

OpenData Praha

Ondřej Profant a kolektiv

2. listopadu 2015

Abstrakt

OpenData (Otevřená data) jsou světovým trendem, který napomáhá otevřenosti ve veřejné správě pomocí poskytnutí strojově čitelných a zdokumentovaných primárních dat. Významným využitím je též business reuse. Mnoho podnikatelů umí data kreativně využít. Velkou uživatelskou skupinou je akademická obec pro kterou jsou tvrdá data nezbytnost.

Obsah

Opendata Praha - studie	3
Záměr	3
Proč OpenData?	3
Technické řešení	7
Licence	9
Datové sety	9
Kvalita dat (pětihvězdičkový model)	9
Rozvoj	9
Seznam použitých zkratk	9
Zdroje	11
Přílohy	11
Implementace	11
Návod: CKAN základní práce	13
Co je to CKAN?	13
Správa organizace	13
Logo	13
Jak zveřejnit data?	13
Dataset	13
Zdroj	14
V jakém formátu uveřejnit?	16
Jak ta data získám?	16

Návod: automatizace	17
1. Získat data	17
2. Nahrát data	17
Rozvoj	17
Návod: získání dat ze softwaru Ginis	18
Úvod	18
Struktura software Ginis	18
Výběr knihy	19
Vlastní filtry – flagy a masky	19
Jak těžit data	20
Formát CSV - Comma separated values	21

Opendata Praha - studie

Záměr

V programovém prohlášení Rady hl. m. Prahy¹ i komise ICT RHMP² jsou OpenData (a dokonce i OpenSource) silně podpořena. Navíc jsou k realizaci přizváni Piráti a spolupracuje se s ministerstvem financí (Lukáš Wagenknecht). Z toho vyplývá, že je opravdu velká šance k úspěšné realizaci otevření pražských dat.

Tento dokument je úvodním stanovením toho, co zveřejnovat, jak to technicky provést a odkud data získat (zodpovědná osoba, informační systém etc.).

Též v úvahách zahrneme jednotlivé městské části (MČ), jelikož počítáme s projektem celopražských otevřených dat.

Budu velmi rád za věcné připomínky, nápady a i další způsoby věcné kritiky textu (nebojte se napsat, že něčemu nerozumíte! Chyba je většinou v překladu).

OpenData musí být:

1. Veřejná (dostupná, bez licenčního zatížení)
2. Strojově čitelná (otevřený formát, strojově zpracovatelný)
3. Uspořádaný v daném schématu (přístup přes dané rozhraní - API, to zajistí celoměstský portál)
4. zpřístupněna za jasně definovaných podmínek užití dat s minimem omezení

Proč OpenData?

Otevřená data je myšlenka, že by k určitým datům mohl mít přístup každý, mohl je sdílet a mohl s nimi dále pracovat. Např. tvořit statistiky, datamining apod. Což může napomoci veřejné kontrole a dokonce i komerčnímu sektoru. Důsledné dodržování metodiky navíc pomůže i samotné instituci, která dnes mnohdy nemá klíčová data ve vhodném formátu (zvláště pak starší data).

Díky otevřenosti mohou vznikat desítky nástrojů, která např. data vizualizují a podobně, což jsou obvykle formy zpracování, které úřad samotný nedělá.

Krásným příkladem je, že navrhované řešení umí zobrazit údaje na mapě. Čili pokud zveřejníme seznam škol, který bude obsahovat údaje o poloze, tak máme automaticky i mapu s vyznačenými školami. A to je samozřejmě jen zlomek možných využití. Vizualizace jsou přiloženy v příslušné kapitole dále.

Více: <http://www.otevrenadata.cz/otevrena-data/>

Londýn Velmi kvalitním příkladem otevřených dat je **Londýn**, který má dokonce tvrdá čísla k prospěšnosti OpenDat.

Děčín Děčín byl jedním z prvních průkopníků otevřených dat v ČR. Bohužel zájem o ně není příliš velký. Což je způsobeno dvěma faktory: 1) atraktivita severočeského města není taková jako světových metropolí, 2) vývojáři s otevřenými daty nepočítají, čili nevznikají aplikace.

Na druhou stranu můžeme říci, že to Děčín nějak nezatížilo, nepřineslo problémy ani nic podobného.

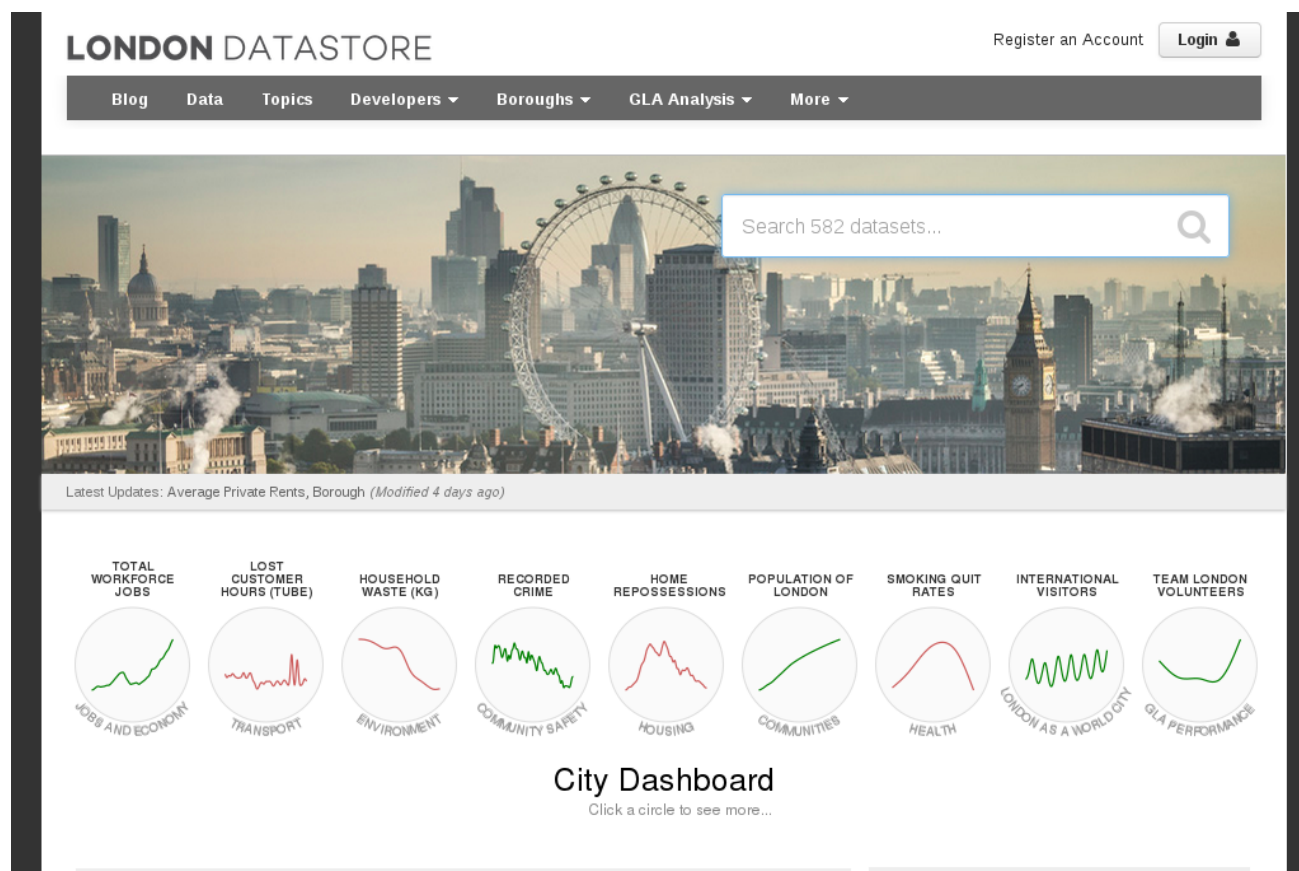
Výraznější úspěch měla **vizualizace** rozložení cen veřejných zakázek, která zcela jednoznačně ukázala, že většina zakázek se zadává pod limitem.

Děčín v současnosti přechází z vlastního řešení na DKAN.

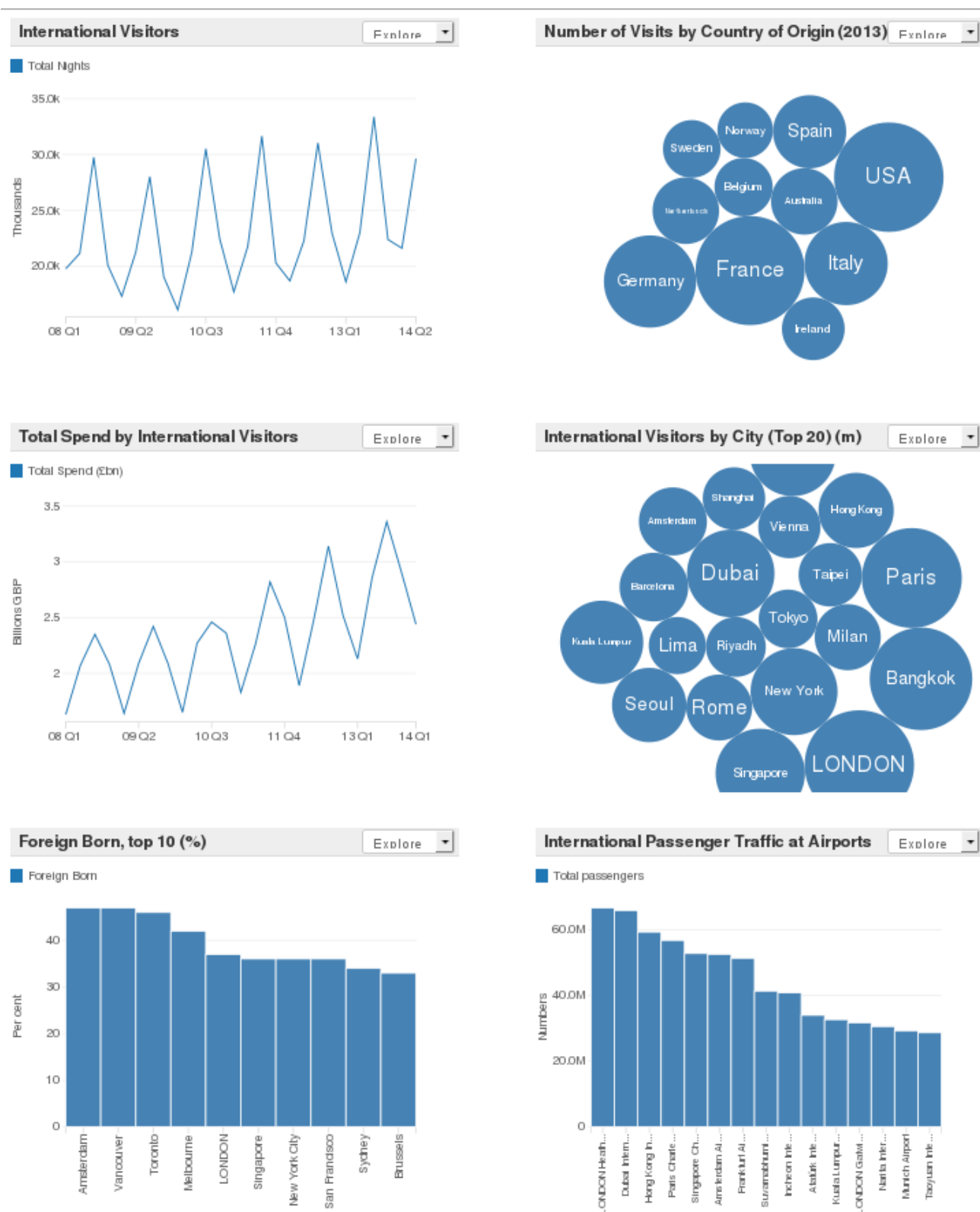
Více: <http://www.mmdecin.cz/component/k2/item/552>

¹Programové prohlášení RHMP.

²Komise ICT RHMP schválila projektový záměr opendat 16. 5. 2015

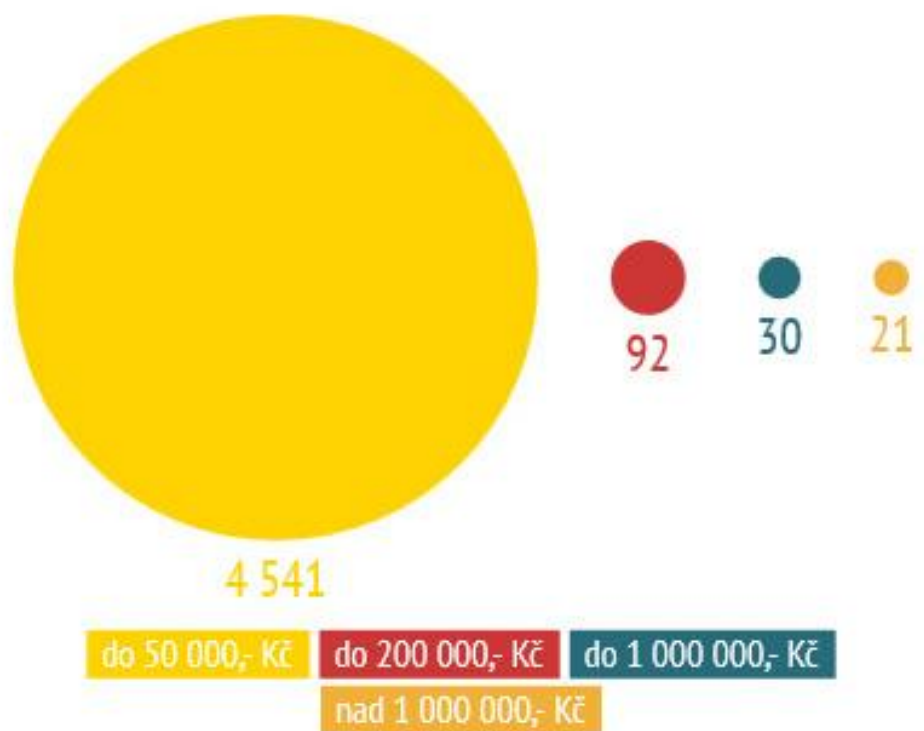


Obrázek 1: Úvodní stránka londýnských OpenDat



Obrázek 2: Turistika v Londýně (vybraná část, data jsou rozsáhlejší)

Počet zakázek dle jejich výše



Obrázek 3: Rozložení počtu veřejných zakázek v Děčíně v roce 2013. Vidíme, že většina je do 50 000 Kč (čili bez povinnosti oslovit více dodavatelů).

IPR Praha Institut plánování a rozvoje Praha otevřel geografická data Prahy, která předtím prodával. Díky tomu je volně dostupné velké množství geografických dat v několika různých formátech.

Více: <http://opendata.iprpaha.cz>

Technické řešení

Řešením **MFČR** a **MSČR**, což je opensource CMS **Drupal** s modulem **DKAN**, se pouze inspirujeme. Tento systém bohužel nemá zapisovatelné API, které je třeba pro budoucí automatizaci.

Namísto toho zvolíme principiálně a funkčně velmi podobné opensource řešení **CKAN**.

Takto postavený systém nevyžaduje žádné extrémní nároky (stačí běžný LAPP: Linux, Apache, **Python 2**, **PostgreSQL**, **SOLR** - engine nad **Lucene**, design **Bootstrap**), je podporovaný komunitou v dostatečném rozsahu. Zároveň je pro uživatele dostatečně přívětivý.

Jedna instance bude stačit pro MHMP, MČ i příspěvkové organizace. Technickým správcem bude odbor informatiky MHMP (INF). Věcným správcem zatím není určen.

Výhodou je, že tím na MČ nepřeneseme téměř žádné nároky. Navíc data budou na jednom místě (s množstvím dat stoupá jejich atraktivita).

Jediné náklady budou na provoz serveru a čas zaměstnanců. Vzhledem k rozsahu IT infrastruktury MHMP to jsou však zcela zanedbatelné náklady.

Je samozřejmě třeba dořešit věci jako monitorování přístupů (oblíbenosti, využívání) a jak tato data distribuovat zpět dalším subjektům.

Formáty a typy dat Nejprve si definujeme 3 skupiny dat:

1. **Souhrnná data:** Seznamy údajů vhodné pro statistické zpracování. Např. seznam faktur, seznam škol, ...
2. **Vnitřní tvorba:** Textové dokumenty tvořené přímo v rámci úřadu. Např. výroční zprávy odborů, právní stanoviska. Zde by bylo dobré, kdyby se z dokumentů vypreparovala statisticky použitelná data (tabulky) a ta se zveřejnila v rámci kategorie I.
3. **Vnější tvorba:** Dokumenty zpracované u dalších subjektů. Např. audity, výroční zprávy etc.

Jak vidíme, tak jen kategorie I. jsou data určená pro rozsáhlý datamining. Zatímco kategorie II. a III. jsou zpracovatelné jen pomocí fulltextového vyhledávání. Nicméně i to je ohromný krok vpřed.

Technické řešení (CKAN) přijímá data ve formátech CSV, XSLX, XML, ODF, DOCX, popřípadě PDF. Pro kategorie I. je ideální CSV, pro kategorii II. a III. pak DOCX, strojově čitelné PDF (dle toho, co lze sehnat).

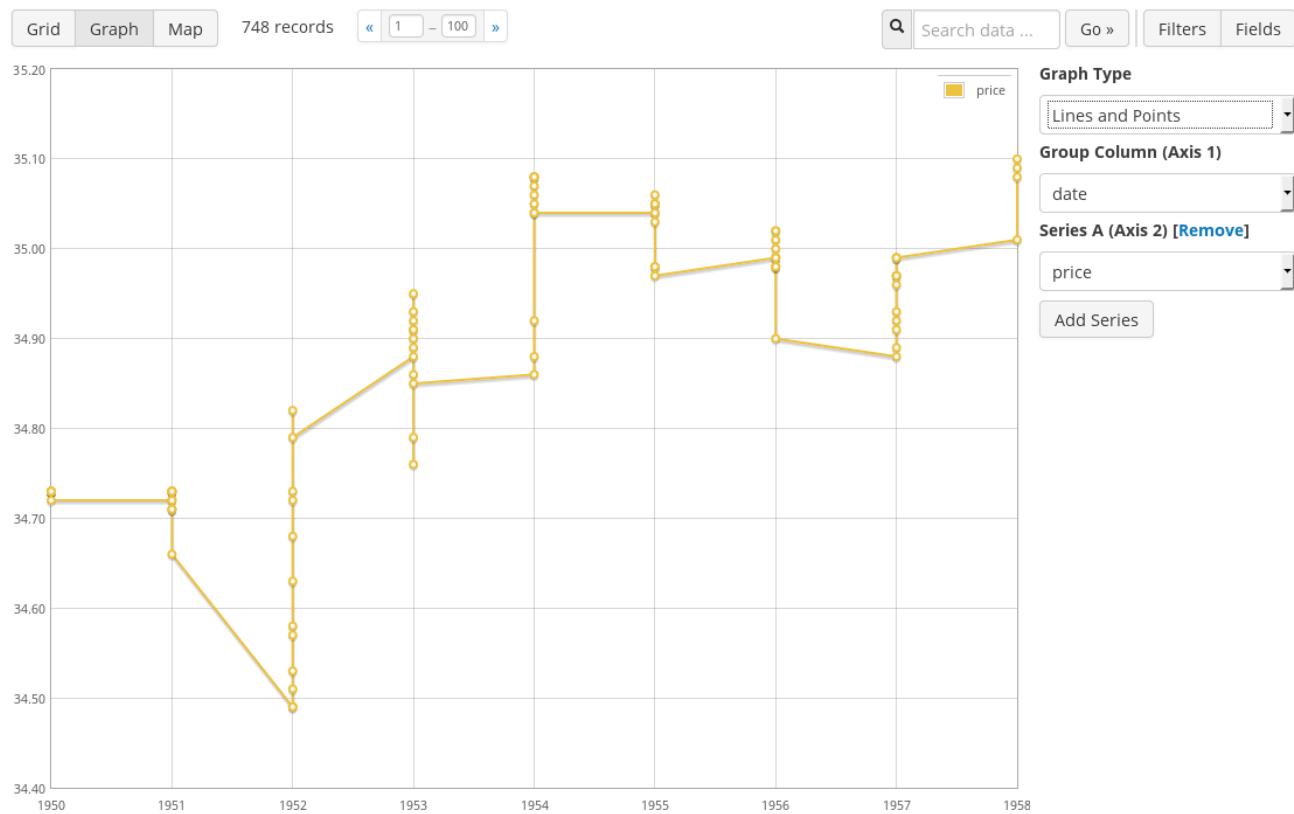
Výstupní formát je shodný se vstupním (nahraju-li csv, dostanu csv). Navíc však CKAN zpracuje u vybraných formátů dle příslušných ontologií formáty RDF a JSON, což jsou nejpoužívanější a nejpraktičtější formáty pro OpenData. DKAN též umí data vizualizovat. Buďto pomocí různých druhů grafů, anebo vyznačit na mapě.

DKAN se automaticky stará o persistentnost odkazů, verzování a vytváří API dle zadaných pravidel.

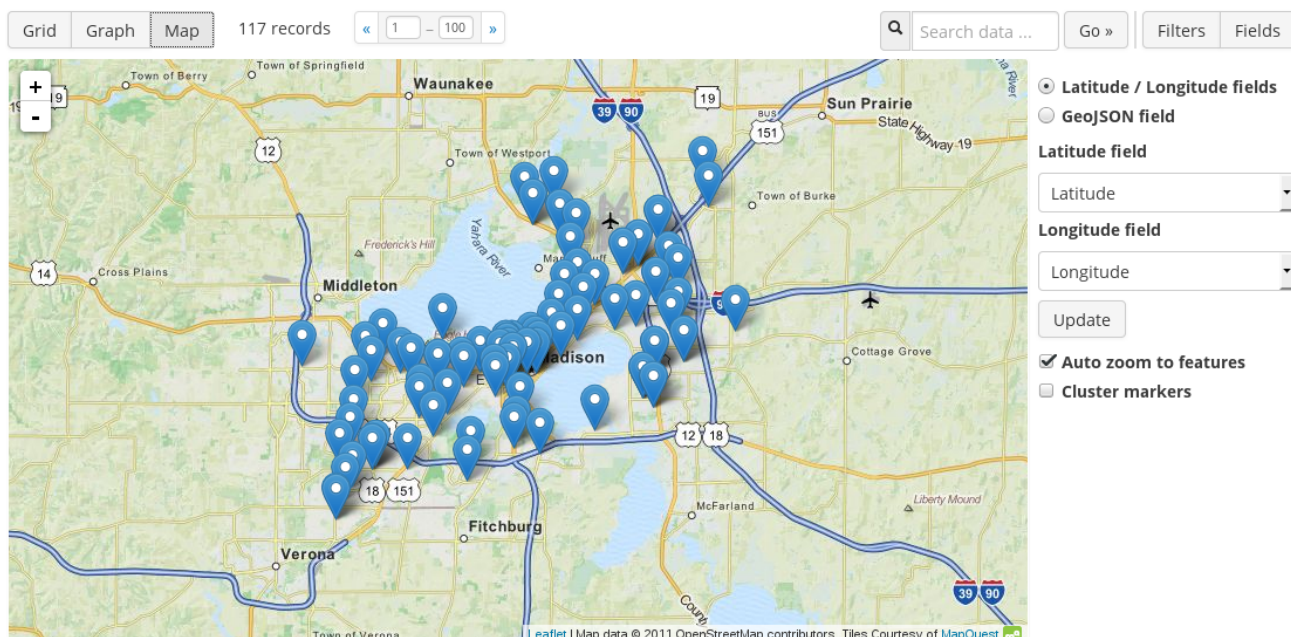
Vizualizace Ukázky defaultní vizualizace dat v modulu CKAN.

Workflow Workflow opět využijeme stejné jako na MFČR. Namísto automatizace zaneseme interně do povinností kurátorů datových sad, aby data zkompletovali a nahráli v daném termínu. V praxi se nebude jednat o velkou zátěž, jelikož většina dat vzniká prostým exportem z IS (cca. 4 kliknutí, následně nahrání daného souboru). Popřípadě se dokonce data odesílají již dnes, ale pouze interně (což lze následně zrušit a interně čerpat z OpenDat).

Do budoucna se jistě počítá s automatizací tam, kde to bude možné. Ale v současnosti nemá cenu se tím zdržovat. Instituce i odbory samozřejmě mohou práci optimalizovat dle uvážení - mají přístup přes webové rozhraní i skrz API.



Obrázek 4: Vizualizace cen zlata v jednotlivých letech. Zdrojem dat je prostý CSV soubor.



Obrázek 5: Vizualizace polohy volebních místností v Madisonu. Zdrojem dat je prostý CSV soubor.

Licence

Je nezbytné, aby data byla uvolněna pod nerestriktivní licenci. MFČR má podmínky různé pro každou sadu. Nicméně většina z nich je uvedena jako “Volné dílo”.

Více: <http://opendatacommons.org>

Datové sety

Primárně chceme zveřejnit již dnes zveřejněná data, popř. data, která všichni považovali za nezajímavá. Tím, že data budou strojově čitelná, vznikají zcela nové možnosti, jak s nimi pracovat.

Datové sady též podrobně předpřipravilo [ministerstvo vnitra](#).

Kvalita dat (pětihvězdičkový model)

Pětihvězdičkový model je dnes již klasický model hodnocení kvality otevřených dat. Značí (každý další stupeň obsahuje předešlé):

1. Jakákoliv data pod otevřenou licencí (sken)
2. Strukturovaná data (excel)
3. Strukturovaná data v otevřeném formátu (csv, xslx)
4. Každý datový set má URI / systém obsahuje API
5. Data jsou prolinkovaná (obsahují kontext)

Praha aspiruje na 4. až 5. hvězdu. Propojení zatím bude jen formou jednoznačných identifikátorů (ID) v daném datasetu.

Více: <http://5stardata.info>

Rozvoj

Nejméně v prvním roce bude třeba rozšiřovat datové sady a dohlížet na jejich řádné propojení. Též bude třeba komunikovat s dalšími institucemi, které mají cenná data (např. MČ, DPP, ROPID ...), aby je otevřely.

Automatizace Automatizace je přirozený další krok. Je však náročný v tom, že je potřeba vytvořit mnoho konektorů na různé systémy. Nicméně tento krok lze rozložit do delšího časového období, dle priorit.

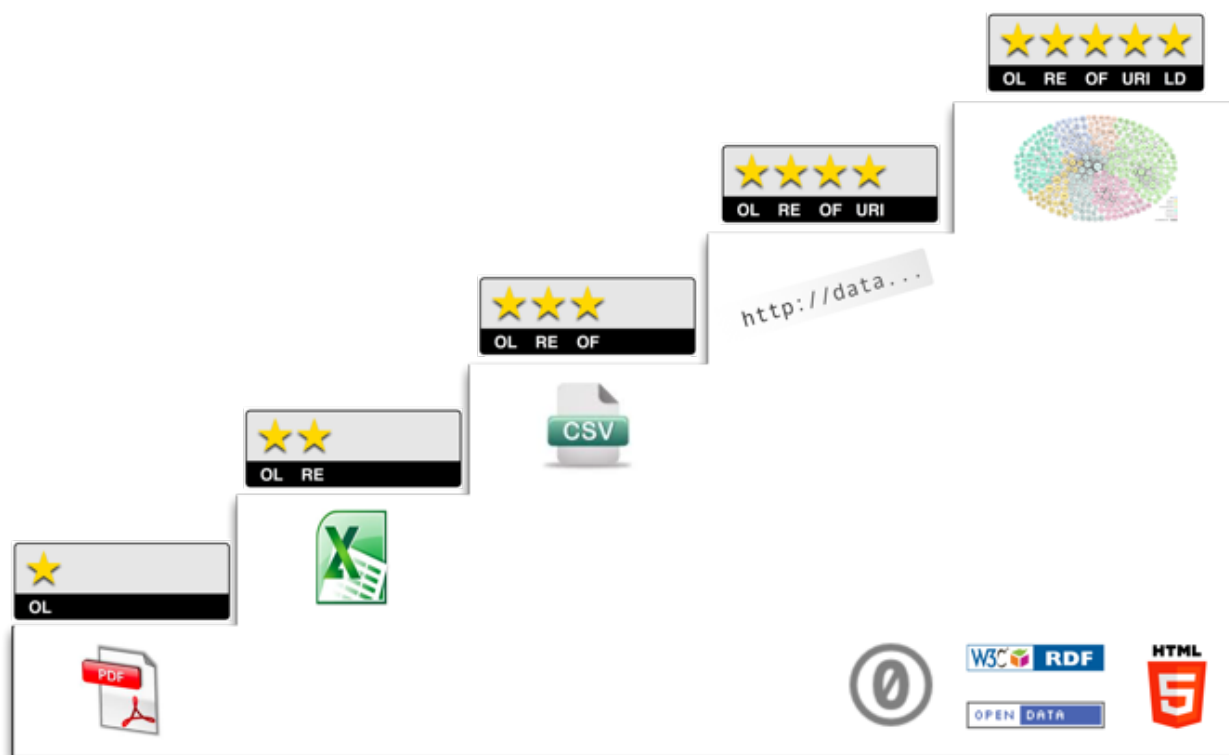
Smlouvy Státní instituce dnes smlouvy nad 500 000 Kč povinně zveřejňují na [portálu veřejné správy](#). Příslušný zákon a s ním i registr se v současnosti novelizuje. Do budoucna se předpokládá stejná povinnost pro samosprávy.

V rámci projektu Otevřená města vzniká opensource centrální registr smluv, který je přímo zaměřen na potřeby obcí. Obsahuje přímo specifikaci metadat smlouvy etc. Hotový by měl být v létě 2015. Předpokládá následnou migraci do centrálního státního registru.

Národní katalog MVČR připravuje metodiku (definici datových setů) pro orgány státní správy. Do budoucna snad i samosprávy. Z předběžných jednání je jasné, že bude počítat se systémy DKAN a CKAN (data se mají centrálně sbírat do jednoho katalogu).

Stejně tak lze uvažovat o zapojení do celoevropských otevřených dat jako např. [PublicDataEU](#).

Seznam použitých zkratek



Obrázek 6: Pětihvězdičkový model Otevřených dat

Zkratka	Význam	Poznámka
HMP	Hlavní město Praha	
MHMP	Magistrát HMP	
ZHMP	Zastupitelstvo HMP	
RHMP	Rada HMP	
MČ	Městská část	
ÚMČ	Úřad MČ	
MP	Městská policie	
MFČR	Ministerstvo financí ČR	
MVČR	Ministerstvo vnitra ČR	
MSČR	Ministerstvo spravedlnosti ČR	
TSK	Technická správa komunikací	Je držitelem většiny informací o automobilové dopravě v Praze
PID	Pražská integrovaná doprava	
ROPID	Regionální organizátor PID	
DPP	Dopravní podnik Praha a.s.	
IPR	Institut plánování a rozvoje	Bývalý ÚRM (útvár rozvoje města)
INF	Odbor informatiky MHMP	
OVO	Odbor volených orgánů MHMP	


Zkratka	Význam	Poznámka
CMS	Content Managment System	Systém pro správu obsahu webových stránek
LAMP	Linux Apache Mysql PHP/Python	Velmi běžná kombinace nastavení serveru pro webové aplikace.
LAPP	Linux Apache PostgreSQL PHP/Python	Běžná kombinace nastavení serveru pro webové aplikace\ (obecně vzato
IS	Informační systém	

Zdroje

- Otevřená data: <http://www.otevrenadata.cz>
- Pět hvězdičkový model kvality OpenDat: <http://5stardata.info>
- Licence k OpenData: <http://opendatacommons.org>
- CKAN: <http://ckan.org>
- Drupal: <http://drupal.com>
- DKAN: <http://nucivic.com/dkan>
- DKAN demo: <http://demo.getdkan.com>
- OpenData MFČR: <http://data.mfcr.cz>
- Public data EU: <http://publicdata.eu>
- Portál veřejné správy: <https://portal.gov.cz>
- London OpenData: <http://data.london.gov.uk>
- OpenCity App: <http://opencityapps.org>
- Prezentace k OpenData v Praze: <http://www.slideshare.net/ondrejprofant/open-data-praha>
- Velmi pěkně zpracované souhrné PDF: <http://bit.ly/1NCDipy>
- Data v dopravě: <http://www.ioda.cz> a <http://www.vyrocenky.cz>

Přílohy


Implementace




[Datové sady - filtr](#)
[Datové sady - seznam](#)
[Skupiny dat](#)

Vítejte na portálu Otevřená data Ministerstva financí ČR!
Najdete zde katalog datových sad, které Ministerstvo zveřejňuje nebo plánuje zveřejnit.

Datové sady můžete buď [filtrovat](#), [procházet po skupinách](#) nebo vyhledávat pomocí klíčových slov:




Otevřená data MFČR @otevrenadatamf 20. úno.
Právě jsme zveřejnili největší dodavatele ICT a čerpání finančních prostředků na ICT za resorty MF, MO, MMR, MZP, MSp a MD.
Rozbalit




Otevřená data MFČR @otevrenadatamf 17. úno.
K seznamu faktur je nově zveřejněn i Seznam partnerů MF, kde je možné dohledat IČ dodavatele!
Rozbalit

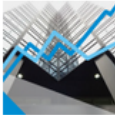
[Sledovat uživatele @otevrenadatamf](#)




Daňové statistiky
Daňová statistika
Daňové úlevy 2011-2012
...




Informace o výsledcích kontrol
Kontroly v oblasti práv duševního vlastnictví
Kontrolní činnost v systému elektronického mýtného...
Kontroly v oblasti zaměstnávání cizinců
...




Makroekonomické statistiky
Fiskální výhled ČR
Makroekonomická predikce
...




Provoz Ministerstva financí
Report provozních nákladů MF
Přehled faktur Ministerstva financí
Seznam smluv Ministerstva financí České republiky
...




Regulace hazardu
Přehled povolených technických herních zařízení
Obecně závazné vyhlášky obcí regulující loterie a jiné...
Výsledky z provozování loterií a jiných podobných her
...



Vývoj státního rozpočtu
Státní rozpočet
Čistá pozice ČR vůči EU
Vládní finanční statistika GFS 2001
...



Monitor Státní pokladny
Monitor zpřístupňuje veřejnosti finanční a účetní informace všech organizací veřejné správy ČR.



Nepublikované datové sady
Výpis datových sad, u kterých se v současné době analyzují možnosti zveřejnění.

Obrázek 7: Pohled na úvodní stránku opendat MFČR

Návod: CKAN základní práce

Co je to CKAN?

CKAN je v současné době nejrozšířenější open source technologií, užívanou pro zveřejňování otevřených dat. CKAN je ve své podstatě speciální webový portál, kam se data nahrávají. Využívá jej i hlavní město Praha. Má dvě části.

Jeho veřejná část je přístupná na adrese `opendata.praha.eu`. Pro nahrání však potřebujete využít neveřejnou část přístupnou pouze ze sítě mepnet na adrese `private.opendata.praha.eu`

Portál otevřených dat nabízí tzv. datasety, tedy sady/složky dat. Tyto sady pak tvoří jednotlivé soubory v různých formátech. Jaké a proč je vysvětleno dále.

Správa organizace

CKAN je organizován po institucích (organizacích). V našem případě se jedná o městské části, příspěvkové organizace a korporace vlastněné Hl. m. Prahou. Každá organizace má v rámci systému vlastní datasety, spravuje si svoje vlastní uživatele etc.

Organizace má především tyto editovatelná pole:

- Název
- Popisek
- Metadata: klíč-hodnota, libovolný počet
- Logo

Dále vlastní datasety, které přidali uživatelé jenž jsou členy dané organizace. S profilem správce organizace můžete zakládat další uživatele v této organizaci.

Logo

Logo je obrázek s rozlišením: 189 × 189px. Výšku lze použít i jinou.

Jak zveřejnit data?

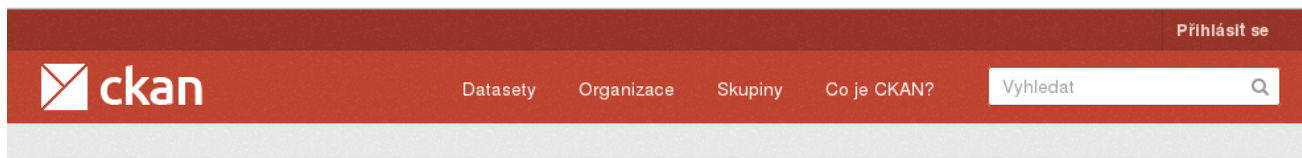
Cílem Prahy je nabídnout co nejvíce otevřených dat ze všech oblastí. Proto se tento návod zabývá tím, jak data do tohoto portálu přidávat.

Dataset

Dataset je jakási složka. Obsahuje název, popis, metadata (např. licenci) a url. Slouží k základní orientaci v datech. Slouží např. k logickému sloučení dat v různých formátech (typicky mapové podklady se dodávají v různých formátech) a doplňujících datech (podrobná legenda, číselníky).

Url datasetu lze měnit. Standardně se použije název upravený do podoby vhodné pro webovou url. Např. jsou transformovány háčky, mezery etc. Vhodné je prefixovat url zkratkou vaší organizace a pokud je název dlouhý, tak není problém ho v url zkrátit.

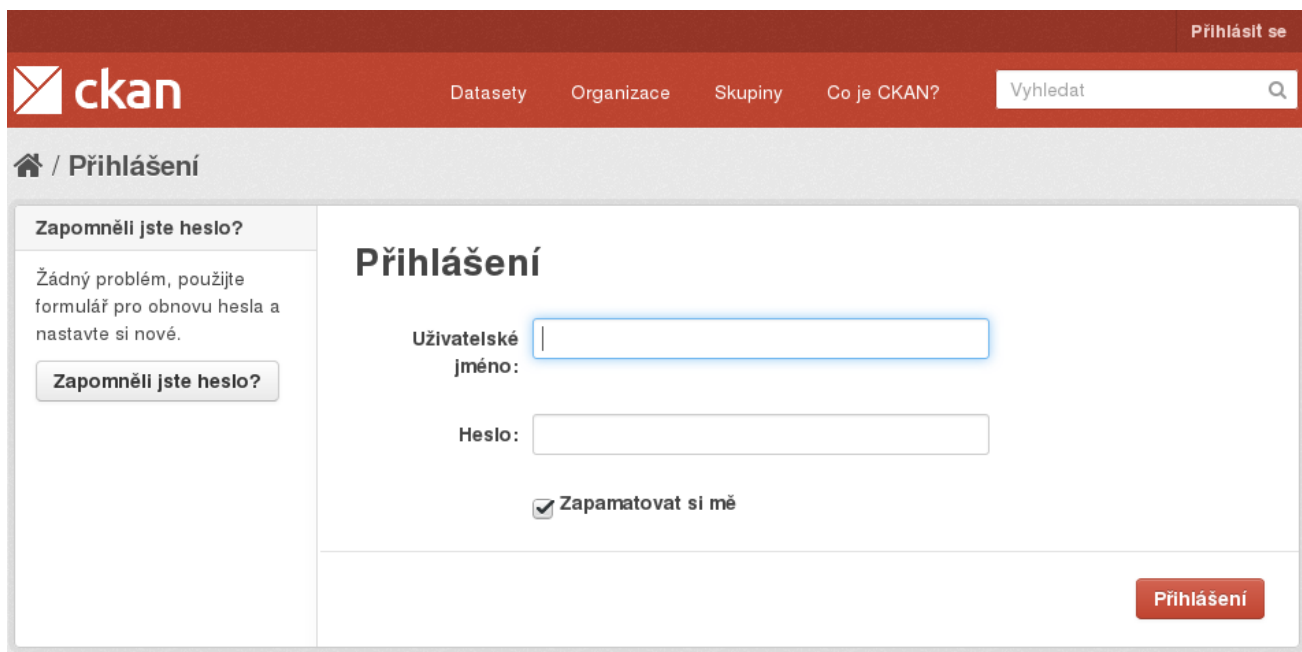
Dataset by neměl obsahovat víc jak 100 zdrojů. Na druhou stranu dataset s 1 zdrojem je obvykle též špatně. Optimální je např. “Výsledky hlasování zastupitelstva za rok 2014” a jednotlivé zdroje by pak měly být “1. zasedání”, ... Popřípadě může být dataset (zvláště v případě menší MČ) za celé volební období.



Obrázek 8: Hlavní menu, vpravo přihlašování

Vytvoření datasetu

1. Vpravo nahoře je tlačítko „Přihlásit se“ (v anglické verzi „Login“).
2. Zobrazí se stránka s přihlašovacím dialogem. Zde zadejte své přihlašovací údaje.



Obrázek 9: Přihlašovací dialog

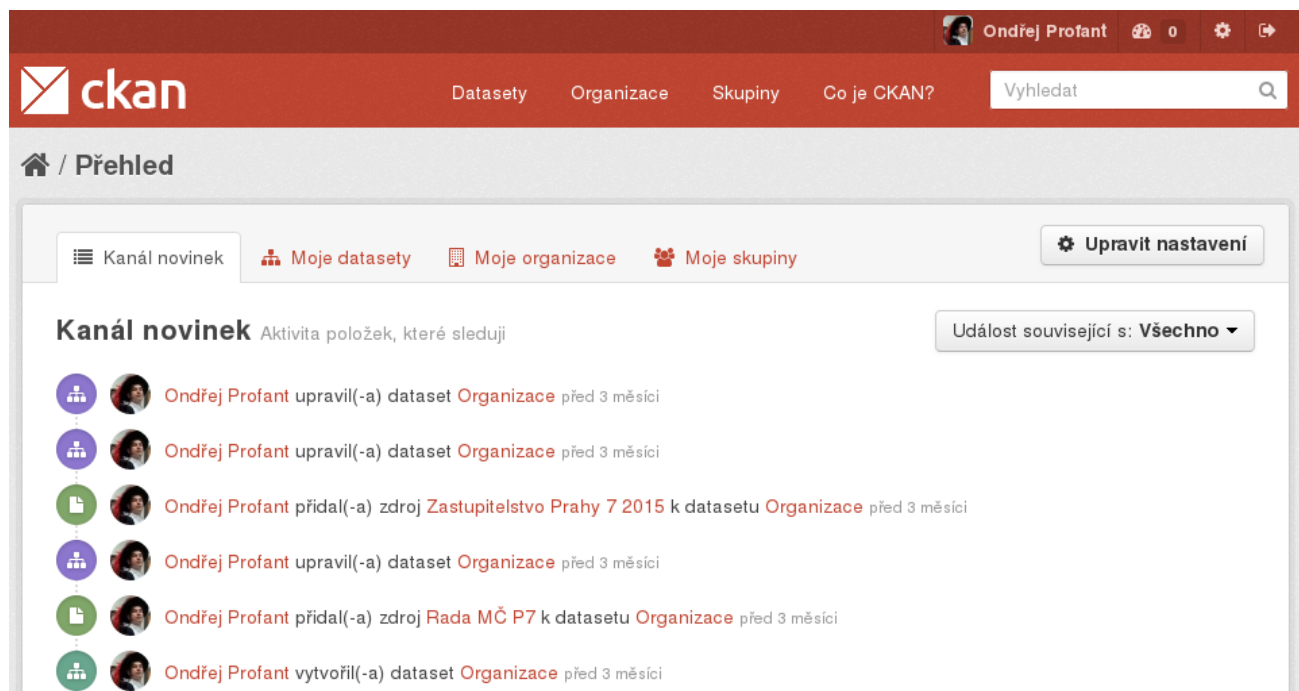
3. Po úspěšném přihlášení se vám zobrazí „Kanál novinek“.
4. Nový dataset založíte na panelu (záložce, tabu) vedle „Moje datasety“.
5. Při vytvoření datasetu musíte zadat nezbytná metadata. Metadata popisují společné znaky daného datasetu.

Dataset je sada/složka, kde se shromažďují dílčí zdroje. Svůj dataset může mít například „Samospráva“, „Školství“, „Majetek“ – ideální je zřídit obdobnou strukturu datasetů jako je tomu v případě odborů. Každý odbor bude tak mít svůj dataset; všichni, kteří v rámci odboru zveřejňují otevřená data pak budou používat pouze jeden svůj konkrétní. Do datasetu se následně přidávají jednotlivé zdroje (soubory). To může být např. konkrétní tabulka („seznam škol“, „seznam nebytových prostor“).

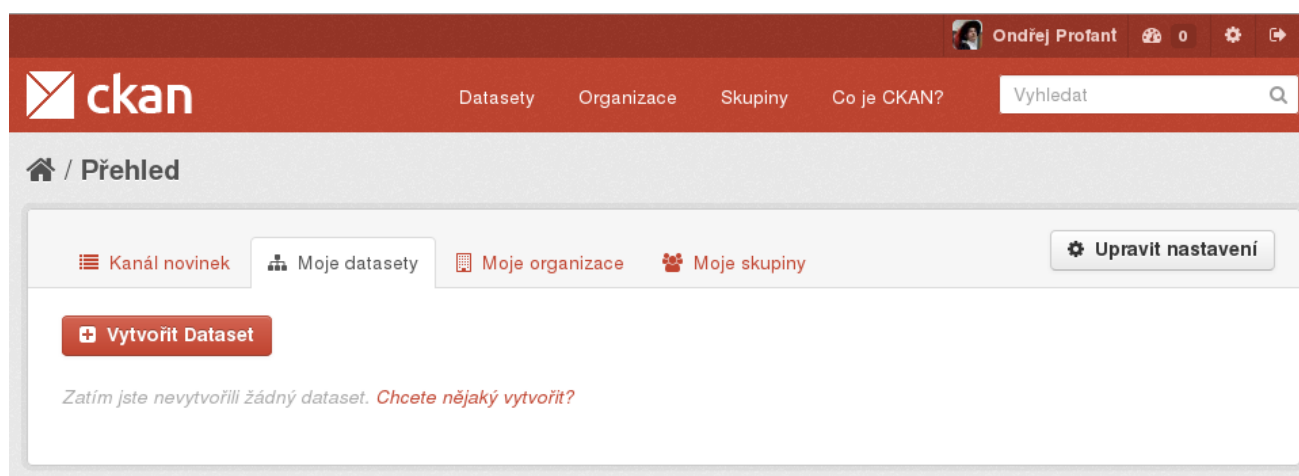
Zdroj

Zdroj je již samotný soubor (jeden!) s daty. Např. soubor `navstevnost.csv` obsahuje tabulku návštěvnosti.

Pokud máte více formátů, tak každý bude zvláštní zdroj (např. IPR zveřejňuje mapové podklady v několika formátech a dvou souřadnicových systémech).



Obrázek 10: Kanál novinek po přihlášení



Obrázek 11: Vytvoření datasetu

Přidání zdroje

V jakém formátu uveřejnit?

Zdroj může být jednoduchá spreadsheetová (excelovská) tabulka. Nicméně v rámci ní musí být dodržet správný formát, nesmí se používat listy etc. Ideální je tabulku uložit ve formátu „comma separated values“ CSV (ten vám nežádoucí věci ani nedoloží uložit). Jedná se o základní a nejlépe čitelný formát pro data uspořádaná do tabulek a proto je velmi důsledně podporován v rámci otevřených dat.

CSV

Jak ta data získám?

Většina informačních systémů umí exportovat. Lze tedy data těžko dostupná z původní aplikace „vyvézt“ tak, aby byla použita nezávisle na programu, ve kterém byla vytvořena. Např. Gordic Ginis (v samostatném návodu) vám dovolí vyexportovat seznam faktur přímo jako CSV). Dané CSV můžete rovnou nahrát jako součást datasetu na portál otevřených dat. Samozřejmě nejdříve se musíte ujistit, že jste exportovali jen žádaná data (např. neobsahují osobní údaje). Kromě toho lze samozřejmě zveřejňovat i jiné typy souborů, nicméně .csv a .txt jsou v tomto případě preferované typy.

Návod: automatizace

Opendata jsou řešena pomocí opensource CMS CKAN. Tento portál má přístupné webové (klikací) rozhraní pro nahrávání souborů. (Viz předchozí kapitola) Mnohdy je však zbytečné práci dělat ručně, když jí můžeme automatizovat.

1. Získat data

Je potřeba získat samotná data. Ideálně ve formátu `csv`. Toho lze docílit např. jednoduchým dotazem na databázi či ručním vytvořením.

2. Nahrát data

CKAN má vlastní programovatelné rozhraní (API) pomocí [REST](#). Čili není složité dopsat si vlastní můstek (plnící skript).

Pro usnadnění jsme připravili ukázkové skripty pro upload v rozšířených skriptovacích jazycích PHP a Python. Skripty naleznete v repozitáři:

<https://github.com/singularita-corp/ckan-examples>

Skripty jsou opravdu jednoduché a slouží jako inspirace.

Neváhejte se ozvat pokud vám bude něco nejasné. Popř. podělit se o ukázky dalších skriptů. Budeme průběžně uveřejňovat další.

Rozvoj

Jelikož není pevně stanoveno, co kdo bude zveřejňovat, tak čekáme na první tvrdá data. Následně vytipujeme nejpoužívanější informační systémy (předpokládáme např. SW Ginis) a pro ty přímo připravíme skripty pro automatizaci.

Návod: získání dat ze softwaru Ginis

Úvod

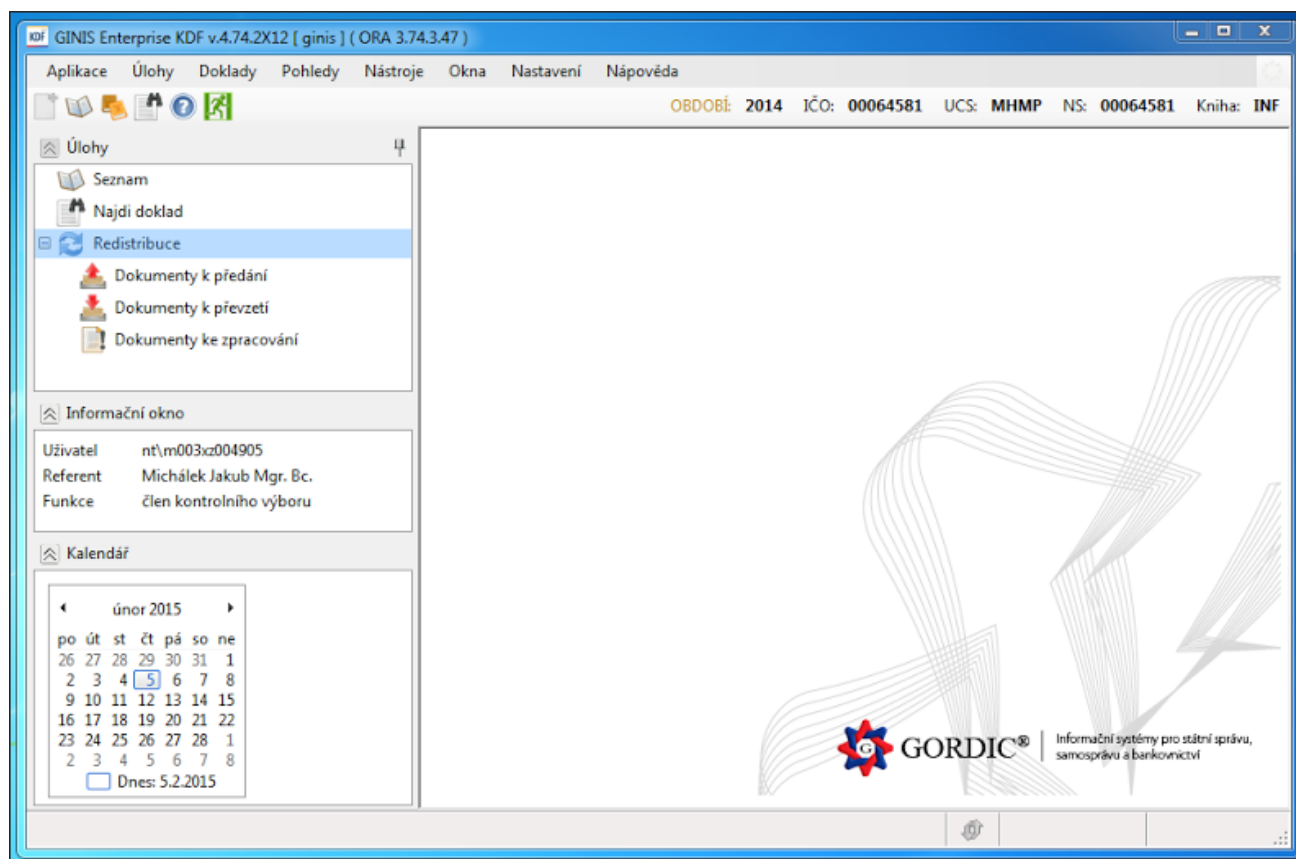
V dnešní době stále častěji pracujeme a chceme pracovat s tzv. otevřenými daty. V případě hlavního města Prahy, jeho městských částí a příspěvkových organizací se však střetáváme s problémem: existuje sice vůle data uvolnit, nicméně software nám v tomto brání. Proč tomu tak je, a jak tento problém lze vyřešit? Tento návod se bude věnovat software Gordic Ginis, který je používám k evidenci celé řady operací (hlavně účetního charakteru). Používá jej Praha, její městské části i příspěvkové organizace. Pracuje s vlastními formáty souborů, které dokáže otevřít pouze program Gordic Ginis a nikdo jiný.

Mluvíme tedy o stavu, známém jako vendor lock-in (proprietární uzavření). Software je totiž navržen tak, aby nám s daty neumožnil pracovat nezávisle na něm. Tím se zvyšuje závislost nás jako pracovníka s tímto software na jeho dodavateli.

Struktura software Ginis

Gordic Ginis není jedním programem, jako je tomu v případě kancelářských balíků, grafických editorů a podobně. Naopak se sestává z celé řady modulů. Agenda úřadu tak může být efektivně rozdělena do jednotlivých modulů. Různí pracovníci úřadu mají přístup pouze do vybraných modulů podle svého pracovního zařazení, čímž je řešena problematika řízení přístupu.

Jednotlivých modulů existuje až několik desítek a jsou označovány třípísmennými zkratkami. Nás bude zajímat modul s názvem Kniha došlých faktur (se zkratkou KDF).



Obrázek 12: Ilustrace 1: Úvodní obrazovka po spuštění Kniha došlých faktur

V tomto modulu jsou jednotlivá data strukturována následujícím způsobem:

Modul obsahuje určité množství účetních knih, které obsahují účetní data týkající se jednotlivých faktur.

- **Modul** je daná část software, se kterou pracujeme (v tomto návodu je to jediné Kniha došlých faktur, tedy KDF)
- **Účetní kniha** je určité seskupení účetních dat, které má stejnou charakteristiku (může to být třeba odbor úřadu nebo daný účetní rok)
- **Účetní data** jsou jednotlivá data v systému Gordic Ginis; v tomto případě odpovídá jedna položka účetních dat jedné faktuře.

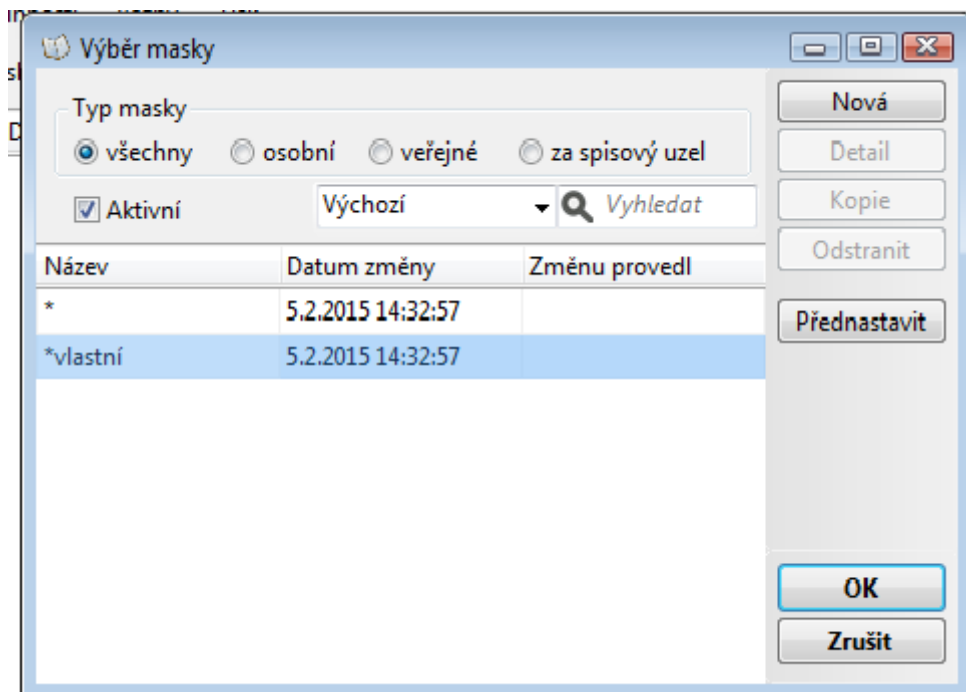
Výběr knihy

První co nás nejspíše bude v tomto případě zajímat, bude přístup k jednotlivým datům. Pomocí tlačítka s dvěma knihami v levé horní části obrazovky se dostaneme do menu „výběr knihy“, kde si vybereme příslušnou účetní knihu.

Vlastní filtry – flagy a masky

Různá účetní data mají nastavena svůj vlastní flag (příznak) viditelnosti. To znamená, že ne každý uživatel, který má přístup k modulu Kniha došlých faktur uvidí všechna data. Někteří uživatelé vidí více účetních dat (a tedy i účetních knih), než ti ostatní.

Kromě tohoto předdefinovaného příznaku viditelnosti účetních dat, potažmo účetních knih, je tu i další, a to uživatelský. Systém Gordic Ginis jej označuje pod názvem maska. Lze jej užívat jako filtr; zobrazit jen ta účetní data (tj. faktury), které jsou jen z určitého roku, mají určitý řetězec textu v názvu, nebo jsou pro nás jiným způsobem zajímavá.



Obrázek 13: Ilustrace 2: Vidíme dva uživatelské filtry. První je předdefinovaný, který nezobrazí nic, a druhý jsme si nastavili sami. Ten nám toho ukáže více

Důležitým faktem také je, že po spuštění systému se automaticky zapíná maska, která nezobrazí primárně žádná data. Je nezbytné jí proto přenastavit tak, že po kliknutí na ikonku knihy vlevo (dvakrát) nastavíme svůj

filtr. Nastavení této masky nám zobrazí veškeré účetní položky, ke kterým máme přístup.

Chceme-li účinně získat veškerá data ze systému Gordic Ginis, musíme vybrat takový uživatelský filtr/masku, která nám bez výjimky **zobrazí veškerá účetní data, ke kterým máme přístup.**

Jak těžít data

Protože jsme došli až sem, znamená to, že už nyní dokážeme bezpečně zobrazit všechny účetní data, která eviduje systém Gordic Ginis a ke kterým máme jako uživatelé přístup. Sám software Gordic Ginis sice není navržen tak, aby poskytoval otevřená data, lze z něj však jednotlivé účetní knihy vyexportovat.

Výše jsme si řekli, že software pracuje s tzv. Účetními daty. Ta seskupuje do jednotlivých knih. Pokud si knihu zobrazíme, vidíme ji jako tabulku a každá položka účetních dat v ní představuje jeden řádek.

Název	Zkratka	Období
01 - Odbor per. a hosp. činn. - PER	01 OPH - PER	2014
01 - Odbor služeb	01 SLU	2014
02 - Odbor Evropských fondů	02-FON	2014
03 - Odbor Archiv hl. m. Prahy	03-AMP	2014
04 - Odbor zdravotnictví, soc. péče a prev.	04-ZSP	2014
05 - Odbor správ činností ve zdrav. a soc. péči	05-SCZ	2014
08 - Odbor bezpečnosti a krizového řízení	08-BKR	2014
14 - Odbor účetnictví	14-UCT	2014
16 - Odbor rozpočtu	16-ROZ	2014
20 - Odbor technické vybavenosti	OTV	2014
21 - Odbor městského investora	OMI	2014
22 - Odbor strategických investic	OSI	2014
23 - Odbor technické vybavenosti	OTV	2014
24 - Odbor personální	24 - PER	2014
28 - Odbor daní, cen a poplatků	DPC	2014
29 - Odbor rozvoje a financování dopravy	29-RFD	2014
29 - účet 149024 - Odbor rozvoje a finan. dopravy	RFD-VH	2014
34 - Odbor stavební a územního plánu	SUP	2014
40 - Odbor informatiky	INF	2014
43 - sek. radního pro oblast informatiky	SE6-Ing. Vašíček	2014
44 - Sek. nám. primátora pro obl. zdrav., soc. a byt.	SE 4 (Dlouhý)	2014
45 - Sek. 1. nám. primátora pro obl. fin. pol.	SE 1 (Vávra)	2014
47 - sek. radního pro obl. kult. pam. p. a ces. r., ZV	SE 2 (Novotný)	2014
48 - Sekret. radní pro obl. maj. EF a podp. podnikání	SE- 8 Vorlíčková	2014

Obrázek 14: Ilustrace 3: Jednotlivé účetní knihy, které používá Magistrát hlavního města Prahy

Nejjednodušší metoda těžby dat z Gordic Ginis spočívá v zobrazení každé účetní knihy zvlášť a v jejím exportu do tabulky ve formátu .xlsx, .ods nebo i .csv. Toto lze provést následujícími způsoby:

1. Automatizovaným skrze API
2. Ručním
 1. Po částech (při větších objemech dat přehlednější)

2. V jednom celku

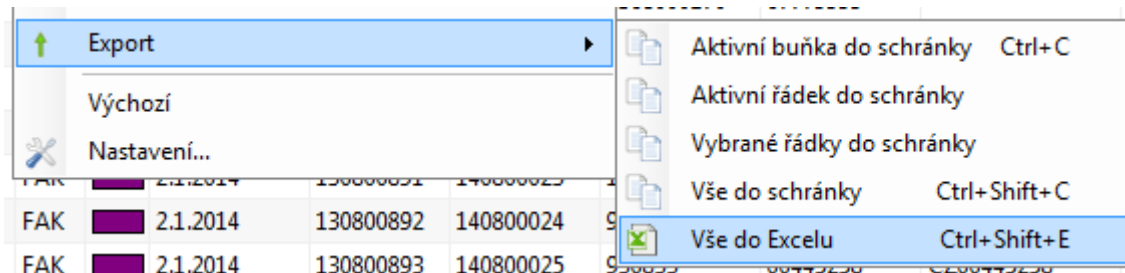
Automatizovaný export může zajistit vaše IT oddělení skrze zdokumentované API. Výsledek je stejný jako, když export provádíte ručně. Pro začátek je ruční export zcela dostačující. V dalších měsících město dodá exportní skript.

Magistrát hlavního města Prahy jako uživatel systému Gordic Ginis má ve své databázi v Knize došlých faktur cca 230 000 účetních dat (tj. faktur). To je opravdu hodně. Software Gordic Ginis umožňuje zobrazit si je všechny najednou...

Pomalý způsob spočívá v exportu každé účetní knihy zvlášť do separátního souboru. Rychlý způsob spočívá v exportu všech účetních dat z každé účetní knihy a každého roku do jednoho souboru. Rychlý způsob vyžaduje jen jediný export, zároveň ale potřebuje tolik paměti při samotné konverzi dat, že vede zcela jistě k havárii systému – ale jen v případech, kdy pracujeme s velkými objemy dat.

Chceme-li bezpečně exportovat data ze systému Gordic Ginis výše uvedeným způsobem po částech, uskutečníme toto prostřednictvím následujícího pracovního postupu:

1. Vybereme konkrétní knihu, kterou chceme exportovat
2. označíme veškeré položky (CTRL+A)
3. pravým tlačítkem vybereme z kontextové nabídky „export → vše do Excelu“
4. otevře se nám automaticky spreadsheetový editor s účetními daty z dané knihy.
5. uložíme danou tabulku
6. postup opětujeme s další knihou.



Obrázek 15: Ilustrace 4: Kontextová nabídka pro export příslušného výběru z účetní knihy

Formát CSV - Comma separated values

Jakmile máme data ze systému Gordic Ginis v tabulkovém editoru, máme v podstatě vyhráno – lze s nimi již totiž pracovat tak, jako s čímkoliv jiným. Nyní si vysvětlíme, proč je dobré je ukládat ve formátu „comma separated values“.

Podobně jako pro čistý text je zdrojový kód samotným základem, i v případě tabulkových editorů existuje jeden, základní formát – zdrojový kód. Je to tzv. „comma separated values“, tedy hodnoty oddělené čárkami. V tomto formátu oddělují jednotlivé buňky čárky a jednotlivé řádky jsou uspořádány tak, jak by se zobrazily v tabulkovém editoru. Z tabulkového editoru lze exportovat data do formátu „comma separated values“ (který má zpravidla příponu .csv). Lze si také vybrat, zda-li budou hodnoty odděleny středníkem, nebo formátovány pro různé operační systémy.

Výhodou tohoto formátu je skutečnost, že jej lze otevřít v internetovém prohlížeči, nejobyčejnějším textovém editoru, ba jej dokonce zobrazit bez jakýchkoliv problémů i v textovém (tj. negrafickém) rozhraní operačního systému. Jedná se proto o dobrý formát pro provedení kontroly a především pro export dat takovým způsobem, aby byla dostupná opravdu všem.