

# Jaká data veřejné správy ČR jsou otevřená?

Dušan Chlapek

Fakulta informatiky a statistiky,  
Vysoká škola ekonomická v Praze

## Cíle projektu a případové studie

- ověřit možnosti přípravy aplikace sémantických technologií při publikování dat veřejné správy ČR
- výběr a přizpůsobení nástroje pro katalogizaci dat VS
- nastavení metodiky pro použití nástroje pro katalogizaci
- prvotní katalogizace dat VS ČR
- pilotní aplikace a experimenty nad katalogizovanými a zpracovanými daty
- Případová studie řešena v rámci první etapy projektu "Sémantické propojování dat ve veřejné správě" financovaného interní grantovou agenturou VŠE Praha (IG407011), hlavními řešiteli jsou (kromě autora příspěvku) Jan Kučera, Jindřich Mynarz, Marek Ovečka, Martin Tajtl, Vojtěch Svátek
- doba trvání celého projektu 9 měsíců od března 2011

## Otevřená data veřejné správy

- jsou data, která zveřejňuje s použitím informačních technologií orgán veřejné správy ve formátu umožňujícím jejich další strojové zpracování.
- Šíření těchto dat není omezeno licenční ani jinou právní podmínkou.
- Otevřená data umožní vytváření propojení a vytváření aplikací nezávisle na poskytovateli dat.

SOFTECON2012 

## Aktivity v oblasti otevřených dat

- rostoucí objem dat opatřených sémantickými metadaty odkazujícími na sdílené slovníky (ontologie) ve světě
- v některých zemích již data systematicky zveřejňována, např.:
  - USA (2009, iniciativa „Open Government“ prezidenta Obamy)
  - Velká Británie
  - Rakousko
  - Nizozemí
  - Francie
  - ...
- EU - Evropský portál otevřených dat - plánovaný termín spuštění v roce 2013
- LOD2.EU - Linked Open Data - projekt spolufinancovaný Evropskou komisí

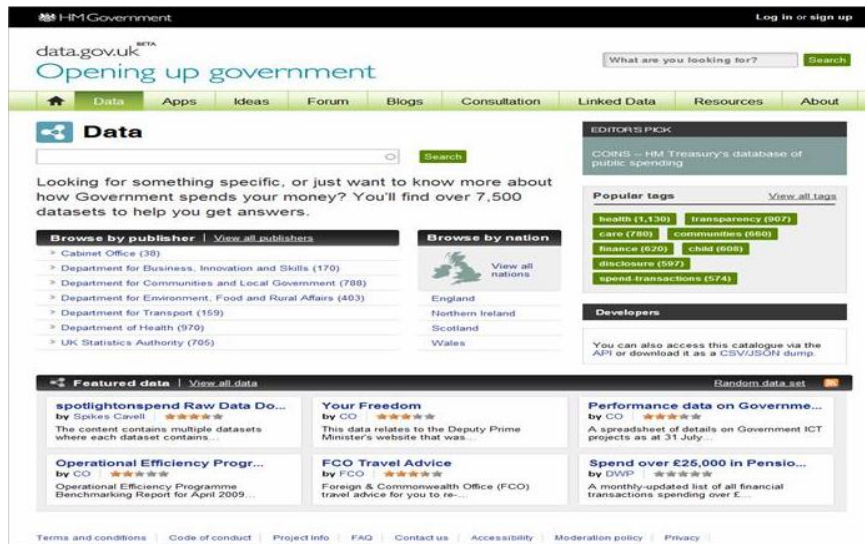
SOFTECON2012 

# Aktivita v oblasti otevřených dat USA



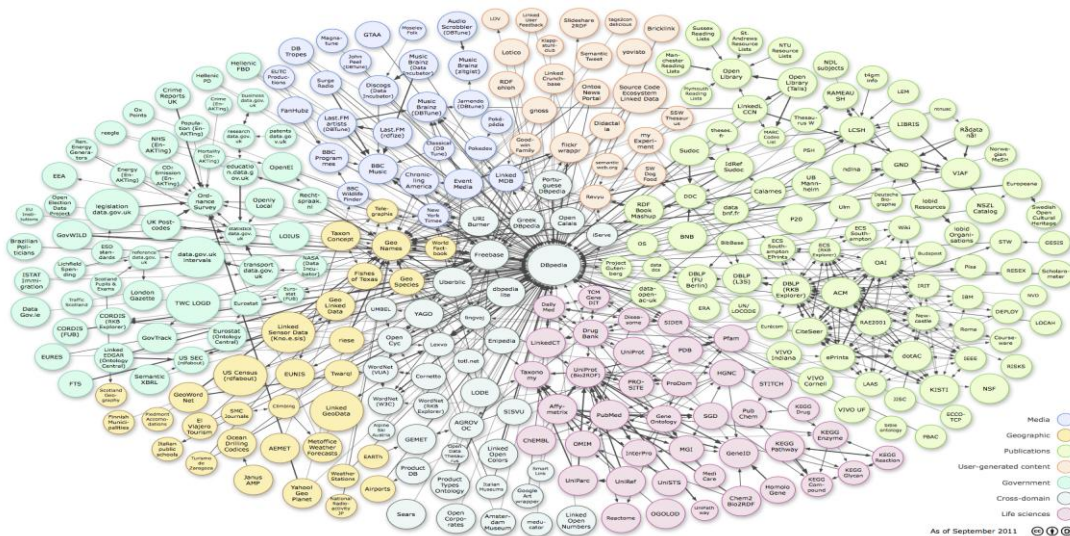
SOFTECON2012

# Aktivita v oblasti otevřených dat VB



SOFTECON2012

# The Linking Open Data cloud diagram



SOFTECON2012

## Příklady aplikací v ČR

- Budování státu - rozklikávací rozpočet, kam jdou moje daně - <http://budovanistatu.cz/>
- Deník politika - profily politiků, politických stran, obecních i krajských zastupitelstev, institucí a voleb v České republice. <http://www.denikpolitika.cz/>
- Inventura demokracie - zadávání veřejných zakázek, zamezení praxe legislativních přílepků, depolitizace kontrolních institucí a omezení poslanecké a senátorské imunity. <http://www.inventurademokracie.cz/>
- zIndex - hodnocení zadavatelů veřejných zakázek <http://www.zindex.cz/>
- Průhledná radnice - zprůhlednit činnost zastupitelstev a navýšit možnost kontroly veřejných výdajů, zakázek a závazků ze smluv. <http://www.pruhlednaradnice.cz/>

SOFTECON2012

## Příklady aplikací v ČR

- OpenData.cz <http://www.opendata.cz>
  - neformální iniciativa pro podporu budování transparentní datové infrastruktury české veřejné správy.
  - podílí se na ní lidé z různých organizací, například Matematicko-fyzikální fakulty Univerzity Karlovy, Vysoké školy ekonomické v Praze
  - výsledky prezentované případové studie jsou zveřejňovány i přes tento portál

SOFTECON2012 

## Licencování dat

- Otevřená data ve smyslu Open Knowledge  
Definition:
  - data dostupná jako celek v podobě umožňující úpravy
  - povoleno provádění úprav dat spolu s šířením vzniklých modifikací
  - využití a další šíření dat není omezeno
  - Detailněji viz <http://opendefinition.org/okd/>

SOFTECON2012 

## Licencování dat - dopady

- Nestačí data pouze technicky zpřístupnit, ale je třeba udělit patřičná oprávnění jejich uživatelům
- Různá právní ochrana v jednotlivých zemích a nejednotný výklad
- V některých anglosaských zemích lze dílo věnovat do Public Domain, kde již není chráněno autorským zákonem.
- V ČR se autor nemůže vzdát osobnostních a majetkových práv.
- Dle zákona č. 121/2000 Sb. je chráněna struktura databáze jako autorské dílo souborné a jsou přiznána *"zvláštní práva k databázi pořizovateli, pokud pořízení, ověření nebo předvedení obsahu databáze představuje kvalitativně nebo kvantitativně podstatný vklad bez ohledu na to, zda databáze nebo její obsah jsou předmětem autorskoprávní nebo jiné ochrany"*

SOFTECON2012 

## Licencování dat - nové licence

- Za účelem publikace dat otevřených v souladu s Open Knowledge Definition vznikají nové licence, např.:
  - Open Data Commons Public Domain Dedication and Licence
  - Open Data Commons Open Database License
  - Open Government Licence
- organizace VS ČR zatím neuvádějí licence ani podmínky za jakých lze se zveřejněnými daty pracovat

SOFTECON2012 

## Mapování dat v ČR - zvolený nástroj

- Comprehensive Knowledge Archive Network (CKAN)
  - sw pro vytváření a správu katalogu dat
  - vyvíjen jako svobodný/open source software
  - Open Knowledge Foundation (OKFN) provozuje instanci CKAN pro ČR (<http://cz.ckan.net>)
  - Hlavní důvody výběru: dostupnost, upravitelnost, možnost využití přispěvateli i mimo grantový projekt

SOFTECON2012 

## Metodika pro využití CKAN

- balíček dat - základní jednotka v CKAN
- do balíčků dat jsou sdružovány zdroje, které spolu souvisí a platí pro ně atributy balíčku dat:
  - **Title** - název balíčku
  - **Name** - jednoznačný identifikátor balíčku
  - **URL** - odkaz na adresu webové stránky, na které je k nalezení popis dat (nejedná se o odkaz na data samotná)
  - **Notes** - popis balíčku
  - **Licence** - licence nebo právní podmínky, které určují, jak lze nakládat s daty v balíčku (nejedná se o licenci k balíčku, ale pouze o licenci vztahující se ke stažitelným souborům uvedeným jako Zdroje (Resources) v rámci balíčku)
  - **Tags** - štítky, věcné/předmětné kategorie, pomocí kterých označujeme balíček
  - **Resources** - zdroje, odkazy na stažitelné soubory s daty nebo na API aplikace
  - **Groups** - skupiny, označení skupin, do kterých balíček náleží
  - **Author** - autor nebo původce dat v balíčku
  - **Author email** - emailová adresa autora nebo původce dat
  - **Maintainer** - správce dat v balíčku
  - **Maintainer email** - emailová adresa správce dat v balíčku
  - **Version** - verze dat v balíčku

SOFTECON2012 



# Metodika pro využití CKAN

- Tags - štítky
  - označení balíčku dat věcnými kategoriemi
  - štítky v EN voleny z Integrated Public Sector Vocabulary <http://doc.esd.org.uk/IPSV/2.00.html>
  - štítky v CZ voleny z CZ-NACE <http://apl.czso.cz/iSMS/klasstru.jsp?kodcis=80004>
  - příp. volně vytvořené

SOFTECON2012

# Ukázka úvodní obrazovky v CKAN

The screenshot shows the CKAN Czech Republic homepage. At the top, there is a navigation bar with links for 'Přidat dataset', 'Vyhledat', 'Skupiny', 'O CKAN', 'Návod', and 'Metodika'. Below this, the main heading reads 'Vítejte na CKAN Czech Republic!'. There are three main action buttons: 'Vyhledat data' (Search for data), 'Sdílení dat' (Share data), and 'Spolupracovat' (Collaborate). The 'Vyhledat data' button includes a search input field and a button to 'Vytvořit datasetu' (Create dataset). The 'Sdílení dat' button includes a text area for sharing and a 'Vytvořit datasetu' button. The 'Spolupracovat' button includes a list of links: 'GetTheData.org', 'DataPatterns.org', and 'Open Data Manual'. Below these buttons, there is a 'Tagy' (Tags) section with a word cloud of various tags in English and Czech, such as 'ministry\_of\_labour\_and\_social\_affairs', 'práce', 'rozpočet', and 'škola'. At the bottom, there is a footer with 'O CKAN', 'Sekce', 'Jazyky', and 'Meta' information.

SOFTECON2012



# Ukázka balíčku dat v CKAN

**Prague tourism statistics – Statistiky cestovního ruchu v Praze**

**English description**  
This dataset contains statistical data about tourism in the City of Prague. Data is available from the Portal of Prague.

**Český popis**  
Tento dataset obsahuje statistická data o oblasti cestovního ruchu v Praze. Data jsou zveřejněna na Portálu hl. m. Prahy.

**Zdroje**

Zdroj	Formát	Akce
Statistics of the overnight stays in accommodation facilities in the City of Prague	XLS	Preview
Statistics of number of guests in accommodation facilities in the City of Prague	XLS	Preview
Conference in accommodation facilities in the City of Prague. Conference	XLS	Preview
Usage of beds in accommodation facilities. Využití lůžek v hromadných ubyt.	XLS	Preview
Comparison of number of guests in accommodation facilities in the City of Prague	XLS	Preview
Comparison of number of overnight stays in accommodation facilities in the City of Prague	XLS	Preview
Number of guests in accommodation facilities in the City of Prague by country	XLS	Preview
Statistics of number of overnight stays in accommodation facilities in the City of Prague	XLS	Preview
Number of guests by country 2009 – 2010. Počet hostů dle zemí 2009 – 2010	XLS	Preview
Number of overnight stays by country 2009 – 2010. Počet přenocování dle zemí	XLS	Preview

**Další informace**

Pole	Hodnota
Language	Czech

**O CKAN**  

- O CKAN
- Twitter @ckanproject
- API
- Dokumentace API
- Kontaktujte nás
- Pravidla pro odvrtnou soukromí

**Sekce**  

- Uživatelé
- Tagy
- Verze
- Autorizační skupiny
- Správce webu

**Jazyky**  

- čeština (Česká republika)
- English
- Deutsch
- français
- italiano
- español
- português
- русский

**Meta**  
 © 2011 Open Knowledge Foundation  
 Licencován v rámci Open Database License  
 Běžná na CKAN v1.5.1.

SOFTECON2012

# Ukázka balíčku dat v CKAN

**State employment policy spendings - Výdaje na státní politiku zaměstnanosti**

**English description**  
Information about spendings on state employment policy. This information is available from the Integrated portal of the Ministry of Labour and Social Affairs of the Czech Republic.

**Český popis**  
Informace o výdajích na státní politiku zaměstnanosti, které jsou k dispozici v rámci integrovaného portálu Ministerstva práce a sociálních věcí.  
Domovská stránka Ministerstva práce a sociálních věcí České republiky.

**Dataset structure (Struktura datasetu)**  
The following picture represents an UML class diagram describing source called State employment policy general overview (1991 - 2009).  
Na následujícím obrázku je uveden UML diagram tříd vytvořený pro zdroj Výdaje na SPZ souhrnný přehled 1991-2009.

```

classDiagram
    class Vydaj_na_statni_politiku_zaměstnanosti {
        -hodnota: float
        -rok: int
    }
    class Vydaj_na_pasivni_politiku_zaměstnanosti {
        -hodnota: float
        -rok: int
    }
    class Vydaj_na_aktivni_politiku_zaměstnanosti {
        -hodnota: float
        -rok: int
    }
    class Dicer_vydaj_na_aktivni_politiku_zaměstnanosti {
        -hodnota: float
        -rok: int
    }
    class Typ_vydaje_na_APZ {
        -typ: string
    }
    Vydaj_na_statni_politiku_zaměstnanosti "1" -- "1" Vydaj_na_pasivni_politiku_zaměstnanosti
    Vydaj_na_statni_politiku_zaměstnanosti "1" -- "1" Vydaj_na_aktivni_politiku_zaměstnanosti
    Vydaj_na_aktivni_politiku_zaměstnanosti "1" -- "1" Dicer_vydaj_na_aktivni_politiku_zaměstnanosti
    Dicer_vydaj_na_aktivni_politiku_zaměstnanosti "1" -- "0..*" Typ_vydaje_na_APZ
    
```

**Tento dataset není Otevřený**  
Bud proto, že licence není otevřená nebo že chybí stažitelné zdroje.  
Spusťte dotaz na [IsOpenData](#).

SOFTECON2012

## Výsledky prvotního mapování

- mapována data zveřejněná na www stránkách a portálech úřadů VS v ČR pro všechna ministerstva, ČSÚ, krajské úřady a 4 okresní města Libereckého kraje.
- mapovány byly pouze zdroje, které jsou ve formátech dobře strojově zpracovatelných (xls, xlsx, xml, csv)
- V průběhu projektu katalogizováno 1470 souborů, které jsou zařazeny do 145 balíčků.
- K 22.2.2012 obsahuje katalog 165 balíčků, které obsahují 1604 souborů.

SOFTECON2012 

## Příklady využití otevřených dat - I

- Analýzy dat o grantech a dotacích
- Sledování vývoje ceny práce dle regionů a typů zaměstnání
- Analýzy dat z registru vozidel
- Interaktivní mapa škol a školských zařízení
- Mapa ekologických zemědělců a vývozců/dovozců biopotravin
- Vývoj cenových indexů na základě dat ČSÚ
- Párování vypsanych veřejných zakázek a možných dodavatelů

SOFTECON2012 

## Příklady využití otevřených dat - II

- Rozklikávací rozpočty a sledování veřejných výdajů
- Ceníky produktů a služeb, cenové mapy
- Vizualizace data na mapovém podkladě – počasí, dopravní informace, významné objekty
- Tematické portály – např. pro oblast čisté energie
- Velká Británie publikovala jako otevřená propojená data sbírku zákonů

SOFTECON2012 

## Další realizované výstupy projektu VŠE Praha

- spravována a udržována česká verze instance CKAN
- instance CKAN zpřístupněna pro další rozšiřování
- připraven návod na použití CKAN
- analýza stran zainteresovaných na otevřených a propojitelných datech VS ČR
- grafický popis obsahu balíčků dat
- využívání dat katalogizovaných v CKAN pro další zpracování (převod do RDF) a propojování v navazujících případových studiích

SOFTECON2012 