data nárok. To vyplývá i ze zákona č. 106/1999 Sb., o svobodném přístupu k informacím.

Jako řešení se nabízí přechod instituce na *otevřená data*. Jde o myšlenku spočívající v poskytnutí vlastních, do vhodné formy uvedených dat k volnému užití široké veřejnosti. To umožní občanům i firmám data kreativně využít a přinést tím užitek celé společnosti. Právě možnost neomezeného využití dat veřejných institucí kýmkoliv a za jakýmkoliv účelem je hlavní motivací této práce.

Cíle práce

Cílem této práce je zhodnotit situaci týkající se publikování dat na úřadě Městské části Praha 8 a navrhnout jeho přechod na *otevřená data*, což by umožnilo zvýšení transparentnosti úřadu, zlepšení informovanosti občanů a zefektivnění chodu samosprávy.

Cíle práce rovněž zahrnují návrh datových souborů zvlášť vhodných ke zveřejnění úřadem a procesů jejich zveřejňování. Návrh je založen na nejnovějších trendech, které byly zjištěny analýzou praktik měst nejvíce využívajících *otevřená data*. Získané poznatky jsou následně využity ke zlepšení situace na úřadě.

Dále se práce zabývá tématem *otevřených dat* jako celkem a objasňuje pojmy s ním spojené.

Struktura práce

Celá práce je rozdělena do tří hlavních kapitol:

Kapitola 1 obsahuje teoretickou část práce, pojednává o *otevřených datech* obecně a objasňuje pojmy s nimi spojené. Formou rešerše dále zpracovává nejnovější trendy v *otevřených datech*, a to hlavně v části 1.6. Poskytuje tím čtenáři potřebné technické základy a povědomí o tématu.

Kapitola 2 se věnuje analýze měst a institucí, které *otevřená data* již publikují. Analýza je zaměřena na tuzemská i zahraniční města a na technologickou realizaci zveřejnění dat i jejich využití veřejností. Výsledky analýzy jsou na konci kapitoly zobecněny.

Kapitola 3 se zabývá návrhem přechodu úřadu MČ Praha 8 na *otevřená data*. Jsou zde navrženy i *datasety* ke zveřejnění a proces jejich zveřejňování. To je následně zobecněno pro využití i na jiných institucích.

Kapitola 1

Rešerše

Tato kapitola se věnuje teoretické části bakalářské práce. Uvádí čtenáře do problematiky *otevřených dat* a vysvětluje pojmy, které se jich týkají. V kapitole je dále rozebráno, jaký mají *otevřená data* smysl a jaké úrovně otevřenosti rozlišujeme. V sekci zabývající se používanými technologiemi jsou navíc uvedeny trendy, které je možno v oboru *otevřených dat* aktuálně sledovat.

1.1 Úvod do problematiky otevřených dat

Termín *otevřená data* označuje data, která byla zpřístupněna veřejnosti k volnému užívání. V praxi to znamená vystavení dříve nezveřejňovaných dat na internet. Zavádět *otevřená data* má největší význam u veřejných institucí.

Po přechodu na *otevřená data* vypadá situace následovně: Data jsou uložena v souborech, které se nazývají *datasety*[1]. Ty jsou zveřejněny na webu, typicky v *katalogu otevřených dat*, kde je pak zájemce o data snadno najde a může je podle svých potřeb využít.

Proces zveřejnění *datasetu* do *katalogu* se nazývá publikace. *Dataset* může mít různou formu – například se může jednat o JPEG obrázek, či to může být strukturovaný XML dokument. Na základě toho rozlišujeme u *otevřených dat* jejich kvalitu a *stupeň otevřenosti*[2].

1.1.1 Definice otevřených dat

Otevřená data, anglickým termínem *open data*, jsou data, která mohou být volně užívána, upravována a šířena, a to kýmkoliv a za jakýmkoliv účelem [15].

[1] *Dataset*, česky též nazývaný datová sada, je soubor obsahující data, která jsou vhodná k publikování veřejnosti. *Dataset* je obecně zavedený termín i v ČR, proto bude i nadále používán v této práci.

[2] Více o otevřenosti dat v sekci 3.2.2

Aby data byla otevřená, musí splňovat následující požadavky [15]:

Použití a šíření: Licence, pod kterou jsou data publikována, musí být otevřená a žádné další podmínky užití nesmí být v rozporu s touto licencí.

Dostupnost a přístupnost: Data musí být dostupná online a dosažitelná po vynaložení rozumného úsilí (ideálně bezplatné stažení z internetu). Pokud licence vyžaduje uvést při šíření autora, musí být u souboru dat uveden.

Formát: Data musí být uložena ve vhodné a modifikovatelné formě. Tím je především myšleno, aby byla strojově čitelná a uložena v otevřeném formátu, což umožní, aby se dala zobrazit a upravovat v běžně dostupném bezplatném softwaru.

1.1.2 Dataset

Jak již bylo zmíněno, termínem *dataset* se označuje datový soubor. Jeden *dataset* by se měl týkat ideálně pouze jedné oblasti či problematiky, kterou popisuje, aby byla data pro uživatele přehledná a přínosná. Většinou se jedná o tabulku hodnot, soubor geografických dat či jiný strukturovaný datový soubor.

Nejdůležitějším požadavkem na *dataset* je, že musí být publikovaný online pod *otevřenou licencí*, což znamená, že musí být bezplatně dostupný všem a bez omezení. Pokud by tato podmínka nebyla splněna, není možné data označovat za otevřená, ačkoliv by jinak splňovala sebevíc požadavků na otevřenost.

Další nutností *otevřených dat* je, aby byl soubor *datasetu* strojově čitelný a data uvnitř něj dobře strukturovaná (viz. kapitola 1.4.1). Tím se dosáhne možnosti je efektivně automatizovaně zpracovat bez nutnosti programování speciálního *parseru*[3] pro každý *dataset* individuálně.

Data by navíc měla být publikována pokud možno surová či při zachování maximální možné míry detailu. Vydavatel dat totiž nemůže dopředu zcela správně odhadnout, jaká najdou jeho data využití, proto by vždy měla být publikována kompletní.

1.1.3 Katalog otevřených dat

Katalog otevřených dat je webový portál, na kterém se soustředí *datasety* publikované jednou či více institucemi. Jde tedy o místo, které slouží zájemci o konkrétní data k vyhledání správného *datasetu*, který data obsahuje. Z toho důvodu by měl *katalog* nabízet možnost vyhledání *datasetu* podle názvu, či filtrování podle oblasti, které se týká. Pro uživatele může být rovněž zajímavá možnost filtrace podle formátu, či řazení podle stáří zveřejněného *datasetu*.

[3]Parser je nástroj pro extrakci informací ze strukturovaného dokumentu.

Buď si může instituce přecházející na *otevřená data* zprovoznit vlastní *katalog* nebo může využít katalogy, které již existují. Tyto katalogy často sdružují více institucí, které spojuje například obor jejich působnosti či jejich poloha.

1.2 Smysl otevřených dat

Každá data, především pokud jich je větší objem, mohou mít potenciál být k něčemu užitečná. Data s místy dopravních nehod mohou například odhalit nebezpečná místa či špatně viditelné dopravní značení. Informace o vytíženosti dopravních spojů mohou pomoci zoptimalizovat dopravu ve městě. Data týkající se státních zakázek zase mohou poukázat na špatné hospodaření státu.

U posledního příkladu je patrné, že zainteresované strany udělají všechno proto, aby taková data ideálně vůbec nespatřila světlo světa. Transparentnost ale zdaleka není jediný důvod, proč by bylo dobré, aby data byla dostupná veřejnosti. Často se totiž stává, že data, která vzešla z různých měření, poslouží účelu, pro který byla získána, a dále se již nepoužijí. Důvodem bývá fakt, že o datech často ví jen úzká skupina lidí a ještě menší počet lidí k nim má přístup.

1.2.1 Motivace k otevření dat

Hlavním důvodem, proč by data měla být otevřená, je fakt, že více než kdekoliv jinde zde platí, že „víc hlav víc ví“. Státním orgánům produkujícím data většinou chybí motivace s daty jakkoliv pracovat nad rámec jejich povinností. Pokud jsou ale data uvolněna, najde se nemálo nadšenců, kteří vytvoří zajímavé vizualizace či vyrobí aplikaci, která veřejnosti usnadní život a možná i napomůže fungování instituce, která data vydala. Pokud se navíc nápadu chopí firmy, což není vůbec neobvyklé, jelikož pro ně budou data dostupná zdarma, vytvoří se i nová pracovní místa. Část prostředků investovaných do dat se tedy vrátí státu ve formě daní. [16].

Dalším důvodem, proč je výhodné svá data otevřít, je *zákon o svobodném přístupu k informacím*. Dle tohoto zákona má každý občan právo získat informace o činnosti veřejných institucí [17]. Ten vychází z faktu, že provoz veřejných institucí je financován z veřejných prostředků, tedy hlavně občany, kteří by z toho důvodu měli mít k vyprodukovaným datům přístup.

Vyhovování žádostem o informace ale může úřady vytěžovat, čímž se zbytečně zvyšují i náklady na jejich provoz. Příkladem může být případ problémového občana, který vytrvale zahlcuje svými dotazy zastupitelstvo obce Seč v Pardubickém kraji [18]. Pokud by obec měla svá data otevřená, pouze by všechny případné dotazy směřovala na svůj web.

Posledním z hlavních argumentů pro zveřejnění dat je zvýšení transparentnosti úřadu, ke kterému tím dojde. To způsobí horší podmínky pro korupci, která je jedním z největších problémů veřejného sektoru a přináší státu nemalé finanční ztráty.

1.3 Licence

Jak bylo zmíněno, *definice otevřených dat* (1.1.1) vyžaduje, aby byla data publikována pod *otevřenou licencí*. K účelu tomu účelu se doporučuje využít například licenci *Creative Commons*[4] ve verzi 4.0 [16]. Je ale možné použít i libovolnou jinou *otevřenou licenci*. Aby se licence dala považovat za otevřenou, musí splňovat následující podmínky [15]:

Užití: Licence musí umožnit bezplatné užití dat bez jakýchkoliv poplatků a bez nutnosti souhlasit s jakýmikoliv dalšími podmínkami užití.

Šíření: Licence musí povolovat šíření dat, včetně prodeje díla založeného na zmíněných datech.

Modifikace: Musí být umožněna úprava dat a následné zveřejnění pod stejnou licencí jako původní data. Stejně tak musí být povoleno šíření pouze části dat, které byly součástí původního publikovaného celku. Data také musí být možno spojovat s jinými daty, která nesmí být nijak limitována.

Diskriminace: Licence nesmí svými podmínkami diskriminovat žádný subjekt a musí umožnit používání dat za jakýmkoliv účelem.

1.4 Druhy publikovaných dat

Publikovaná data mohou mít mnoho podob. Ty lze rozdělit do tří hlavních skupin:

- strukturovaná data,
- geografická data,
- ostatní formy dat.

Pro každou skupinu je vhodná odlišná reprezentace a způsob uložení dat. Od toho se odvíjí i více možných formátů *datasetů*. Uvedené druhy dat jsou níže rozebrány.

1.4.1 Strukturovaná data

Jde o asi nejběžnější data se kterými se dá setkat. Data mají textovou formu, která se navíc vyznačuje i pevně danou strukturou, která se dodržuje v celém dokumentu.

Speciálním typem strukturovaných dat je tabulka, což je seznam položek orientovaný po řádcích. Ve sloupcích se pak nachází atributy, které jsou

[4] Creative Commons je americká organizace, která si klade za cíl rozšířit množství autorských děl dostupných veřejnosti k legálnímu využívání a sdílení.

standardně definované prvním řádkem tabulky. První datová položka se tedy zpravidla nachází na druhém řádku tabulky.

Níže jsou rozebrány formáty strukturovaných dat, které jsou ve spojení s *otevřenými daty* nejčastější.

1.4.1.1 Formát XLS a CSV

Formáty XLS a CSV jsou tabulkové formáty. Prvním z nich je formát programu Excel vyvinutí společností Microsoft. Tento formát je uzavřený, a proto nevhodný k použití pro *otevřená data*. Ačkoliv byl tento v roce 2007 nahrazen částečně otevřeným formátem XLSX, na který je doporučováno přejít, zejména na úřadech stále přetrvává zastaralý formát XLS.

Formát CSV je textový formát hodnot, které jsou odděleny čárkami (případně středníky). Je otevřený, a proto spolu s XLSX mnohem vhodnější pro *otevřená data* než formát XLS.

1.4.1.2 Formát XML a HTML

XML i HTML slouží k nesení strukturovaných informací. Hlavním rozdíl mezi oběma formáty je, že první z nich byl vyvinut k popisu dat, zatímco HTML přednostně k jejich zobrazení a zaměřuje se tedy spíše na to, jak data vypadají [19]. K použití pro *otevřená data* je tedy výhodnější použít formát XML.

1.4.1.3 Formát RDF

Speciální pozornost si zaslouží formát RDF, jehož smyslem je vytvořit data, která budou strojově čitelná a srozumitelná, tedy přidat jim význam [20]. Jedná se o jednoduchý datový model, kdy RDF přiřazuje každé určité věci (webovému zdroji) její *URI*[5] a následně jí přidává vlastnosti (odborně *atributy*) s jejich hodnotami [21]. Tím „spojuje“ informace z více webových zdrojů a je stavebním kamenem tzv. *propojených dat* (anglicky linked data). Na tato data se lze následně dotazovat jazykem SPARQL, který dokáže pomocí RDF dokumentů zpracovat data z různých zdrojů.

1.4.2 Geografická data

Geografická data jsou vyčleněna ze strukturovaných proto, že jde typ dat se specifickým použitím, a stojí tedy za to se mu věnovat zvlášť. Hlavní účel geografických dat je přiřadit určitému místu na Zemi jeden nebo více *atributů* s hodnotami a následně tato data vizualizovat na mapě.

Rozšířenými geografickými formáty dat jsou například DWG, DXF, KML, SHP, GML, GeoTIFF nebo GeoJSON.

[5]URI, neboli Uniform Resource Identifier (česky též jednotný identifikátor zdroje) je způsob popisu určitého objektu z hlediska jeho typu a umístění.

1.4.2.1 Formáty KML a GML

Společným rysem formátů KML a GML je to, že jsou otevřené a oba jsou rozšířeními jazyka XML. Oba dva formáty dokáží k souřadnici a mapě přiřadit značku, popis, či geometrický tvar, v praxi se však využívá spíše kombinace obou zároveň. V jazyku GML se obsah vytvoří a pomocí KML se pak zobrazí.

1.4.2.2 Formáty SHP, DWG a DXF

Z formátů SHP, DWG a DXF není ani jeden plně otevřený, přestože formát SHP se k tomu velmi blíží. Autorem tohoto formátu je společnost Evri a slouží k modelování prostorových dat. Podobně tomu je i u formátu DWG, který je vlastněn společností Autodesk. Formát DXF byl vytvořen jako formát pro export dat z formátu DWG pro použití v jiném softwaru.

1.4.2.3 Formáty GeoTIFF a GeoJSON

Formáty GeoTIFF i GeoJSON oba vznikly jako geografická rozšíření formátů TIFF a JSON, aby k nim bylo možné přidat *metadata*[6] s geografickými informacemi. Oba formáty jsou otevřené.

1.4.3 Ostatní druhy dat

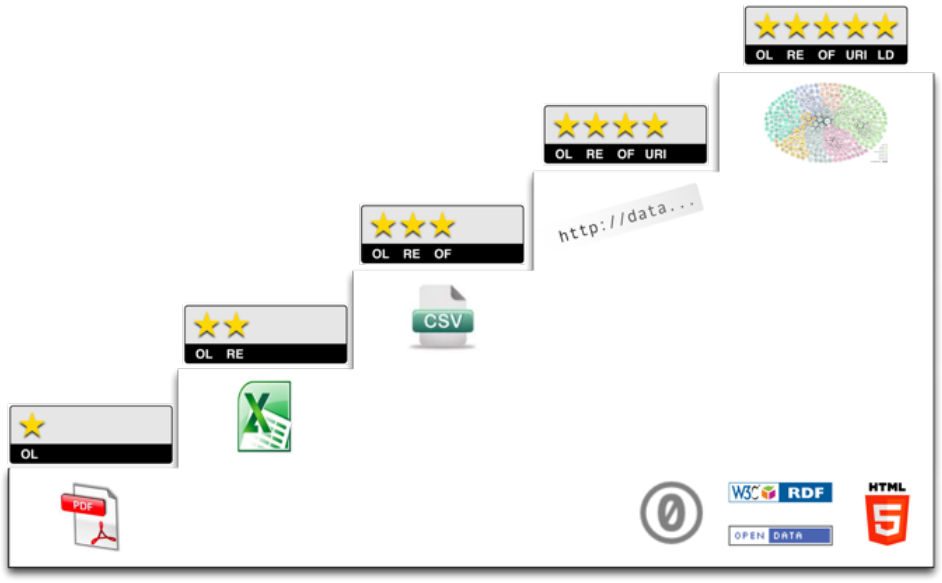
Jelikož existují i jiné formy zveřejňovaných dat než dosud zmíněné, je jim věnována zvláštní sekce. Patří mezi ně například textové dokumenty či obrázky. Standardními formáty dokumentů jsou otevřené TXT, PDF, ODT, RTF či DOCX a již mizící proprietární formát DOC. Z formátů pro obrázky jsou nejběžnějšími JPEG a PNG.

Je důležité se vyvarovat publikování textového dokumentu, který byl naskenován a následně uložen do formátu PDF ve formě obrázku, jelikož by dokument pozbyl strojové čitelnosti [16].

1.5 Stupně otevřenosti dat

Úroveň otevřenosti dat, neboli jejich kvalitu, lze ohodnotit podle *stupnice oteřenosti dat*, kterou stanovil Sir Tim Berners-Lee. Dle ní je možné data ocenit 1-5 hvězdičkami podle toho, do jaké míry jsou otevřená. Zjednodušeně je to znázorněno na obrázku (1.1).

[6] Metadata, neboli data o datech, jsou informace popisující soubor a doplňující jeho hlavní obsah. V kontextu *otevřených dat* jde o informace o *datasetu*, které zahrnují například jeho název, autora, formát, umístění či licenci, pod kterou je *dataset* publikován.



Obrázek 1.1: Pětihvězdičková *stupnice otevřenosti dat* podle Sira Tima Berners-Leeho [1]

Data, která jsou oceněna daným množstvím hvězdiček, musí zároveň splňovat i nároky z nižší kategorie ohodnocení. Jednotlivé kategorie jsou definovány následovně [22]:

* – Data jsou dostupná online a mají otevřenou licenci. Stačí, aby byla data publikována a bylo možné je dohledat. Příkladem je naskenovaný dokument PDF umístěný na webu.

** – Data jsou *strojově čitelná*, což znamená, že jsou reprezentována pomocí znaků a soubor dodržuje určitou strukturu. To umožnuje prohledávat obsah *datasetu* a získávat z něj informace. Tyto podmínky splňuje například XLS tabulka.

*** – Data musí být uložena v souboru, který má otevřený formát. To umožní uživateli s daty pracovat s použitím běžně dostupného softwaru. Příkladem může být soubor soubor ve formátu CSV.

**** – Takto ohodnocená data musí mít vlastní *URI* a být tak jednoznačně identifikována, aby na ně bylo možné odkazovat odjinud.

***** – Data s pěti hvězdičkami musí být navíc tzv. propojená. Toho se docílí odkazováním na jiné *datasety*, čímž se data získají kontext. K tomuto účelu slouží formát RDF.

1.6 Technologie pro otevřená data

Tato část je zaměřena na technologickou stránku zavedení *otevřených dat* a nejnovější trendy v této oblasti. Technologických překážek při nasazování *otevřených dat* není mnoho. Prakticky stačí mít data ve správné formě uložená na místě, ke kterému lze přistupovat ze sítě internet, a poté je publikovat v *katalogu otevřených dat*. Po splnění těchto požadavků je navíc možné ještě nabídnout uživatelům přístup k těmto datům pomocí *API*[7].

1.6.1 Forma dat

Jak již bylo naznačeno dříve, data je nutné mít ve správné podobě. Správnou podobou se rozumí to, aby data na *stupnici otevřenosti* dosáhla co nejvyššího počtu hvězdiček. V praxi se tedy jedná hlavně o to, aby data byla publikována pod *otevřenou licencí*, byla *strojově čitelná* a dostupná v *otevřeném formátu*. Splnění těchto podmínek stačí pro to, aby data dosáhla v otevřenosti hodnoty tří hvězdiček.

První požadavek lze snadno splnit publikováním *datasetu* v textové reprezentaci. Toto zajistit není v případě nově vznikajících *datasetů* problém – stačí publikovat původní soubor, nikoliv sken vytištěného dokumentu.

U historických dokumentů, které jsou archivovány pouze v papírové podobě, může být digitalizace poměrně pracná a časově náročná, hlavně, pokud je jejich počet vysoký. Zde je na místě zjistit, o jaká data by byl ze strany veřejnosti největší zájem a ta digitalizovat jako první [16]. Při digitalizaci je pro převod do textové podoby nejvhodnější využít dnes již běžně dostupnou technologii OCR, která minimalizuje nutnost manuálního přepisu textu.

Co se týká *otevřeného formátu*, neměl by být problém data před publikací do některého z nich převést. Díky otevřenému standardu by měl každý kvalitní proprietární software, který je na instituci použit pro tvorbu či čtení dat, být schopen je uložit do odpovídajícího *otevřeného formátu*. Může to být ale problém u některého geografického softwaru či specializovaných zařízení, která produkují data pouze ve vlastním formátu. Příkladem by mohla být zařízení pro měření hluku či kvality ovzduší. Velkou část dat ale obvykle představují například tabulky programu MS Excel, kde není nic jednoduššího, než je převést do formátu CSV.

1.6.2 Úložiště

Ať už mají data jakoukoliv podobu, musí být někde uložena. Při výběru úložiště máme standardně na výběr ze dvou možností. Tou první je použít server, na kterém již běží webové stránky publikující instituce a případně poběží i *ka-*

[7] API je rozhraní služby nebo aplikace, které je poskytnuto dalším programátorům pro snadnou programovou komunikaci se zmíněnou aplikací či službou.

talog otevřených dat, a nahrávat *datasety* přímo na něj. Druhou možností je použít jako úložiště jiný server.

Obě varianty mají své klady i zápory. Zvolení první možnosti se zdá jednodušší a přímočařejší. Je třeba spravovat pouze jeden server, vše je na jednom místě a *datasety* není problém nahrávat přímo do *katalogu*, pokud se zvolí některý s touto funkcionalitou. Takto jednoduché to je v případě, že se jedná o vlastní webový server.

Pokud se o web ale stará nějaká firma, k serveru nebývá ze strany instituce fyzický přístup a rozšíření o podporu pro ukládání *datasetů* pak může být ze strany firmy zpoplatněno. Další nevýhodou využití pouze jednoho serveru může být právě centralizace, kdy v případě selhání serveru dojde k výpadku webu, *katalogu* i úložiště.

Druhá možnost zahrnuje nutnost pořízení a správy dalšího serveru, pokud ale už nějaký datový server existuje, má již pravděpodobně vyřešeno i zálohování a zvolení této varianty nepřinese žádné další výdaje. Úložiště může být také poskytováno třetí stranou, například prostřednictvím *cloudové služby*[8].

1.6.3 Katalogy

Katalogy jsou velice důležitým prvkem *otevřených dat*, jelikož umožňují zveřejnění *datasetů* a jejich doplnění o *metadata*, díky kterým následně může zájemce o data vyhledat ten správný *dataset*, ve kterém jsou data obsažena. Jak již bylo zmíněno v úvodní části této kapitoly, instituce otevírající svá data má na výběr, zda využije jeden z již existujících centrálních *katalogů*, kam svá data bude přidávat, nebo si zprovozní svůj vlastní. Tato část se zabývá právě výběrem vhodného *katalogu* ke zprovoznění.

1.6.3.1 Požadavky na katalog otevřených dat

Aby bylo možné *katalog* efektivně používat, měl by splňovat několik základních funkčních požadavků. Hlavním funkčním požadavkem, tedy to, co dělá *katalog katalogem*, je schopnost vést záznamy o *datasetech* doplněné o *metadata* a poskytnout je uživateli. To ale přestává stačit v momentě, kdy počet *datasetů* vzroste. Další vhodné funkcionality *katalogu* jsou proto následující:

- vyhledávání podle názvu *datasetu*,
- filtrování podle formátu a kategorie *datasetu*,
- řazení podle názvu *datasetu*, či doby od jeho zveřejnění,
- možnost nahrání *datasetu* přímo do *katalogu*,

[8] Označení cloud (z anglického slova pro oblak) zahrnuje internetové technologie nabízející uživatelům určité služby (například webové úložiště) nezávisle na fyzickém umístění hardwarových komponent poskytovatele těchto služeb, které pro uživatele přestává být podstatné.

- možnost integrace *katalogu* do CMS,
- API pro snadný programový přístup.

Další podmínkou pro bezproblémový chod *katalogu* je zajištění jeho bezpečnosti. Pokud by tomu tak nebylo, hrozila by kompromitace či ztráta dat v *katalogu*. To by mohlo mít dopad i na dobré jméno instituce, která *katalog* provozuje. Základní bezpečnostní požadavky zahrnují například:

- jednoznačná autentizace a autorizace uživatele pověřeného k manipulaci s *katalogem* (např. přidávání, modifikaci, či mazání *datasetu*) [23],
- zachování konzistence přidaného *datasetu*,
- ochrana před bezpečnostními útoky (např. *XSS* či *SQL injection*[9]).

Dalšími volitelnými bezpečnostními požadavky může být například záznam událostí, které v rámci *katalogu* proběhly [23], či pravidelná záloha *datasetů*. To však lze ale provádět i externě, nemusí to tedy být zahrnuto přímo ve funkcionalitách *katalogu*.

1.6.3.2 Katalog CKAN

Katalog CKAN (obr. C.1) je nejpopulárnějším *katalogem* pro *otevřená data* [24]. Jedná se o *otevřený software*, což znamená, že je poskytován zdarma, a díky svým vlastnostem se stal velmi populární, ať už ve své originální podobě, či upravený tak, aby vyhovoval konkrétnímu použití. Národní instance *katalogu* CKAN pro ČR je v provozu na adrese `http://cz.ckan.net`.

CKAN nabízí mnoho funkcionalit, jak pro uživatele, tak pro správce *katalogu*. Z uživatelských funkcí stojí za zmínku možnost vyhledávání v *katalogu*, filtrování podle štítků, formátu, skupiny či vydavatelské organizace. V *katalogu* lze též vyhledávat podle území označeného na mapě. Výsledky vyhledávání je dále možné seřadit podle názvu, relevance či doby uplynulé od zveřejnění *datasetu*. Administrátorovi *katalog* umožňuje přidat nový *dataset* (proces přidání *datasetu* zachycen na obrázku C.2), následně ho editovat, či smazat. *Katalog* navíc zaznamenává aktivitu uživatelů a poskytuje správu uživatelských profilů.

CKAN poskytuje integrované datové úložiště, kromě přidání odkazu na *dataset* ho tedy lze i nahrát na server. Pokud je nahrána tabulka ve formátu CSV či XLSX, je navíc k dispozici grafická vizualizace obsažených dat. CKAN disponuje i bohatým *API* pro snadnější programový přístup k publikovaným*datasetům*. [25].

Katalog podporuje i integraci do *CMS systémů*[10] Drupal a Wordpress, pro které jsou dostupná instalovatelná rozšíření. Kromě toho existují i distribuce

[9] Jde o útoky namířené proti webové aplikaci podstrčením škodlivého kódu s cílem získat nad ní nebo její částí kontrolu.

[10] CMS jsou systémy pro správu obsahu publikovaného na webových stránkách.

zmíněných systémů, které již mají CKAN integrovaný v sobě. Tou založenou na Drupalu je DKAN, což je rovněž *otevřený software*, který navíc podporuje i vizualizaci geodat na mapě pomocí *javascriptové*[11] technologie s názvem Leaflet [26]. V ČR je DKAN využíván například Ministerstvem financí ČR (viz 2.1.3.2). Distribuce založená na Wordpressu se nazývá DataPress, byla vyvinuta britským Open Data Institutem a využívá ji pro svá data například město Londýn (viz 2.2.2).

1.6.3.3 Katalog Junar

Junar je komerční *katalog*, jehož předností je kromě standardních funkcí *katalogu* hlavně vizualizace tabulkových a geografických dat. Kromě toho poskytuje i *API* či možnost vložení do vlastních webových stránek pomocí *iframe*[12]. Cena za provoz *katalogu* Junar začíná na částce $200 měsíčně [27]. Právě povedené vizualizace a uživatelská podpora je zřejmě jeden z hlavních důvodů, díky kterým si Junar našel své zákazníky. Jedním z nich je město Sacramento. Na obrázku C.3 je zobrazen příklad koláčového grafu, který je možné pomocí *katalogu* Junar vytvořit právě z dat, publikovaných městem Sacramento.

1.6.3.4 Katalog Socrata Open Data Portal

Posledním z nejpoužívanějších *katalogů otevřených dat* je *katalog* od společnosti Socrata (obr. C.4). Je stejně jako *katalog* Junar komerční, ale bohužel není známa jeho cena, která je zřejmě individuální. Socrata Open Data Portal je schopen v podstatě všeho, co umí *katalog* Junar, s tím rozdílem, že navíc nabízí internetový obchod s mobilními a webovými aplikacemi, které byly vytvořeny speciálně pro *katalog* společnosti Socrata. Tyto aplikace využívají k připojení ke *katalogu* Socrata Open Data API (SODA) [28]. Aplikací není v současné době mnoho a některé z nich jsou placené. To se ale může brzy změnit, jelikož vyvíjet a umístit aplikaci do obchodu může každý.

1.6.3.5 Vlastní implementace katalogu

Jednou z možností při volbě *katalogu* je také implementace svého vlastního. Tento přístup však nebývá příliš častý. Jsou s ním spojeny náklady, které se zdají být zbytečné, když bezplatná a funkční řešení již existují.

V některých případech však není nutné sahat po komplexním řešení, proto když už přeci jen vznikne *katalog* implementovaný samotným vydavatelem dat, bývá velice jednoduchý a s minimem funkcí.

[11] JavaScript je oblíbený skriptovací jazyk používaný především na webu, kde se vkládá přímo do HTML kódu.

[12] Iframe je HTML tag, který umožňuje do webové stránky vložit tzv. rám a zobrazit v něm obsah jiné stránky.

Kapitola 2

Analýza

V této kapitole jsou popsány aktuální trendy používání *otevřených dat*, jak ve světě, tak v České republice. Zjišťování současných praktik bylo zaměřeno na veřejné instituce, především na státní správu. Výsledky analýzy jsou nakonec zobecněny pro aplikaci na jiné instituci a situace v ČR je porovnána se situací v zahraničí.

2.1 Otevřená data v ČR

Tato sekce je soustředěna na aktuální vývoj v oboru *otevřených dat* na veřejných institucích v České Republice.

2.1.1 Městská část Praha 8

MČ Praha 8 se rozkládá na území o rozloze $21,8\ km^2$ a čítá více než 100.000 obyvatel, díky čemuž se řadí na 5. místo co se týče rozlohy a 3. místo co do počtu obyvatel vzhledem k ostatním pražským částem [29]. Jde tedy o poměrně velkou samosprávu, s čímž je spojen i velký potenciál pro využití *otevřených dat*. Pro zajištění kvality analýzy a reflektování skutečného stavu, proběhla analýza úřadu MČ Praha 8 ve spolupráci s jeho pracovníky.

2.1.1.1 Otevírání úřadu

Trend zpřístupňování informací na Prahu 8 dorazil dne 1. 3. 2013 zahájením projektu „Otevřený úřad“. Jeho cílem bylo zefektivnit poskytování informací občanům a zvýšit tím transparentnost úřadu. Celkové náklady na projekt byly stanoveny na $2.883.646,24\ Kč$ a termín dokončení na 31. 8. 2014 [30].

Z dostupných zdrojů vyplývá, že hlavní část projektu byla orientována na školení, kterých se zúčastnilo přes 350 pracovníků úřadu. Školení se týkala především komunikačních dovedností úředníků a dovedností projektového řízení pro vedoucí pracovníky [31]. V rámci projektu také vznikla „Metodika

pro otevřená data (open data) Městské části Praha 8“, na základě které byly v rámci subprojektu „Otevřená data“ publikovány tři *datasety* do národního *katalogu* „CKAN Czech Republic“. Národní *katalog* CKAN s nahranými *datasety* MČ Praha 8 je zobrazen na obrázku C.7.

Publikované *datasety* však nejsou příliš kvalitní, jelikož jde pouze o odkazy na webové stránky MČ, kde data nejsou vhodně strukturována a není uvedeno, pod jakou licencí jsou publikována. Ačkoliv byla dne 7. 8. 2014 radou města avizována příprava plánu dalšího rozvoje *katalogu*, zatím se o žádné další *datasety* nerozrostl [31].

2.1.1.2 Stávající technologické řešení

Jak již bylo zmíněno, MČ Praha 8 momentálně využívá k publikaci svých *datasetů* celonárodní *katalog* CKAN, který však zatím není na *datasety* vydané úřadem příliš bohatý. Data, která chce úřad zveřejnit totiž primárně nahrává přímo na webový portál MČ Praha 8, většinou formou příloh k publikovaným příspěvkům. Na portále funguje vyhledávání v příspěvcích, které lze omezit pouze na přílohy, takže lze *datasety* i poměrně snadno vyhledat podle názvu.

Problém nastává v případě, že se uživatel chce pouze informovat, jaká data úřad nabízí. Při vyhledávání totiž není možné zobrazit všechny publikované přílohy a uživatel tedy musí znát její název nebo alespoň jeho část. Rovněž není k dispozici žádné další filtrování. Právě z těchto důvodů by bylo vhodnější použití datového *katalogu*.

Webová technologie, kterou úřad používá k provozu svého portálu, se jmenuje Publix ^{4}He, neboli Publix Helium. Jde o redakční systém vyvinutý pomocí technologie *.NET*[13] společností CVS Development s.r.o. [32], která k systému poskytuje úřadu i technickou podporu. Jde o systém se standardními funkcemi, mezi které patří například publikace příspěvků s možností přidání příloh, úprava struktury webu či správa uživatelů systému [33].

Po přihlášení přenese systém uživatele do administračního prostředí webu, které je pro lepší představu zachyceno na obrázku C.8. Na levé straně se zobrazí stromová struktura webu, která zde setrvává po celou dobu průchodu administračním rozhraním. To je poměrně užitečné, jelikož je pak možné kdykoliv snadno přejít do jiné sekce webu a vše je srozumitelně hierarchicky uspořádáno.

Příliš uživatelsky přívětivý ale není systém vyskakujících oken, který je použit k provádění téměř všech administrativních úprav webu. Tím se systém Publix značně odlišuje od moderních *CMS* typu Wordpress či Drupal, a většina úkonů prováděných v systému se tak stává poměrně složitými. O administraci webu, ale i publikování nových přípěvků, se proto stará určený pracovník úřadu.

[13] Názvem .NET se označuje soubor technologií určených k vývoji pro operační systém MS Windows.

2.1.1.3 Proces publikace dat

Jak již bylo uvedeno, úřad momentálně publikuje svá data prostřednictvím systému Publix, který je určen především pro provozování webových stránek. Data v něm jsou tedy publikována jako přílohy k příspěvkům. Kvůli obsluze systému, která je pro běžného úředníka poměrně komplikovaná, je bohužel prakticky každý publikovaný příspěvek nahráván jeho administrátorem. Proces zveřejňování dat typicky probíhá následovně:

1. vytvoření *datasetu* ve vhodné, strojově čitelné podobě;
2. kontrola *datasetu* před zveřejněním (není pravidlem);
3. odeslání *datasetu* připraveného ke zveřejnění e-mailem administrátorovi webových stránek,
4. publikace *datasetu* administrátorem na webové stránky formou přílohy příspěvku.

Tento proces publikace není příliš vhodný, protože je zdlouhavý a zbytečně zaměstnává administrátora, který funguje pouze jako prostředník mezi úředníky a webem. Mnohem efektivnější by bylo přiřadit pracovníkům úřadu uživatelské účty s právy publikovat příspěvky a zaškolit je v používání základních funkcionalit redakčního systému Publix.

2.1.1.4 Data publikovaná úřadem

Množství dat, které pro občany MČ Praha 8 na svém webu zveřejňuje, je již nyní poměrně vysoké. Mezi nejzajímavější publikované *datasety* patří například:

- harmonogram svozu nebezpečného odpadu,
- harmonogram přistavení velkoobjemových kontejnerů,
- povodňové mapy,
- harmonogram blokových čištění komunikací,
- parkovací zóny,
- výsledky hospodaření MČ Praha 8.

Z uvedeného seznamu *datasetů* a po pečlivém prozkoumání webových stránek MČ je zřejmé, že úřad je již co se týká zpřístupňování dat svým občanům poměrně vstřícný. Pokud uživatel ví, co ho konkrétně zajímá, jsou informace i poměrně snadno dohledatelné. *Datasety* však nejsou katalogizovány, je tedy poměrně obtížné zjistit, co všechno úřad nabízí.

Problém může nastat při pokusu o získávání dat programově, například mobilními či webovými aplikacemi, které jsou aktuálním trendem prezentace

dat. Data úřadu totiž ve většině případů nejsou příliš dobře strojově zpracovatelná. Tím je myšlena především nekonzistentní struktura *datasetů*, která znemožňuje automatizaci procesu získávání dat.

Tento neduh se často objevuje zejména u tabulek, ať už u umístěných přímo v těle webových stránek, tak u uložených ve formátu XLS. Příkladem může být *dataset* s harmonogramem blokového čištění ulic MČ, který je zobrazen na obrázku C.9. *Dataset* je ve formátu XLS, ani z daleka ale nedodržuje pravidla zápisu dat do tabulky.

Dalším problémem publikovaných dat úřadu jsou uzavřené formáty, a to především XLS či DOC. Dále jsou data dostupná často pouze v už zpracované formě, prezentovaná například webovými aplikacemi vloženými na městském webu. To se týká především geodat a jejich zobrazení pomocí interaktivní mapy.

V sekci s informacemi o umístění kontejnerů na elektroodpad jsou například k dispozici jen obrázky map, na kterých je umístění kontejneru označeno křížkem. To by nevadilo, kdyby byl zveřejněn i *dataset* se seznamem zeměpisných souřadnic těchto kontejnerů, tak tomu ale bohužel není.

2.1.1.5 Aktuální využití dat

Jelikož je kvůli absenci datového *katalogu* poměrně těžké zjistit, jaká data vůbec úřad nabízí, není jejich používání příliš podpořeno. Spolu se strojově nečitelnými nebo špatně strukturovanými *datasety* pak další využitelnost dat klesá k nule.

Úřad ale nechal pro občany vyvinout některé aplikace sám. Z těch mobilních jsou to aplikace „Mapový portál MČ Praha 8“, „Měsíčník Osmička“ a „Osmička – Měsíčník Prahy 8“. Všechny tři jsou vytvořené pro operační systémy iOS a Android a počet jejich stažení nepřevyšuje ani v jednom případě hodnotu 500. První ze zmíněných aplikací nabízí mapu s body zájmu, jako jsou dětská hřiště, sportoviště nebo kontejnery [34]. Vzhledem k tomu, že aplikace pro systém Android bohužel nefungovala na telefonech, na kterých byla testována, nebylo možné funkcionality ověřit. Druhá a třetí zmíněná aplikace plní ten samý účel – jde o elektronickou verzi měsíčníku, který MČ Praha 8 vydává. Každá z nich má jiného autora, pro aplikaci „Osmička - Měsíčník Prahy 8“ je to společnost 4WORKS Solutios (`http://www.posunemevasvys.cz/project/ios-aplikace-osmicka/`) a pro „Měsíčník Osmička“ jde o společnost Aitom (`http://www.aitom.cz/reference/praha-8`). Proč byla vyvinuta další aplikace, která slouží ke stejnému účelu, se bohužel nepodařilo zjistit.

Co se týče webových aplikací, úřad na adrese `http://www.praha8.cz/Mapovy-portal-mestske-casti-Praha-8.html` nabízí „Mapový portál“, který obsahuje body zájmu a dále mapy pro krizové řízení. Ty zahrnují povodňové mapy nebo krizovou mapu pro ochranu obyvatelstva.

Další webovou aplikací je stránka „Veřejné zakázky“ (dostupná na adrese `https://zakazky.praha8.cz`), která umožňuje zobrazit přehled všech veřej-

ných zakázek MČ Praha 8. Přehled zakázek je velmi detailní a někdy je možno zobrazit i smlouvu zakázky ve formátu PDF. Pro efektivnější používání je v seznamu možné vyhledávat či ho filtrovat.

Poslední zajímavou webovou aplikací na webu MČ Praha 8 je „Klikací rozpočet“ (dostupný na adrese `http://rozpocet.praha8.cz`), který nabízí vizualizaci hospodaření MČ formou koláčových a sloupcových grafů.

Kromě výše zmíněného jsou veřejností zřejmě využívány i dokumenty s daty nahrané na web, to však není podložené žádnou konkrétní statistikou, jako například počtem stažení.

2.1.2 Děčín

Statutární město Děčín otevřelo svá data v druhé polovině roku 2013 a stalo se tím jedním z prvních českých měst, která se do této iniciativy zapojila. Za tento přístup získalo rok na to zvláštní cenu za *otevřená data* v soutěži „OTEVŘENO x ZAVŘENO“, kterou vyhlašuje Otevřená společnost, o.p.s. [35].

2.1.2.1 Katalog

Město Děčín nepoužívá pro *otevřená data* žádný specializovaný *katalog*, ale publikuje svá data pomocí redakčního systému přímo na svůj web do vyhrazené kategorie. To má své výhody i nevýhody. Hlavní výhodou je jednoduchost řešení a integrace s webem města, mezi nevýhody pak patří hlavně omezení funkcionalit pouze na vyhledávání v *datasetech*.

2.1.2.2 Zpřístupněná data a jejich využití

Hlavním formátem, který pro svá data město Děčín používá, jsou formáty programu Excel pro tabulková data a formát SHP pro data geografická. U tabulkových dat je využito obou formátu, XLSX i XLS, přičemž druhý jmenovaný stále převládá hlavně u historických dat. Z hlediska kvality tedy děčínská data dosahují dvou až tří hvězdiček.

Z obsahového hlediska publikuje město Děčín hlavně data týkající se hospodaření města a data geografická. Příkladem jsou účetní výkazy nebo majetek a rozpočet města, dále pak geografická data týkající se částí obce a komunikací, dopravního značení, veřejného osvětlení či sběrných kontejnerů.

Publikovaná data našla prozatím využití hlavně v mobilní aplikaci „Děčín“ (`http://www.mmdecin.cz/urad-on-line/mobilni-aplikace-mesta`), díky které možné zjistit novinky z dění ve městě, zobrazit placená i neplacená parkoviště (parkovné lze pomocí aplikace rovnou i zaplatit), či se dokonce objednat na úřad pro vyřízení nějaké záležitosti.

2.1.3 Projekty a instituce

Otevřená data začala být v ČR v poslední době podporována některými institucemi a klíčovými státními orgány, což je pro jejich rozšíření velmi důležité. Příklady takových orgánů jsou například Ministerstvo vnitra ČR a Ministerstvo financí ČR.

2.1.3.1 Ministerstvo vnitra ČR

Podpora *otevřených dat* ze strany Ministerstva vnitra ČR se v poslední době výrazně zvýšila a vyvrcholila v dubnu 2015 spuštěním „Národního katalogu otevřených dat“[14] (obrázek C.5) v rámci Portálu veřejné správy. Tento projekt má za cíl soustředit *otevřená data* všech českých veřejných institucí do jednoho *katalogu otevřených dat* [36].

Publikace dat do tohoto *katalogu* zpravidla probíhá buď nahráním *datasetu* přímo na Portál veřejné správy nebo odesláním *metadat* o *datasetu* do její datové schránky formou datové zprávy. Pokud instituce, která má zájem do *katalogu* publikovat, již využívá vlastní *katalog* dat založený na *katalogu* CKAN, je možné napojit Národní katalog otevřených dat na jeho CKAN *API*. Po úspěšném napojení se pak budou data publikované v používaném *katalogu* CKAN automaticky propisovat i do *katalogu* Ministerstva vnitra.

2.1.3.2 Ministerstvo financí ČR

Začátkem roku 2015 otevřelo svá data i Ministerstvo financí [37]. Používá k tomu katalog DKAN (obrázek C.6) dostupný na adrese `http://data.mfcr.cz/cs/dataset`, ve kterém je momentálně publikováno 35 datasetů [8]. V přidávání dalších datasetů ministerstvo plánuje pokračovat. Tento postup kladně hodnotí Fond Otakara Motejla, který se *otevřenými daty* ve státní správě dlouhodobě zabývá [37].

2.2 Otevřená data v zahraničí

Nadcházející část se soustředí na současné trendy zveřejňování dat a jejich využití v zahraničních městech.

2.2.1 Vancouver

Město Vancouver svá data otevřelo veřejnosti spuštěním svého *katalogu* již v roce 2009 a od té doby ho plní svými *datasety* [38].

[14] Národní katalog otevřených dat je dostupný na webu Portálu veřejné správy na adrese `https://portal.gov.cz/portal/rejstriky/data/97899/`

2.2.1.1 Otevírání dat

Po spuštění *katalogu* nastalo v oblasti *otevřených dat* ve Vancouveru několik zásadních událostí. V květnu 2010 přišlo město s plánem pro *otevřená data*, ve kterém si definovalo krátkodobé a dlouhodobé cíle. Ty zahrnovaly publikaci konkrétních *datasetů* a angažování uživatelů *otevřených dat*. V roce 2011 již bylo za přechod na *otevřená data* oceněno cenou pro nejinovativnější organizaci ve státě Britská Kolumbie.

Koncem února 2013 uspořádal Vancouver svůj první „Open Data Hackathon“[15], jehož cílem bylo povzbudit občany k vývoji aplikací a tvorbě vizualizací využívajících vancouverská *otevřená data* [38]. *Hackathony* od té doby získávají ve Vancouveru stále více na popularitě a město se do nich aktivně zapojuje.

2.2.1.2 Katalog otevřených dat

Vancouver ke katalogizování používá vlastní *katalog*, který má umístěn na webových stránkách (obr. C.10). *Katalog* je uspořádán do tabulky s nabídkou různých formátů *datasetů* a není možné v něm vyhledávat, pouze přejít na odpovídající písmeno abecedy. Je ale možné využít vyhledávání pomocí internetového prohlížeče, jelikož zde není použito stránkování a celý seznam *datasetů* je tedy zobrazen na jedné stránce.

Vancouver dosud publikoval více než 170 *datasetů*, které jsou vetšinou aktualizovány jednou týdně, a průběžně přidává další [14]. Mezi nimi se nachází taková data jako například seznam parkovacích míst, umístění stojanů na kola, či místa posprejovaná graffiti.

Data, která *katalog* obsahuje, je tedy možné rozdělit na dvě hlavní skupiny: geodata a tabulky hodnot. *Datasety* geodat jsou ve většině případů k dispozici v *otevřených formátech* KML a SHP a proprietárním formátu DWG. Některé *datasety* obsahující geodata je navíc možné zobrazit i v Google mapách či Bing mapách. V případě tabulek si uživatel může většinou vybrat z otevřeného CSV a proprietárního XLS formátu.

V tabulce 2.2.1.2 je vyčísleno zastoupení jednotlivých formátů *datasetů* v *katalogu*. Z tabulky je patrné, že při celkovému počtu 170 *datasetů* jich je většina poskytnuta ve více formátech a díky vysokému zastoupení *otevřených formátů* téměř všechny *datasety* dosahují kvality tří hvězdiček.

[15] Pojem hackathon vznikl spojením anglických pojmů hack a marathon, a označuje veřejnou událost (zahrnující obvykle i soutěž), na které se sejdou programátoři s cílem vytvořit aplikace na konkrétní téma či systém, často s využitím nekonvenčních přístupů.

Formát	Počet *datasetů*	Poznámka
KML	128	Otevřený formát pro geodata postavený na XML
SHP	126	Částečně otevřený geometrický formát
DWG	107	Uzavřený 2D/3D formát společnosti Autodesk
XLS	42	Uzavřený formát tabulek společnosti Microsoft
CSV	39	Otevřený formát hodnot oddělených čárkami
Ostatní	23	XML, JSON, TXT, Atom, GeoRSS, MR SID

Tabulka 2.1: Zastoupení *datasetů* podle formátů v *katalogu otevřených dat* města Vancouver [14]

2.2.1.3 Využití dat

Díky tomu, že Vancouver poskytuje *otevřená data* již delší dobu a jejich využívání podporuje, vzniklo z těchto dat již mnoho aplikací. Velký podíl na tom má právě podpora ze strany města a *hackathony*, které město spolufinancuje. Některé z aplikací, ačkoliv vznikly na datech Vancouveru, našly následně uplatnění v celé Kanadě.

Prvním příkladem využití dat Vancouveru je mobilní aplikace „Park 'in' Spot“, která na mapě přehledně zobrazí parkovací místa, ať už jde o placené či neplacené parkování.

Další mobilní aplikací je „Tap Map“[16], která zobrazí polohu veřejných kohoutků s pitnou vodou, kterých je ve městě více než 550.

Zajímavá je i aplikace „Vancouver Collection Schedule“[17], která informuje o harmonogramu svozu odpadu ve městě. V aplikaci si je možné nastavit upozornění na blížící se plánovaný odvoz odpadu v konkrétní lokalitě, či nahlásit případný problém s touto službou.

Vancouver dále poskytuje a využívá data o stavu pitné vody (`http://app.vancouver.ca/EngWaterQuality_Net/Default.aspx`). Na základě těchto dat pak provádí pravidelná čištění a výplachy potrubí pro udržení její vysoké kvality.

Jednou z nejnovějších a zároveň i nejzajímavějších webových aplikací, která je postavená na *otevřených datech*, je aplikace „Career Path.“ Aplikace vznikla a vyhrála první místo v rámci *hackathonu* „CODE 2015“, který se konal 20. – 22. 2. 2015 napříč celou Kanadou a byl zaměřen na data z „Canada's Open Government portal“ (`http://open.canada.ca`).

Jde o webovou aplikaci, která má pomoci mladým lidem s volbou svého profesního zaměření a budoucího zaměstnání. Na základě volby profese zobrazí uživateli její popis a obvyklé požadavky. Tím nejzajímavějším je zřejmě graf

[16] Aplikace „Tap Map“ je dostupná na adrese `https://itunes.apple.com/ca/app/tap-map/id391108983`.

[17] Aplikace „Vancouver Collection Schedule“ je dostupná na adrese `https://play.google.com/store/apps/details?id=net.recollect.vancouver`.

počtu pracovních míst v dané profesi v průběhu času a odhad jejich budoucího vývoje. To je srovnáno s odhadovanými počty uchazečů o dané zaměstnání a napojeno na web s nabídkou pracovních míst.

2.2.2 Londýn

Jako hlavní město Velké Británie má i Londýn velké množství dat, ze kterých je možno získat zajímavé informace, což vedlo k jejich otevření veřejnosti. Londýn spustil svůj první *katalog otevřených dat* v lednu 2010 [39]. V listopadu 2014 pak město přešlo na *katalog* London Datastore (obr. C.11), který používá technologii DataPress.

2.2.2.1 Katalog otevřených dat

Katalog DataPress dokáže v *datasetech* vyhledávat podle názvu, filtrovat podle témat a formátu či podle vybrané oblasti na mapě, které by se měl *dataset* týkat. Výsledky vyhledávání či filtrování lze řadit podle názvu, relevance, či stáří [40].

London Datastore aktuálně obsahuje 586 *datasetů* [13]. Z tabulky 2.2.2.1 je vidět, že v *katalogu* je výrazně více dat tabulkových než geodat. Z těch jich je většina ve formátu XLS, který má z hlediska otevřenosti hodnotu pouze dvou hvězdiček. Dalo by se tedy usoudit, že Londýn má poměrně hodně *datasetů*, avšak data nejsou co do otevřenosti příliš kvalitní.

Formát	Počet *datasetů*	Poznámka
XLS	426	Uzavřený formát tabulek společnosti Microsoft
CSV	177	Otevřený formát hodnot oddělených čárkami
PDF	89	Otevřený formát pro publikaci dokumentů
HTML	60	Otevřený webový značkovací jazyk
XML	28	Otevřený strukturovaný značkovací jazyk
SHP	21	Částečně otevřený geometrický formát
Ostatní	38	TSV, JPEG, RTF

Tabulka 2.2: Zastoupení *datasetů* podle formátu v *katalogu* London Datastore [13]

2.2.2.2 Využití dat

Velký vliv na využití londýnských *otevřených dat* mělo zpřístupnění informací o dopravních spojích v reálném čase. Ty v roce 2010 uvolnil k bezplatnému používání státní podnik Transport for London [41], díky čemuž vzniklo velké množství aplikací pro hromadnou dopravu na mobilní operační systémy. Hodnota celkového času, který londýnští cestující za 3 roky od zpřístupněním dat díky nim ušetřili, byla odhadnuta na 15 až 18 milionů liber.

Díky londýnským *otevřeným datům* vzniklo i mnoho vizualizací, z nichž je zajímavý například animovaný vývoj ceny londýnských nemovitostí mezi lety 2000 a 2013[18].

Většina dalších užití *otevřených dat* se zaměřuje celou na Británii a některé aplikace, které díky nim vznikly jsou k nalezení v *katalogu* na adrese `http://data.gov.uk/apps`.

Zajímavým příkladem je aplikace „Walkonomics“ dostupná na adrese `http://www.walkonomics.com`, která hodnotí britské chodníky dle jejich kvality, stoupání či bezpečnosti.

Další aplikací, která je užitečná je například „CrashMap“ dostupná na adrese `http://www.crashmap.co.uk`, která poskytuje přehled o všech dopravních nehodách ve Velké Británii v letech 2005 až 2013. Ze stejné oblasti je i aplikace „Roadworks.org“ (`http://roadworks.org`), která na mapě zobrazuje všechny aktuální uzavírky a práce na silnici v zemi.

Významným uplatněním *otevřených dat* je i analýza 37 milionů vydaných lékařských předpisů, díky které se nyní ušetří až 200 milionů liber ročně na předražených lécích [42].

2.3 Poznatky z analyzovaných institucí

Tato sekce zobecňuje poznatky, které byly pomocí analýz získány, a popisuje tím stav, ve kterém se *otevřená data* aktuálně nachází.

Z případů analyzovaných institucí je vidět, že velká města pochopila, že otevírat svá data se jednoznačně vyplatí. Zásadní zrychlení otevírání úřadů po celém světě nastalo hlavně v posledních několika letech, kdy se téměř dá mluvit o situaci, kdy mezi sebou jednotlivé instituce doslova závodí, aby byly otevřeny co nejdříve. To je patrně způsobeno i honbou za politickými body, kdy zpřístupňování informací a zvyšování transparentnosti úřadů bývá veřejností obecně vnímáno velmi kladně.

2.3.1 Dopady otevření dat

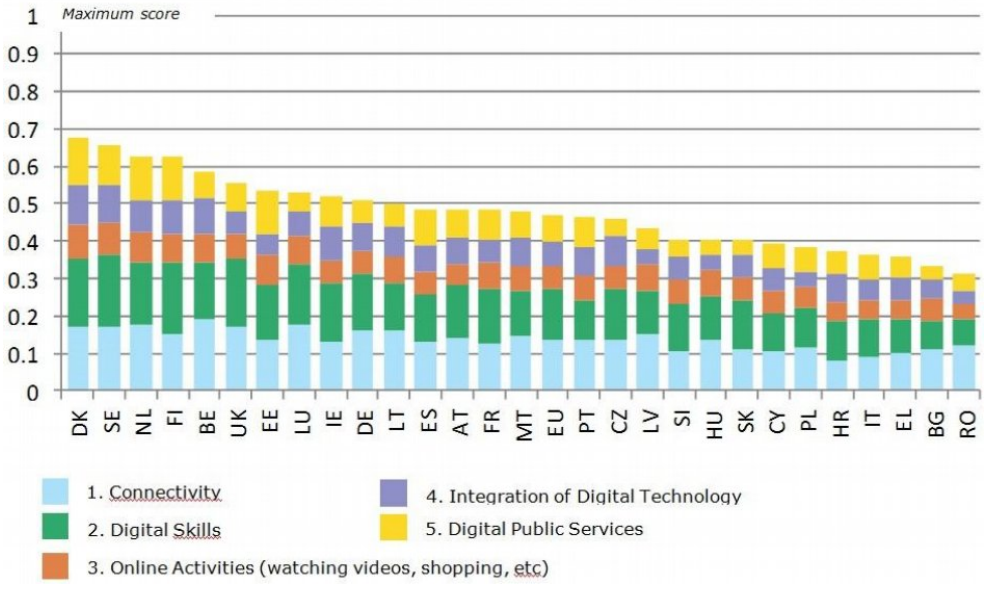
Zpřístupněním dat bylo v mnoha případech dosaženo ušetření finančních prostředků, zefektivnění chodu města a zvýšení zájmu a informovanosti obyvatel o dění ve městě. S tím souvisí i jejich motivace k využívání dat či tvorbě vlastních aplikací na nich postavených. Tuto činnost města často podporují organizováním *hackathonů*, což se velmi osvědčuje a posouvá vývoj kupředu. Nápady využití *otevřených dat* se často transformují i na výdělečnou činnost.

[18] Animace je dostupná na adrese `http://illustreets.co.uk/blog/maps-and-apps/london-house-prices-evolution-over-13-years/`

2.3.2 Porovnání ČR se zbytkem světa

Z analýzy je zřejmé, že veřejný sektor v ČR za světovými špičkami co se týče *otevřených dat* stále výrazně pokulhává. Důsledkem toho se zbytečně prodražuje chod institucí a snižuje jejich efektivita i transparentnost.

Jedním z hlavních důvodů je špatný stav IT ve veřejné správě ČR obecně. Aktuálně probíhá drtivá většina komunikace občana se státní správou výhradně osobně a papírovou formou. Místy se sice objevují snahy toto změnit a některé segmenty státní správy zdigitalizovat, každý pokus však zpravidla naráží při realizaci. Jak si ČR v tomto směru vede zachycuje i nejnovější žebříček informatické vyspělosti států Evropské unie (obrázek 2.1) sestavený začátkem roku 2015. Zatímco celkově ze všech sledovaných ukazatelů Česká republika obsadila 17. místo, což je už samo o sobě podprůměrný výsledek, v oblasti veřejných služeb, kam například spadá právě elektronická komunikace s úřady, je ČR až na 26. místě ze všech 28 členských států Evropské unie.



Obrázek 2.1: Žebříček informační vyspělosti všech členských států EU [2]

Dalším důvodem, proč do České republiky trend *otevřených dat* nepřišel už dříve, je zřejmě obecná nevole státní správy k aktivitě nad rámec povinností, které jí ukládá zákon. Začátkem roku 2015 však přišlo Ministerstvo vnitra a Ministerstvo financí se svými projekty pro podporu otevřených dat (viz. sekce 2.1.3), a tak lze v dohledné době očekávat v této oblasti zlepšení.

Kapitola 3

Návrh

Tato kapitola je zaměřena na aplikaci poznatků získaných analýzou měst, která jsou v *otevřených datech* pokročilá, na úřadě MČ Praha 8. To zahrnuje návrh technologického řešení zavedení *otevřených dat*, či návrh *datasetů* zvlášť vhodných ke zveřejnění. V kapitole je dále navržen i proces zveřejňování *datasetů* se snahou o omezení zbytečné byrokracie a důrazem na kvalitu výsledných dat.

Aktuální situace týkající se zveřejňování dat na úřadě MČ Praha 8 již byla nastíněna v části 2.1.1 této práce. Bylo v ní poukázáno na nedostatky týkající se hlavně formátu a způsobu publikace dat. Problematickým se jevil i současný proces zveřejňování dat. Je proto na místě se zabývat otázkou změny aktuálního stavu a návrhu nového, efektivnějšího řešení. Při návrhu nového řešení si je třeba uvědomit, že každá větší změna zároveň znamená přeškolení velkého množství osob. To s sebou nese i jisté náklady a překážky při přechodu na nové řešení.

3.1 Technologie použité pro otevřená data

Tato sekce se zabývá návrhem technologického řešení publikace *otevřených dat* úřadu MČ Praha 8.

3.1.1 Způsob publikace dat

Jak již bylo v analytické části uvedeno, z technologického hlediska je situace taková, že úřad používá pro zveřejňování informací výhradně své webové stránky. V minulosti byla v rámci projektu „Otevřený úřad“ snaha na data nahraná na web odkazovat z národní instance *katalogu* CKAN, což bohužel dopadlo zveřejněním pouze 3 *datasetů* ve zmíněném *katalogu*.

Jelikož byl předchozí pokus o katalogizaci dat neúspěšný, je třeba se o ni pokusit znovu, avšak s jiným přístupem. Jelikož do webu MČ již byly investovány finanční prostředky a svou funkci plní poměrně dobře, není důvod od

tohoto řešení upouštět. Příliš se ale neosvědčilo následné ruční přidávání *datasetů* do *katalogu* CKAN, jelikož se jich ve výsledku publikovalo velmi málo.

3.1.1.1 Použitý katalog

Pokud mají být data skutečně kvalitně katalogizována, je třeba tuto činnost co nejvíce usnadnit a důsledně vyžadovat přidávání všech publikovaných *datasetů* včetně *metadat* do *katalogu*.

Řešením by bylo nechat do aktuálně používaného systému Publix doprogramovat podporu pro automatické přidání *datasetu* do *katalogu otevřených dat*. Pro uživatele by to pak znamenalo jen doplnit *dataset* nahrávaný na web o *metadata*, která jsou nutná k exportu do *katalogu*. *Dataset* by do něj pak byl automaticky zveřejněn. Zvyknout si na tento nový postup by pro pracovníky úřadu nemělo být příliš náročné a i doprogramování požadované funkcionality do systému se zdá být poměrně snadné. Nejsou proto očekávány zvlášť vysoké náklady na realizaci tohoto řešení.

Vzhledem k tomu, že se objevila oficiální snaha o centrální správu *otevřených dat* orgánů veřejné správy pomocí Národního katalogu otevřených dat (2.1.3.1), stálo by za to se do této iniciativy zapojit a odkazovat na *datasety* MČ Praha 8 právě z tohoto *katalogu*.

Přidání do Národního katalogu otevřených dat se skládá ze dvou částí. Tou první je vygenerování XML dokumentu, který obsahuje *metadata* o *datasetu*. Vytvořit generátor tohoto dokumentu za zadaných hodnot nepředstavuje žádný technický problém a je velmi snadné.

Druhým krokem je odeslání vygenerovaného XML dokumentu s *metadaty* o *datasetu* formou datové zprávy do datové schránky Národního katalogu otevřených dat. Před odesláním datové zprávy je vyžadováno se nejdříve přihlásit do datové schránky úřadu MČ Prahy 8. Při implementaci této funkcionality tedy bude nutné využít její API.

3.1.2 Forma datasetů

Zdaleka největším problémem dat MČ Praha 8 je jejich špatná struktura. Díky tomu data do velké míry ztrácí na hodnotě, protože je není možné automaticky strojově zpracovat. K dosažení vyšší kvality dat přitom stačí dodržovat základní strukturu *datasetů*. Zejména se jedná o tabulky, které je vhodné opatřovat záhlavím, aby byl jasně definován význam jednotlivých sloupců.

Dalším problematickým typem dat jsou téměř všechna geografická data publikovaná úřadem, jelikož jde v drtivé většině o výřezy map ve formě obrázků. Dále jsou mapová data prezentována formou o webové aplikace. K zajištění využitelnosti dat je nutné publikovat i soubory geodat či alespoň tabulky s výčtem zeměpisných souřadnic a jejich významu.

Dalším nedostatkem, který často zbytečně snižuje kvalitu dat MČ Praha 8, je publikace pouze v *uzavřených formátech*. Převedení *datasetu* do *otevřeného formátu* je přitom zpravidla otázka chvilky.

Při dodržení těchto požadavků lze bez větších problémů dosáhnout pro data MČ Praha 8 na stupnici otevřenosti tří hvězdiček. To pak umožní využívat data více, než tomu bylo doposud. Zavedení této datové politiky se však neobejde bez odpovídajícího proškolení zaměstnanců, kteří budou data k publikaci připravovat.

3.2 Datasety vhodné ke zveřejnění

Při návrhu dat, která by stálo za to zveřejnit, je vhodné se inspirovat tam, kde už *otevřená data* nějakou dobu fungují. Pro tuto potřebu poslouží zahraniční města analyzovaná v části 2.2.

3.2.1 Data k publikaci obecně

Z analýzy je patrné, že největší smysl má otevírání dat v těchto oblastech dopravy, kartografie zdravotnictví a ekonomiky.

Asi nejvíce užitečným a v praxi využitelným se zdá zveřejňování maximálního množství dat o dopravě, díky jejíž optimalizaci lze ušetřit společnosti velké množství času a peněz.

Hlavními a také nejsnadnějšími daty ke zveřejnění jsou v této oblasti jízdní řády veřejné dopravy. Tato data již existují a jsou využívána, často jsou ale licencována a za jejich používání je třeba platit, což od jejich využívání odrazuje malé vývojáře.

Dalšími užitečnými daty z oblasti dopravy jsou například přehledy uzavírek, informace o dopravních nehodách nebo místa častého porušování dopravních předpisů. S využitím těchto dat lze dopravu zoptimalizovat, odhalit na komunikacích nebezpečná místa či místa se špatně viditelným dopravním značením.

Význam zpřístupnění dat z ostatních zmíněných oblastí také není zanedbatelný. Otevřené mapové podklady mají například velký význam pro použití při vizualizaci dat z ostatních oborů. Data za zdravotnictví mohou být užitečná pro zlepšení životního stylu a dosažení zdravějšího obyvatelstva. Detaily hospodaření instituce zase může pomoci ušetřit náklady a směřovat finance tam, kde jsou nejvíce potřeba.

3.2.2 Datasety ke zveřejnění na úřadu MČ Praha 8

Vzhledem k tomu, že je Praha 8 pouze městskou částí, má přístup jen k některým druhům dat. Jak bylo zmíněno v části 2.1.1.4, velkou část těchto dat již úřad publikuje, je však nutné se soustředit na jejich kvalitu co do otevřenosti (viz. kapitola), aby bylo možné data dále zužitkovat.

Nejvhodnějšími daty k tomuto zkvalitnění se zdají být geografická data všech možných bodů zájmu, které se v Praze 8 nacházejí a jsou již na jejím webu dostupná formou obrázků, adres či mapových aplikací. Tyto body zájmu zahrnují například:

- parky,
- dětská hřiště,
- sportoviště,
- psí louky,
- kamerový systém,
- objekty integrovaného záchranného systému,
- úkryty obyvatel před nebezpečím,
- kontejnery na různé druhy odpadu,
- parkoviště.

Uvedený seznam by ještě mohl pokračovat a být rozšířen o spousty dalších, dosud nezveřejněných dat, jako jsou například místa dopravních nehod či místa spáchání přestupků.

Dalšími daty vhodnými ke zveřejnění či převedení do vhodné formy jsou zejména:

- harmonogram vyvážení komunálního odpadu,
- harmonogram blokového čištění komunikací,
- stav kvality ovzduší,
- kompletní přehled hospodaření MČ Praha 8,
- výsledky hlasování v místních referendech,
- aktuální stavy obsazení škol a školek,
- kalendář kulturních akcí.

3.3 Návrh procesu zveřejňování datasetů

Jak z analýzy v sekci 2.1.1.3 vyplynulo, proces zveřejňování dat na úřadě MČ Praha 8 není příliš optimální. Hlavním problémem je zbytečné vytěžování administrátora webového portálu, jehož úkolem nyní je pouze mechanicky umisťovat data na web a nepřidává tak datům žádnou přidanou hodnotu.

Dle „Metodiky publikace otevřených dat“ veřejné správy ČR, se proces publikace skládá z následujících bodů [43]:

1. analýza a výběr dat k uveřejnění,
2. výběr vhodného formátu dat,
3. návrh způsobu přístupu k datům,
4. export dat do vhodného formátu,
5. publikace dat,
6. katalogizace dat.

Pro použití v případě úřadu MČ Praha 8 je vhodné proces upravit a přizpůsobit ho aktuální situaci na úřadě. Oproti současnému stavu, kdy administrátor *katalogu* plní i funkci redaktora, je vhodné rozdělit web na oblasti zodpovědnosti a ty přidělit vedoucím pracovníkům jednotlivých oddělení úřadu, kteří pak budou za obsah na jednotlivých částech webu zodpovídat.

Aby mohli data zveřejňovat přímo pracovníci, kteří data vytvářejí či upravují, měli by projít školením, během kterého se naučí ovládat administrační rozhraní systému Publix a dozví se, jak články s *datasety* na web publikovat. Zvláštní pozornost je třeba věnovat rozšíření pro publikaci do Národního katalogu otevřených dat, jehož implementace byla navržena v sekci 3.1.1.1.

Tím se proces zásadně zjednoduší a umožní častější zveřejňování většího množství dat. Standardně by pak měl proces publikace *otevřených dat* na úřadě MČ Praha 8 vypadat následovně:

1. vytvoření *datasetu* ve vhodné, *strojově čitelné* podobě;
2. kontrola *datasetu*, který je připraven ke zveřejnění (volitelně);
3. publikace článku s *datasetem* na web. Během publikace jsou vyplněna i *metadata* o *datasetu*, který se díky tomu automaticky přidává i do Národního katalogu otevřených dat formou odkazu na soubor publikovaný na webu MČ Praha 8.

Do procesu je zařazen i volitelný krok, kdy vedoucí pracovník oddělení kontroluje *dataset* před zveřejněním. Tento krok lze vynechat podle potřeby a praktik konkrétního oddělení.

Obdobně lze zavést proces zveřejňování *otevřených dat* i na ostatních veřejných institucí, jak v ČR, tak i v zahraničí. Důležitý je prvek rozdělení zodpovědnosti za publikovaný obsah, což motivuje k produkci kvalitnějších a užitečnějších dat a ulehčuje práci administrátorovi webu.

Zobecněný proces publikace *otevřených dat* by tedy měl vypadat takto:

1. Vytvoření *datasetu* ve vhodné podobě,
2. kontrola *datasetu* před zveřejněním (volitelně),
3. publikace *datasetu* do *katalogu otevřených dat*.

Z důvodu využití více osobami by měl být katalogizační systém co nejjednodušší na obsluhu, aby ji zvládli i méně technicky zdatní uživatelé. V tomto směru se velmi osvědčil systém CKAN, ve kterém je proces přidání nového *datasetu* velmi přímočarý.

Závěr

Hlavním cílem této bakalářské práce bylo zhodnotit situaci na úřadě Městské části Praha 8 (provedeno v části 2.1.1) a navrhnout pro něj způsob nasazení *otevřených dat* (kapitola 3), včetně návrhu *datasetů* vhodných ke zveřejnění a procesů jejich zveřejňování. Toho bylo dosaženo zjišťováním aktuálně použitých technologií a procesů publikace dat přímo od pracovníků úřadu.

Pro účel zhodnocení situace na úřadě MČ Praha 8 byly také nastudovány a formou rešerše zpracovány nejnovější trendy v oboru *otevřených dat* (kapitola 1). Dále byla pro tento účel v kapitole 2 analyzována města a instituce, kde jsou *otevřená data* nejvíce využívána. U těchto měst byly popsány jejich „dobré praktiky“ při využití *otevřených dat*, které byly následně aplikovány při návrhu řešení pro úřad MČ Praha 8. Návrh byl následně zobecněn pro využití i na jiných institucích.

Zadání práce bylo tedy splněno ve všech jeho bodech.

Přínosy práce

Hlavním přínosem této práce je návrh přechodu úřadu MČ Praha 8 na *otevřená data*. Vzhledem k tomu, že snaha o zavedení *otevřených dat* na úřadě již jednou dopadla neúspěšně, práce nabízí jiné řešení, které by se mohlo setkat s lepším přijetím a úspěšnější realizací.

Pokud by na úřadě skutečně došlo k zavedení *otevřených dat*, je zde reálná šance na zlepšení fungování MČ, informovanosti obyvatel a jejich zapojení do veřejného života.

Dalším přínosem práce je zjištění aktuální situace v oblasti *otevřených dat* ve veřejném sektoru v České republice a v zahraničí. Dále byly navrženy *datasety* vhodné ke zveřejnění a procesy jejich zveřejňování.

Poučení z praxe

Poučením z praxe je, že mít svá data veřejně přístupná se veřejným institucím skutečně vyplatí. Místo investic do zpracování dat a nabízení služeb občanům pak mnohdy stačí data pouze zveřejnit. Jejich uplatnění pak často najdou sami občané či firmy, které na datech mohou postavit i svůj způsob podnikání. Tím navíc dojde i k vytvoření nových pracovních míst, na čemž stát následně profituje výběrem daní.

Zveřejněním maxima informací též dochází ke zvýšení transparentnosti úřadu, a tím k omezení prostoru pro korupci ve veřejném sektoru, která bývá jedním z nejpalčivějších témat v mnoha zemích světa.

Další vývoj

Dalším stádiem zpřístupňování a využívání informací po úspěšném nasazení *otevřených dat* je řízení se podle konceptu propojených dat. Tento koncept ve své podstatě spočívá v dosažení pěti hvězdiček stupnice *otevřených dat*, čímž se z nich stanou *propojená otevřená data* (v angličtině linked open data). Toho lze dosáhnout propojením *datasetů* ve formátu RDF s jinými daty pomocí jejich *URI*, čímž data získají na významu a kontextu.

Literatura

[1] Hausenblas, M.: *5 Star Open Data [online].* [cit. 2015-04-20]. Dostupné z: http://5stardata.info/

[2] Evropská komise: *Jak pokročila ČR a další země na cestě k digitální Evropě? ČR zatím spadá do skupiny středně výkonných států, trápí ji e-government [online].* [cit. 2015-05-01]. Dostupné z: http://ec.europa.eu/ceskarepublika/news/150224_digitalni_agenda_cs.htm

[3] Datasets [online]. [cit. 2015-05-09]. Dostupné z: http://demo.ckan.org/en/dataset

[4] Create Dataset [online]. [cit. 2015-05-09]. Dostupné z: http://demo.ckan.org/en/dataset/new

[5] Sacramento Crime Reports (Last 3 Years) [online]. [cit. 2015-05-09]. Dostupné z: http://data.cityofsacramento.org/datastreams/90192/sacramento-crime-reports-last-3-years

[6] Police Stations - Map [online]. [cit. 2015-05-09]. Dostupné z: https://data.cityofchicago.org/Public-Safety/Police-Stations-Map/gkur-vufi

[7] Ministerstvo vnitra ČR: *Národní katalog otevřených dat - Datové sady [online].* [cit. 2015-05-01]. Dostupné z: https://portal.gov.cz/portal/rejstriky/data/97898/2015-04.html

[8] Ministerstvo financí ČR: *Otevřená data Ministerstva financí [online].* [cit. 2015-05-11]. Dostupné z: http://data.mfcr.cz/cs/dataset

[9] CKAN Czech Republic: Praha 8 [online]. [cit. 2015-05-09]. Dostupné z: http://cz.ckan.net/group/praha-8

[10] Publix [online]. [cit. 2015-05-09]. Dostupné z: http://test.publix.cz/servis/index.aspx

[11] Blokové čistění komunikací MČ Praha 8 v roce 2015 [online]. [cit. 2015-05-09]. Dostupné z: http://www.praha8.cz/file/mwq1/Blokove-cisteni-komunikaci-2015.xls

[12] Open Data Catalogue [online]. [cit. 2015-05-09]. Dostupné z: http://data.vancouver.ca/datacatalogue/index.htm

[13] The City Data Team: *London Datastore [online]*. [cit. 2015-04-24]. Dostupné z: http://data.london.gov.uk/dataset

[14] City of Vancouver: *Open Data Catalogue [online]*. [cit. 2015-04-24]. Dostupné z: http://data.vancouver.ca/datacatalogue/index.htm

[15] Open Knowledge Foundation Network: *Open Definition [online]*. [cit. 2015-04-07]. Dostupné z: http://opendefinition.org/od/

[16] Fond Otakara Motejla: *Jak otevírat data [online]*. [cit. 2015-04-18]. Dostupné z: http://www.otevrenadata.cz/res/data/001/003498.pdf

[17] Parlament České republiky: *106/1999 Sb. Zákon o svobodném přístupu k informacím*. [cit. 2015-04-18]. Dostupné z: http://portal.gov.cz/app/zakony/download?idBiblio=47807&nr=106~2F1999~20Sb.&ft=pdf

[18] Půlpán, D.: *Stěžovatel trápí patologickými dotazy malou obec, pomůže jen změna zákona [online]*. [cit. 2015-04-18]. Dostupné z: http://pardubice.idnes.cz/obec-sec-stezovatel-jan-patera-dea-/pardubice-zpravy.aspx?c=A150403_151554_pardubice-zpravy_msv

[19] W3Schools: *Introduction to XML [online]*. [cit. 2015-04-19]. Dostupné z: http://www.w3schools.com/xml/xml_whatis.asp

[20] W3Schools: *Introduction to RDF [online]*. [cit. 2015-04-19]. Dostupné z: http://www.w3schools.com/webservices/ws_rdf_intro.asp

[21] Xuan, D. L.: *Diplomová práce: Webové aplikace s využitím Linked Open Data [online]*. [cit. 2015-04-19]. Dostupné z: https://www.vse.cz/vskp/show_file.php?soubor_id=1247567

[22] Berners-Lee, T.: *Linked Data [online]*. [cit. 2015-04-20]. Dostupné z: http://www.w3.org/DesignIssues/LinkedData.html

[23] Chlapek, D.; Kučera, J.; M., N.: *Koncepce katalogizace otevřených dat VS ČR [online]*. [cit. 2015-04-21]. Dostupné z: http://www.mvcr.cz/soubor/koncepce-katalogizace-otevrenych-dat-vs-cr-pdf.aspx

[24] The Open Knowledge Foundation: *About CKAN [online]*. [cit. 2015-04-21]. Dostupné z: http://ckan.org/about/

[25] The Open Knowledge Foundation: *Feature Tour [online]*. [cit. 2015-04-21]. Dostupné z: http://ckan.org/features/

[26] Nucivic: *DKAN [online]*. [cit. 2015-04-21]. Dostupné z: http://nucivic.com/dkan/

[27] Junar: *Junar [online]*. [cit. 2015-04-22]. Dostupné z: http://www.junar.com/

[28] Socrata: *Socrata Open Data Portal [online]*. [cit. 2015-04-22]. Dostupné z: http://www.socrata.com/products/open-data-portal/

[29] Magistrát hlavního města Prahy: *Mapa městských částí [online]*. [cit. 2015-04-25]. Dostupné z: http://www.praha.eu/jnp/cz/o_meste/mestske_casti/index.html

[30] Vlk, T.: *Začíná projekt „Otevřený úřad“ [online]*. [cit. 2015-04-25]. Dostupné z: http://www.praha8.cz/Zacina-projekt-Otevreny-urad.html

[31] Vlk, T.: *Končí projekt otevřený úřad, pokračuje otevřenost úřadu směrem k občanovi [online]*. [cit. 2015-04-25]. Dostupné z: http://www.praha8.cz/Konci-projekt-otevreny-urad-pokracuje-otevrenost-uradu-smerem-k-obcanovi.html

[32] CVS Development s.r.o.: *Redakční systém Publix: Technické informace [online]*. [cit. 2015-04-26]. Dostupné z: http://www.publix.cz/technika

[33] CVS Development s.r.o.: *Redakční systém Publix: Základní postupy práce se systémem [online]*. [cit. 2015-04-26]. Dostupné z: http://www.publix.cz/prace

[34] eMan s.r.o: *Google Play: Mapový portál MČ Praha 8 [online]*. [cit. 2015-05-02]. Dostupné z: https://play.google.com/store/apps/details?id=cz.eman.comte

[35] Otevřená společnost,o. p. s.: *Výsledky soutěže Otevřeno x Zavřeno 2014 [online]*. [cit. 2015-05-10]. Dostupné z: http://www.otevrenaspolecnost.cz/dokumenty/pro-media/tz_oxz2014_vysledky_1.docx

[36] Ministerstvo vnitra ČR: *Národní katalog otevřených dat (metodika) [online]*. [cit. 2015-05-01]. Dostupné z: https://portal.gov.cz/portal/ovm/rejstriky/ogd/x-nkod.html

[37] Fond Otakara Motejla: *Ministerstvo financí otevírá svá data veřejnosti [online]*. [cit. 2015-05-11]. Dostupné z: http://www.otevrenadata.cz/aktualni-deni/ministerstvo-financi-otevira-pristup-k-fakturam-pro-verejnost/

[38] City of Vancouver: *Open Data catalogue [online]*. [cit. 2015-04-24]. Dostupné z: http://vancouver.ca/your-government/open-data-catalogue.aspx

[39] Arthur, C.: *Boris Johnson to launch London 'Datastore' with hundreds of sets of data [online]*. [cit. 2015-04-24]. Dostupné z: http://www.theguardian.com/technology/2010/jan/06/london-datastore-launch-johnson-mashups

[40] Open Data Institute: *Datapress [online]*. [cit. 2015-04-24]. Dostupné z: http://datapress.io/

[41] The Telegraph: *London transport data opened to app developers [online]*. [cit. 2015-05-02]. Dostupné z: http://www.telegraph.co.uk/technology/news/7830412/London-transport-data-opened-to-app-developers.html

[42] *Prescribing Analytics [online]*. [cit. 2015-05-02]. Dostupné z: http://www.prescribinganalytics.com/

[43] Chlapek, D.; Kučera, J.; M., N.: *Metodika publikace otevřených dat veřejné správy ČR [online]*. [cit. 2015-05-09]. Dostupné z: www.mvcr.cz/soubor/metodika-publ-opendata-verze-1-0-pdf.aspx

Příloha A

Seznam použitých zkratek

ČVUT České Vysoké Učení Technické

FIT Fakulta Informačních Technologií

ČR Česká republika

EU Evropská unie

MČ Městská část

IT Informační technologie

iOS iPhone Operating System

CMS Content Management System

URI Uniform Resource Identifier

API Application Programming Interface

RDF Resource Description Framework

PDF Portable Document Format

XLS Microsoft Excel Spreadsheet

XLSX Microsoft Excel Open Xml Spreadsheet

JSON JavaScript Object Notation

TIFF Tagged Image File Format

DXF Data Exchange Format

DWG Drawing

SHP Shapefile

GML Geography Markup Language

KML Keyhole Markup Language

JPEG Joint Photographic Experts Group

XML Extensible Markup Language

CSV Comma-Separated values

HTML HyperText Markup Language

SQL Structured Query Language

SPARQL SPARQL Protocol and RDF Query Language (rekurzivní zkratka)

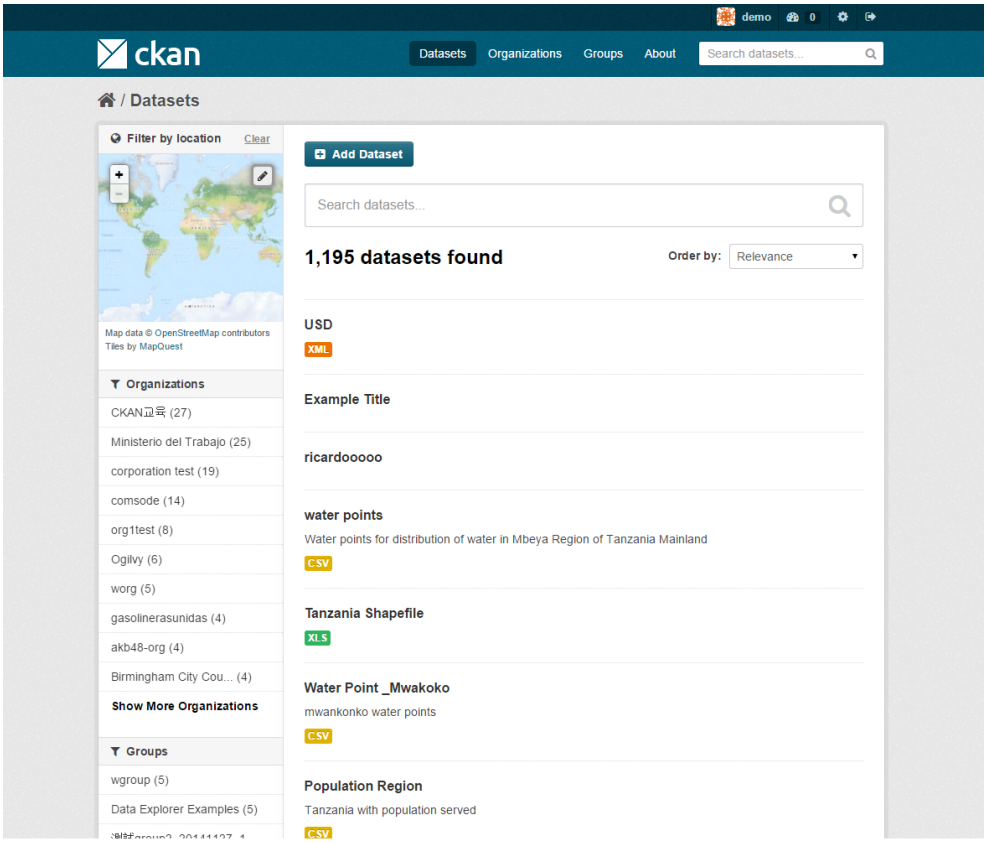
Příloha **B**

Obsah přiloženého CD

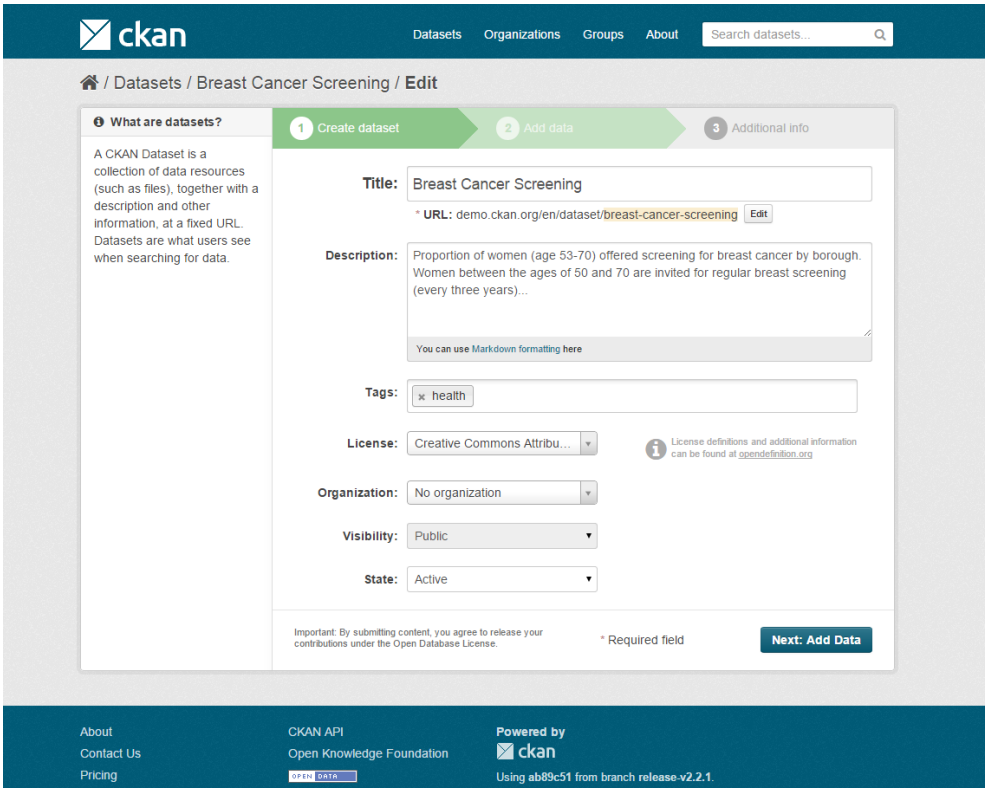
- `readme.txt` ..... stručný popis obsahu CD
- `src`
  - `BP_Smrz_Josef_2015.tex` .... zdrojová forma práce ve formátu LaTeX
- `text` ..... text práce
  - `BP_Smrz_Josef_2015.pdf` ..... text práce ve formátu PDF
  - `BP_Smrz_Josef_2015.ps` ..... text práce ve formátu PS

Příloha C

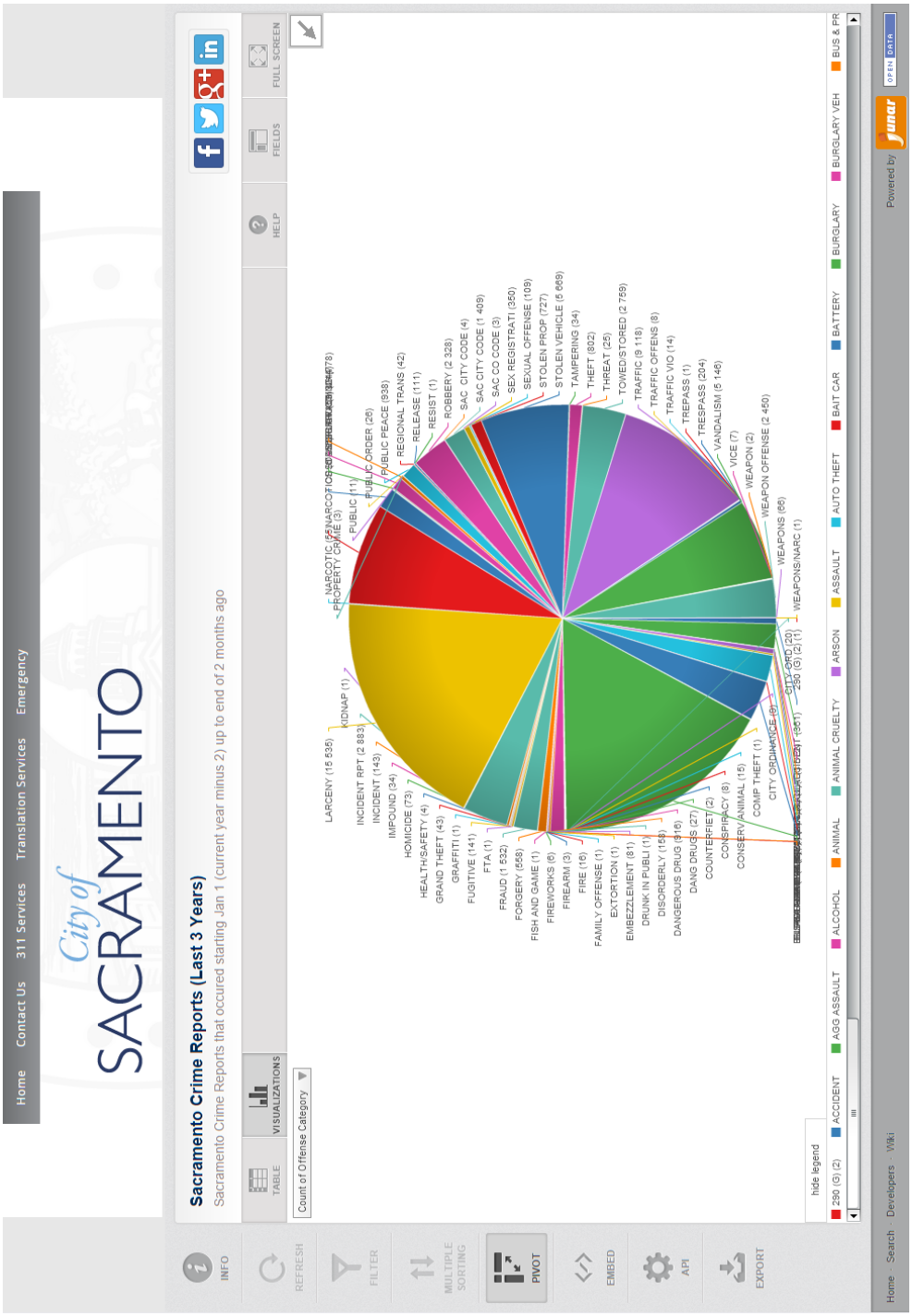
Obrazová příloha



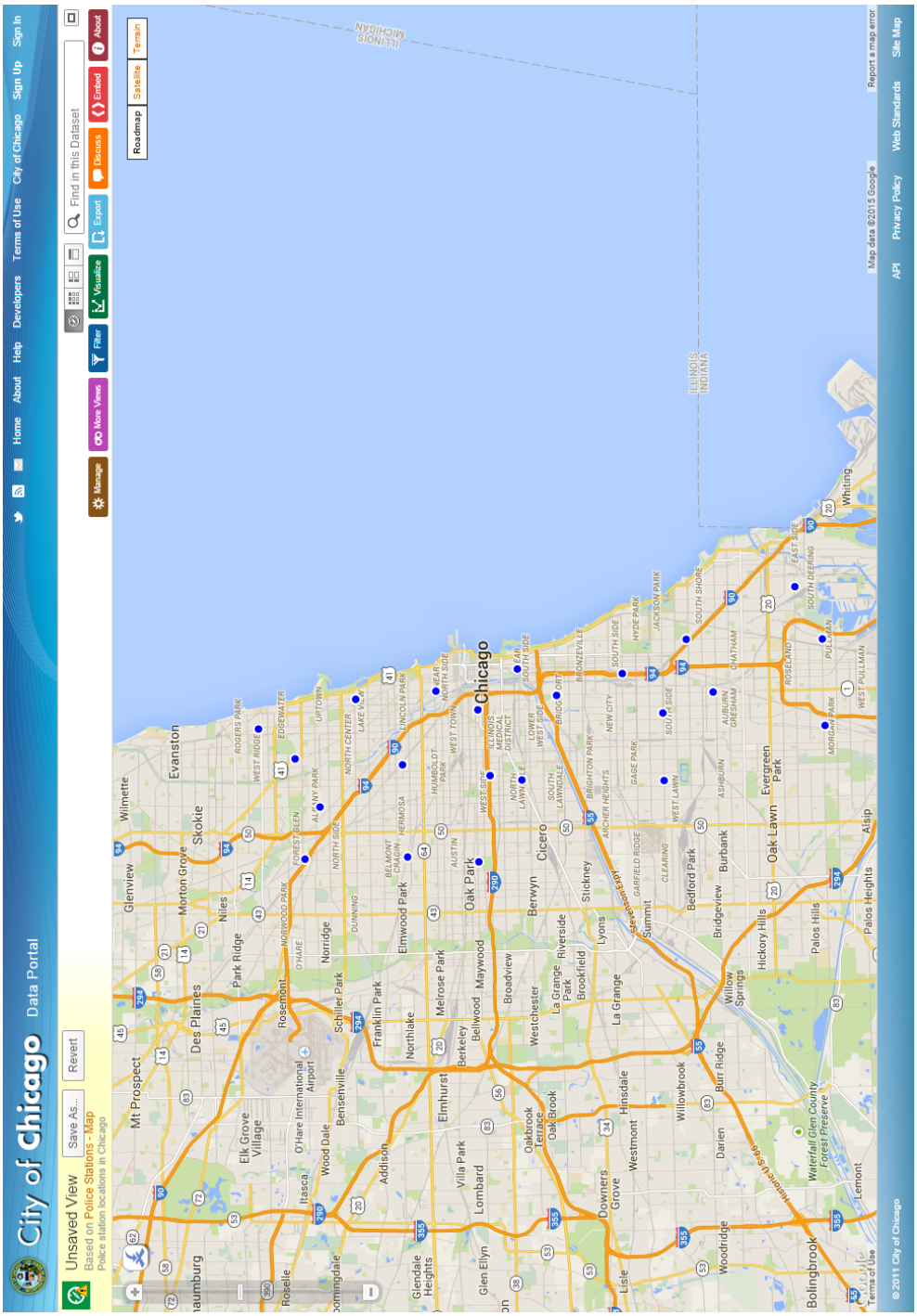
Obrázek C.1: Náhled obrazovky *katalogu* CKAN [3]



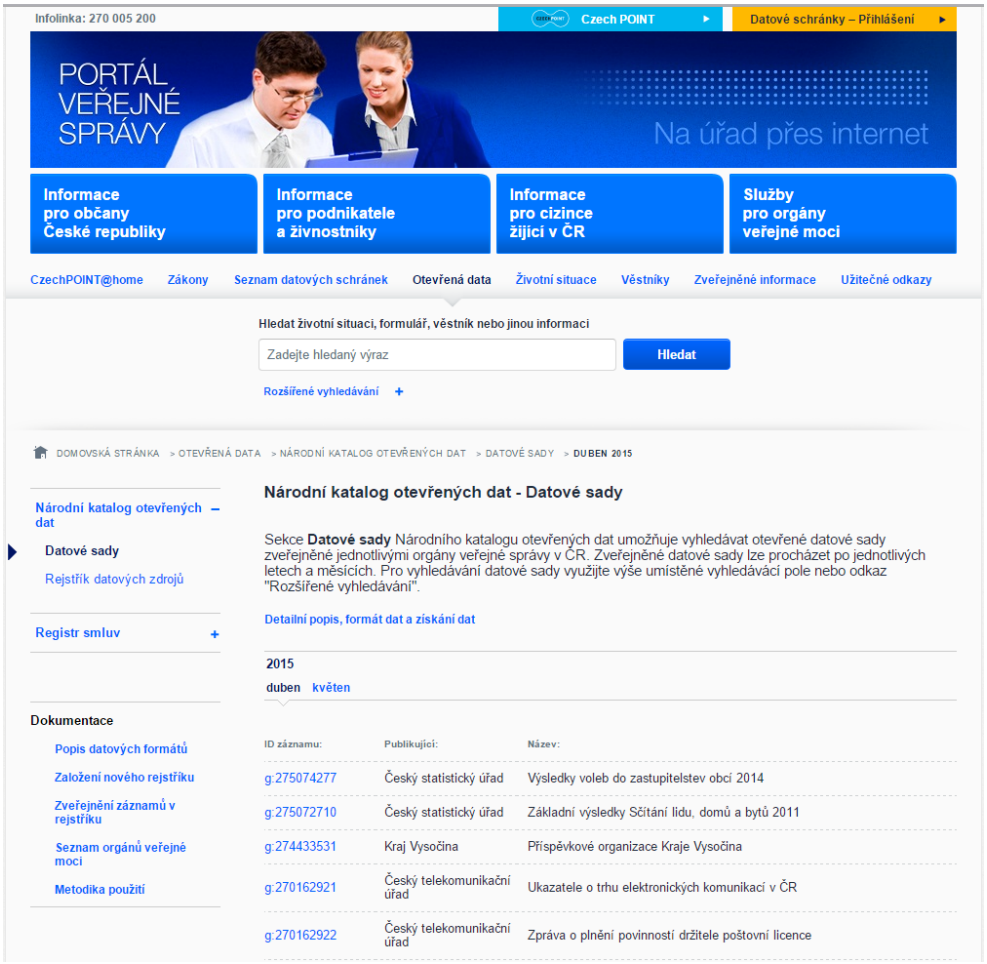
Obrázek C.2: Proces přidání *datasetu* do *katalogu* CKAN [4]



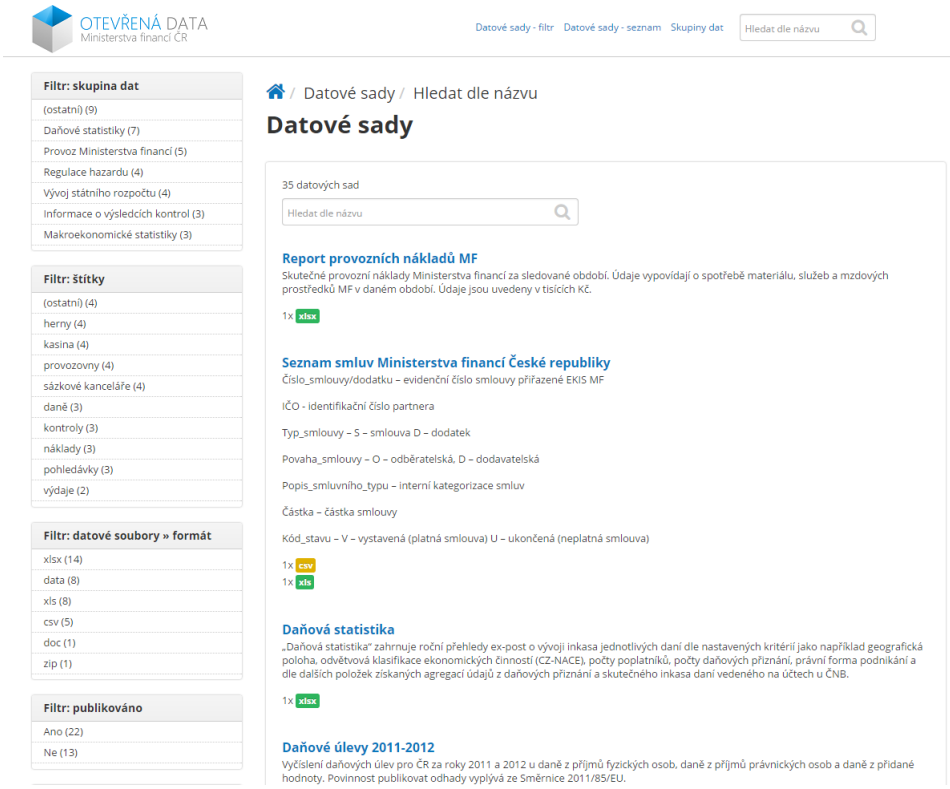
Obrázek C.3: Náhled obrazovky *katalogu* Junar používaného městem Sacramento [5]



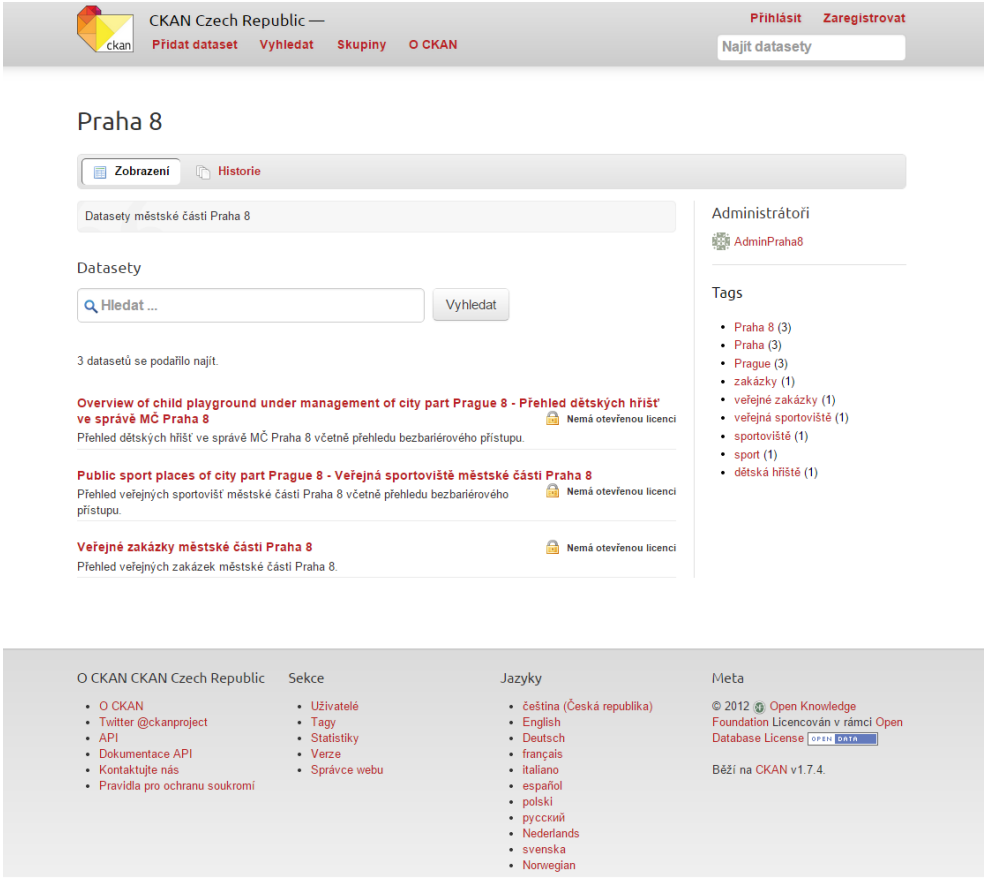
Obrázek C.4: Náhled obrazovky *katalogu* Socrata používaného městem Chicago [6]



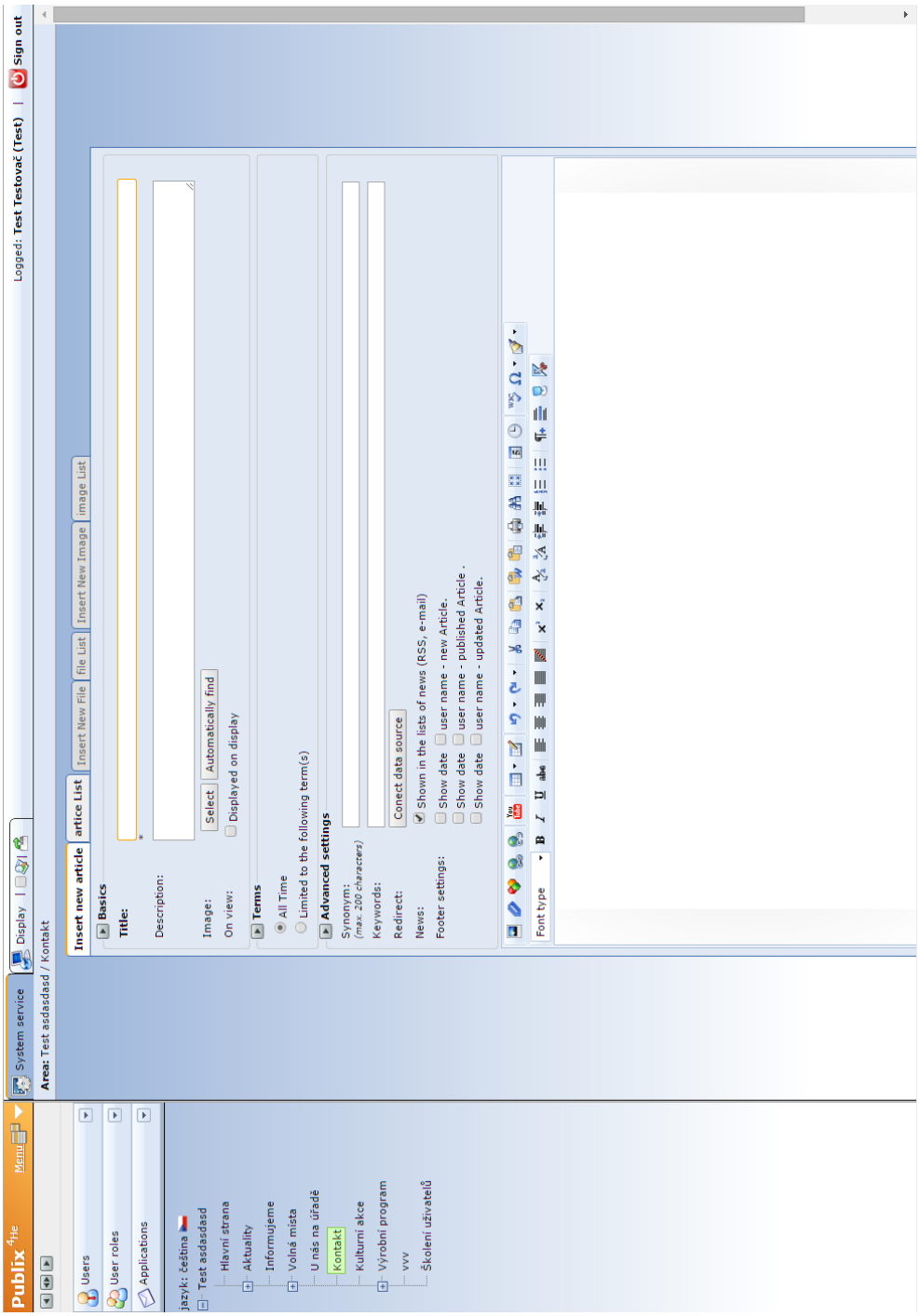
Obrázek C.5: Národní katalog otevřených dat [7]



Obrázek C.6: *Katalog* DKAN Ministerstva financí ČR [8]



Obrázek C.7: *Datasety* MČ Praha 8 v *katalogu* CKAN Czech Republic [9]



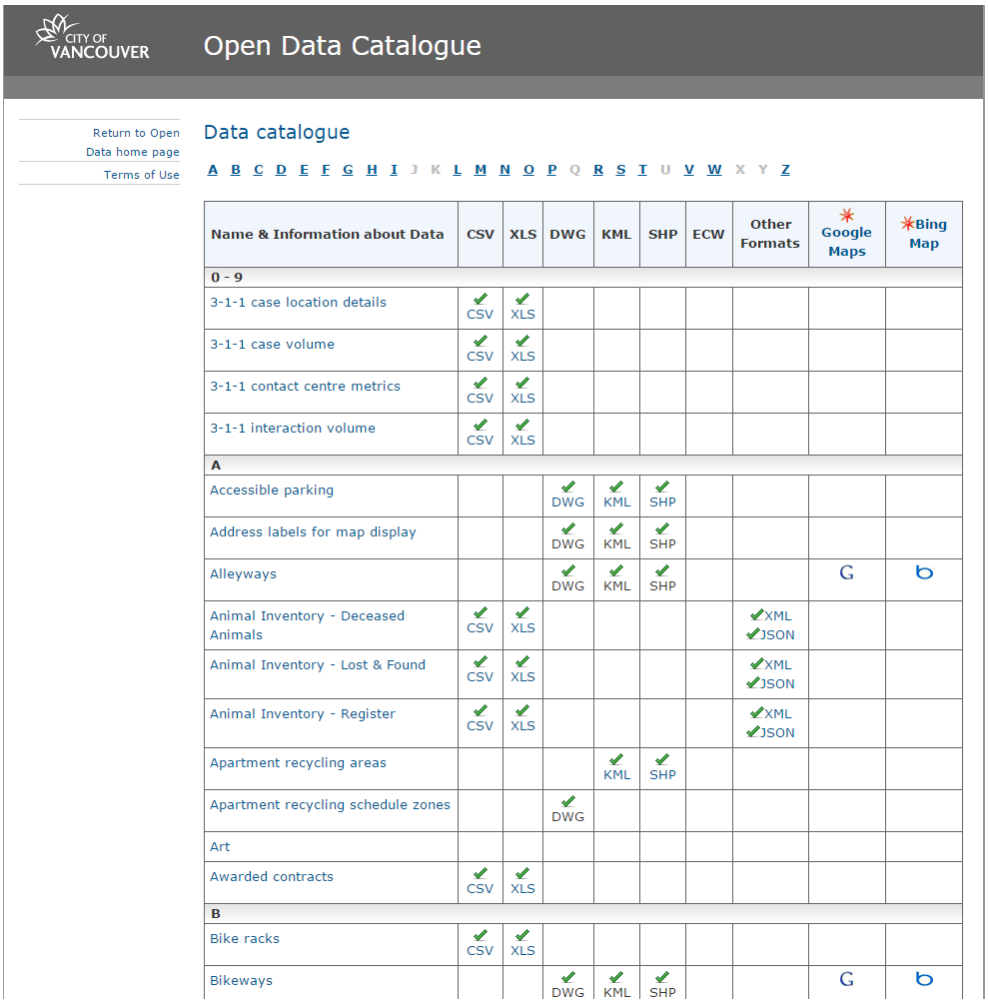
Obrázek C.8: Obrazovka přidání příspěvku do systému Publix [10]

Blokové čištění komunikací MČ Praha 8 v roce 2015

Blok:		1	Blok:		2		Blok:		
Termíny čistění:		8.4.	Termíny čistění:		9.4.		Termíny čistění:		
		27.10			26.10.				
Jirsíkova			Karlínské nám. (východní část)				Březinova		
Karlínské nám.(západní část)+parkoviště			Kolárova				Hybešova		
Křižíkova (Ke Štvanici - Peckova)			Křižíkova (Karl.nám. - Šaldova)				Kaizlovy sady		
Malého			Pernerova (Peckova - Šaldova)				Korybutova		
Peckova			Pobřežní(U nádr.lávky-Šaldova)				Křižíkova (Šaldova -Hybešova)		
Pernerova (Prvního pluku - Peckova)			Sokolovská (U nádr.lávky -Šaldova)				Kubova		
Pobřežní (Ke Štvanici - U nádr. Lávky)			Šaldova				Lyčkovo nám.		
Prvního pluku			Thámova				Na Střelnici		
Sokolovská (Ke Štvanici - U nádr.lávky)							Na Špitálsku		
U Nádražní Lávky							Pernerova (Šaldova - U Invalidovny)		
Vítkova							Petra Slezáka		
Za Poříčskou Branou							Sovova		
							U Invalidovny		
							Urxova		
Blok:		5	Blok:		6				
Termíny čistění:		16.4.	Termíny čistění:		15.4.		Blok:		
		21.10.			20.10.		Termíny čistění:		
Heydukova			Andrštova						
Na Žertvách			Bednářská				Braunerova		
Novákových			Horovo nám.				Budínova		
Palmovka			Chocholouškova				Chlumčanského		
Turnovská			Kotlaska				Kandertova		
U Balabenky			Koželužská				Kašparovo Náměstí		
U Pošty			Ludmilina				Konšelská		
V Mezihoří			Na Hrázi				Krejčího		
Vacínova			Na Rokytce				Lindnerova		
			Na Žertvách - parkoviště				Na Korábě		
			Náměstí dr.Holého				Nad Košinkou		
Blok:		9	Pivovarnická				Podlipného		
Termíny čistění:		22.4.	Pod Hájkem				Ve Vrchu		
		14.10.	Pod Kotlaskou						
K Okrouhlíku			Ronkova						
Ke Koupališti			Stejskalova						

bloky 1 - 17 | bloky 18 - 37 | Seznam ulic P-8

Obrázek C.9: Příklad špatně strukturovaných informací v *datasetu* s harmonogramem blokového čištění MČ Praha 8 [11]



CITY OF VANCOUVER

Open Data Catalogue

Return to Open Data home page

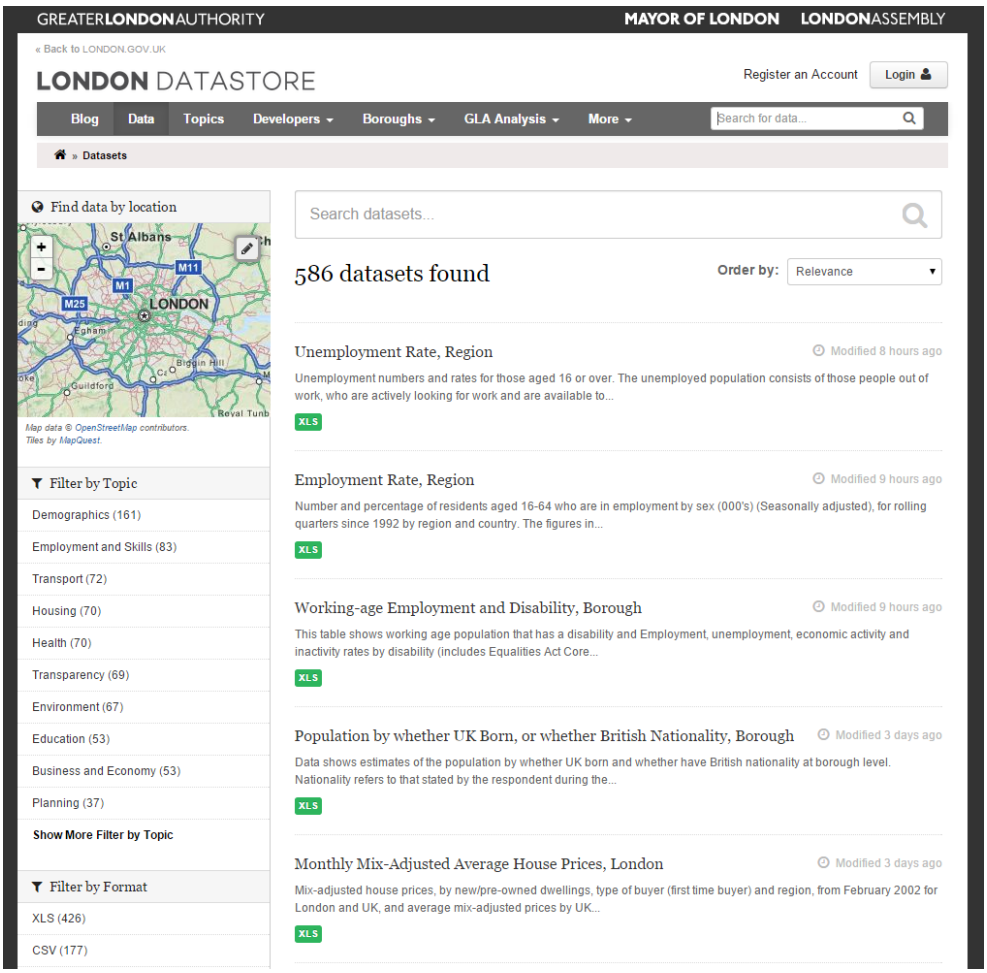
Terms of Use

Data catalogue

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Name & Information about Data	CSV	XLS	DWG	KML	SHP	ECW	Other Formats	Google Maps	Bing Map
0 - 9									
3-1-1 case location details	✔ CSV	✔ XLS							
3-1-1 case volume	✔ CSV	✔ XLS							
3-1-1 contact centre metrics	✔ CSV	✔ XLS							
3-1-1 interaction volume	✔ CSV	✔ XLS							
A									
Accessible parking			✔ DWG	✔ KML	✔ SHP				
Address labels for map display			✔ DWG	✔ KML	✔ SHP				
Alleyways			✔ DWG	✔ KML	✔ SHP			G	b
Animal Inventory - Deceased Animals	✔ CSV	✔ XLS					✔XML ✔JSON		
Animal Inventory - Lost & Found	✔ CSV	✔ XLS					✔XML ✔JSON		
Animal Inventory - Register	✔ CSV	✔ XLS					✔XML ✔JSON		
Apartment recycling areas				✔ KML	✔ SHP				
Apartment recycling schedule zones			✔ DWG						
Art									
Awarded contracts	✔ CSV	✔ XLS							
B									
Bike racks	✔ CSV	✔ XLS							
Bikeways			✔ DWG	✔ KML	✔ SHP			G	b

Obrázek C.10: *Katalog otevřených dat* města Vancouver [12]



Obrázek C.11: *Katalog otevřených dat* města Londýn [13]