

Inspirujme se
Bratislava, 27. 11. 2013

Open Data v ČR

Martin Nečaský¹,
Dušan Chlapek²,
Jan Kučera²

¹Matematicko-fyzikální fakulta,
Univerzita Karlova v Praze

²Fakulta informatiky a statistiky,
Vysoká škola ekonomická v Praze

Obsah

1. Co jsou otevřená data
2. Aktivity v oblasti otevřených dat v ČR
3. Přínosy a výzvy využívání otevřených dat v ČR

Co jsou otevřená data?

- **Data zveřejněná na Internetu, která jsou**
 - **úplná,**
 - **snadno dostupná,**
 - **strojově čitelná,**
 - **používající standardy s volně dostupnou specifikací,**
 - **zpřístupněna za jasně definovaných podmínek užití dat s minimem omezení,**
 - **dostupná uživatelům při vynaložení minima možných nákladů.**
- **Požadavky na podmínky užití otevřených dat:**
 - **Neomezují jejich uživatele ve způsobu použití dat.**
 - **Opravňují uživatele k jejich dalšímu šíření.**
 - **Musí být uveden autor dat (i při dalším šíření).**
 - **Při dalším šíření musí i ostatní uživatelé mít stejná oprávnění s daty nakládat - během šíření dat nesmí dojít např. k omezení jejich využití pouze pro nekomerční účely.**

Ukázka datového katalogu

The screenshot displays the DATA.GOV.UK website interface. At the top, it features the HM Government logo and a 'Log in or sign up' link. The main header includes the 'DATA.GOV.UK (beta)' logo with the tagline 'Opening up Government'. A navigation bar contains links for Home, Data, Participate, Data requests, Apps, Location, Linked Data, Library, and About. Below this is a secondary navigation bar with links for Search, Map Search, Publishers, Tags, Public Roles & Salaries, Spend Browser, Spend Reports, and Site Usage.

The main content area is divided into several sections:

- Search Datasets:** Shows 14199 results. Includes a search input field and a 'Search' button.
- Tags:** Lists various tags such as 'communities disclosure', 'health-well-being-and-care', 'Environment', 'children finance', 'Spending Data', 'care', 'school NERC_DDC', 'health', 'spend-transactions', 'local-government', 'population', 'health-and-social-care', 'transparency', 'nhs', 'Health child', and 'education'. A 'View all tags' button is present.
- Publishers:** Lists various publishers with their respective dataset counts, such as 'Office for National Statistics (1003)', 'Department for Communities and Local Government (813)', 'NHS Information Centre for Health and Social Care (634)', 'Environment Agency (503)', 'Department for Environment, Food and Rural Affairs (465)', 'British Geological Survey (377)', 'Her Majesty's Revenue and Customs (362)', 'Department for Transport (343)', 'Northern Ireland Spatial Data Infrastructure (317)', and 'Ministry of Defence (311)'. A 'View all publishers' button is included.
- UK Location:** Promotes the 'UK Location Programme' which has introduced over 1000 location data records into data.gov.uk. It includes a 'Conduct map based search' button and a 'uk location' logo.
- Data Requests:** Provides information on making data requests and their progress, with a 'Request new data' button.
- Featured dataset:** Highlights the 'National Public Transport Access Nodes (NaPTAN)' dataset, published by 'Transport Direct' and updated on '28 Sep 2013'. It includes a brief description and feedback icons.
- Developers:** Offers information on accessing the catalogue via API or downloading it as a CSV or JSON dump.

Obsah

1. Co jsou otevřená data
2. Aktivity v oblasti otevřených dat v ČR
3. Přínosy a výzvy využívání otevřených dat v ČR

Aktivity v oblasti otevřených dat v ČR

OpenData.cz

OPENDATA CZ

- Akademická iniciativa MFF UK a FIS VŠE (od r. 2011)
- Hlavní aktivity:
 - Propagace myšlenky otevřených dat na úrovni státní správy i místních samospráv
 - Práce s konkrétními datovými množinami veřejných institucí
 - extrakce ze stávajících formátů (často HTML či XLS)
 - konverze do otevřené a propojitelné podoby (Linked Data, RDF)
 - <http://linked.opendata.cz/>

Partnerství pro otevřené vládnutí

- v roce 2012 vytvořen Akční plán, který stanovil
 - přijetí zákona o úřednících veřejné správy zajišťující odpolitizování, profesionalizaci
 - a stabilizaci veřejné správy
 - zefektivnění systému svobodného přístupu k informacím
 - zpřístupnění dat a informací
 - zajistit právní a technickou otevřenost dat veřejné správy
 - vytvořit infrastrukturu otevřených dat ČR a jejich pravidel v oblasti veřejných zakázek
 - publikovat vybranou kolekci 10 datových množin v otevřené podobě (množiny v gesci různých orgánů)
 - zprovoznit a udržovat katalog otevřených dat, který umožní snadné vyhledávání datových množin různých orgánů VS ČR publikovaných v otevřené podobě

Zdroj: <http://www.korupce.cz/assets/partnerstvi-pro-otevrene-vladnuti/Akcni-plan-OGP.pdf>

Partnerství pro otevřené vládnutí

- reálné výsledky Akčního plánu realizované do dnešního dne
 - [Koncepce katalogizace otevřených dat VS ČR](#)
 - [Metodika publikace otevřených dat veřejné správy ČR](#)
 - Koncepce a Metodika jsou dostupné na portálu Korupce.cz
 - definice datové množiny a kritérií otevřenosti
 - stanovení postupů katalogizace, odhady pracností a nákladů
 - definice procesů publikace datových množin v otevřené podobě
 - Z 10 datových množin byla publikována v otevřené podobě zatím pouze 1 (Český statistický úřad začal v otevřené podobě zveřejňovat datové množiny s výsledky voleb)
 - Katalog nebyl do dnešního dne zprovozněn

Zdroj: <http://www.korupce.cz/assets/partnerstvi-pro-otevrene-vladnuti/Akni-plan-OGP.pdf>

Fórum pro otevřená data

SPOLEČNĚ OTEVÍRÁME DATA

soutěž pro aplikace a projekty nad otevřenými daty

- Spolupráce **Fondu Otakara Motejla, FIS VŠE a MFF UK** za účelem koordinace propagace myšlenky otevřených dat a poskytování služeb státní správě, místním samosprávám i soukromým subjektům v této oblasti
- **Aktuální aktivity**
 - propagace na seminářích, workshopech a konferencích,
 - soutěž o nejlepší aplikaci nad otevřenými daty v ČR,
 - koordinace měst a obcí při publikaci jejich vybraných datových množin v otevřené podobě

Zdroj: <http://www.otevrenadata.cz/>

Instituce publikující otevřená data

- Mimo Akční plán partnerství pro otevřené vládnutí existuje v ČR řada institucí, které začaly během roku 2013 publikovat svá data jako otevřená
 - **Česká obchodní inspekce (ČOI)** na svých webových stránkách publikuje data o provedených kontrolách, jejich výsledcích a udělených sankcích
 - <http://www.coi.cz/cz/spotrebitel/open-data-databaze-kontrol-sankci-a-zakazu/>
 - **Český telekomunikační úřad (ČTÚ)** zpracoval komplexní analýzu svých datových množin a rozhodl, že řadu z nich bude publikovat v otevřené podobě během let 2013-2015
 - <http://www.ctu.cz/aktuality/tiskove-zpravy.html?action=detail&ArticleId=10859>
 - **Ministerstvo financí** provozuje aplikaci Monitor státní pokladny, která mimo jiné zpřístupňuje svá data ve strojově čitelné podobě
 - <http://monitor.statnipokladna.cz/2013/data/>
 - <http://monitor.statnipokladna.cz/2013/webove-sluzby/>

Výzkumný projekt COMSODE



- zaměřuje se na vývoj softwarové platformy a metodik pro publikaci otevřených dat v co nejvyšší kvalitě
 - transformace a čištění, propojování, katalogizace a prohledávání datových množin
 - postupy, návody k prioritizaci, odhady pracnosti publikace datových množin
- ověření nástrojů a metodik na konkrétních datových množinách
- spolupráce s orgány veřejné správy
- Slovensko, Česko, Itálie, Nizozemí

Aplikace – www.lekovaencyklopedie.cz

LÉKOVÁ ENCYKLOPEDIÉ

Aspirin

Účinné látky

The prototypical analgesic used in the treatment of mild to moderate pain. It has anti-inflammatory and antipyretic properties and acts as an inhibitor of cyclooxygenase which results in the inhibition of the biosynthesis of prostaglandins. Aspirin also inhibits platelet aggregation and is used in the treatment of thrombosis. (From Martindale, The Extra Pharmacopoeia, 30th ed, p5)

Indikace	Prevence	Kontraindikace
Bolest	Bolest	Astma
Dna (nemoc)	Cerebrální infarkt	Heemoragické poruchy
Horečka	Infarkt myokardu	Koagulopatie
Osteoartrida	Preeklampsie	Kojenec
Revmatická horečka	Tranzitorní ischemická ataka	Léková alergie
Revmatoidní artritida		Mladiství
Záněť		Nesnášlivost
		Rinida
		Trombocytopenie
		Těhotenství - třetí trimestr
		Vrovné nemoci
		Young child

Farmakologické účinky

Mechanismy účinku

Fyziologické účinky

Farmakokinetika

Použití v těhotenství

Kontraindikováno ve třetím trimestru (zdroj: ndf-rt)

LÉKOVÁ ENCYKLOPEDIÉ

ANOPYRIN 100 MG

Čeští léčivého přípravku

Indikace

Nestabilní angina pectoris – jako doplněk standardní terapie. Akutní infarkt myokardu – jako součást standardní terapie. Profylaxe reinfarktu. Po arteriálních cévněchirurgických nebo intervenčních výkonech (např. po aortokoronárním přemostění, při perkutánní transluminální koronární angioplastice). Profylaxe tranzitorních ischemických atak a mozkových infarktů, když se vyskytla prodromální stadia. Upozornění: Přípravek Anopyrin 100 mg není vhodný k terapii bolestivých stavů. Přípravek není určen pro použití u dětí a mladistvých do 18 let.

Indikační skupina

ANTICOAGULANTIA (FIBRINOLYTICA, ANTIFIBRINOL.)

Účinné látky

Aspirin

Farmakologické účinky

Antiflogatika nesteroidní

Antipyretika

Cyclooxygenasa - inhibitory

Fibrinolytika

Inhibitory agregace trombocytů

Použití v těhotenství

Kontraindikováno ve třetím trimestru (zdroj: ndf-rt)

ATC skupiny

Informace o balení

Doplněk názvu	POR TBL NOB 3X20X100MG	Balení	3X20
Síla	100MG	Podání	Perorální podání
Registrační kód SÚKL	16/ 882/94-C	Léková forma	
Jednotková cena	1.22 Kč	Maximální doplatek	73.16 Kč

Ekonomicky výhodnější alternativy

GODASAL 100

POR TBL NOB 100

Maximální doplatek : 13.89 Kč

Jednotková cena : 0.14 Kč

STACYL 100 MG ENTEROSOLVENTNÍ TABLETY

POR TBL ENT 100X100MG II

Maximální doplatek : 13.89 Kč

Jednotková cena : 0.14 Kč

STACYL 100 MG ENTEROSOLVENTNÍ TABLETY

POR TBL ENT 100X100MG

Maximální doplatek : 13.89 Kč

Jednotková cena : 0.14 Kč

STACYL 100 MG ENTEROSOLVENTNÍ TABLETY

POR TBL ENT 100X100MG I

Maximální doplatek : 13.89 Kč

Jednotková cena : 0.14 Kč

Obsah

1. Co jsou otevřená data
2. Aktivity v oblasti otevřených dat v ČR
3. Přínosy a výzvy využívání otevřených dat v ČR

Přínosy otevřených dat

- (P1) posílení transparentnosti,
- (P2) podpora ekonomického růstu,
- (P3) zlepšení služeb veřejné správy,
- (P4) podpora opětovného použití dat,
- (P5) zlepšení vnímání veřejné správy veřejností,
- (P6) zlepšení procesů a dat veřejné správy,
- (P7) zamezení chyb vzniklých při práci s daty
- (P8) snížení počtu dotazů dle zák. č. 106/1999 Sb.
- (P9) pořádek ve vlastních datech
- (P10) možnost transparentně informovat veřejnost o subjektech, které neplní své zákonné povinnosti
- (P11) zvýšení hodnoty dat
- (P12) budování dobrého jména úřadu
- (P13) podpoření kreativity (odborné) veřejnosti

Výzvy pro veřejnou správu ČR

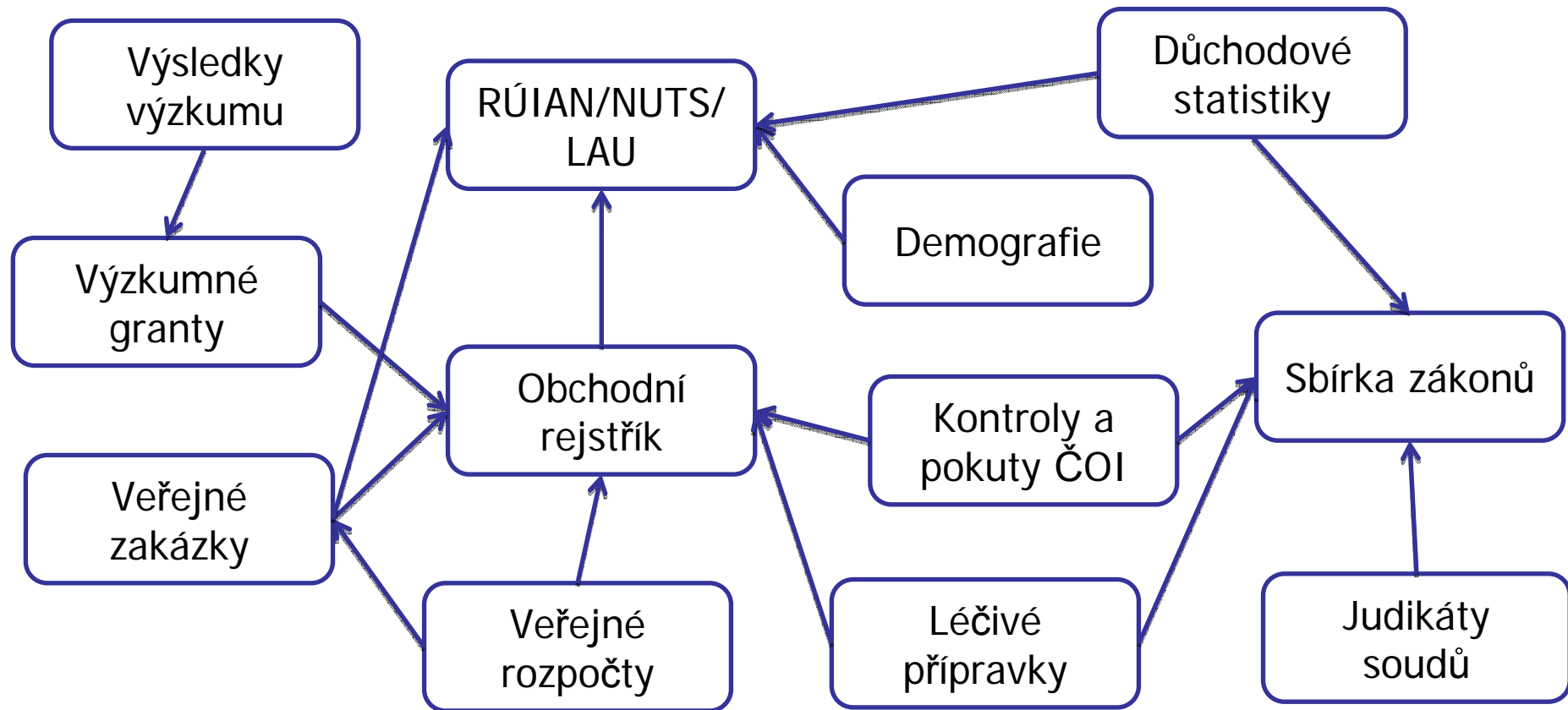
1. Publikace prioritních datových množin
2. Sjednocení způsobu publikace
3. Analýzy potenciálu otevřených dat pro konkrétní orgány VS
4. Katalog otevřených dat ČR
5. Implementace novely směrnice 2003/98/ES
6. Komunikační strategie
7. Propojitelná data (Linked Data)

Propojitelná data (Linked Data)

- Linked Data = 4 principy efektivní publikace dat na webu
 - obdoba publikace dokumentů ve formě webových stránek
 - místo dokumentů však publikujeme datové entity a místo odkazů mezi webovými stránkami publikujeme vztahy mezi entitami
 - základem je datový model RDF (Resource Description Framework)
- Proč jsou Linked Data vhodná pro veřejnou správu?
 - Různé instituce publikují na svých webových stránkách data, která spolu vzájemně úzce souvisejí
 - Linked Data umožňují nejenom publikovat data v otevřené podobě na stránkách jednotlivých institucí, ale umožňují je také propojit
 - propojení jsou opět strojově čitelná

Propojitelná data (Linked Data)

- Experiment na MFF UK v rámci EU projektu COMSODE a několika národních projektů TAČR



Děkuji za pozornost

- **BACKUP**

Otevřená data ve světě

- Otevřená data se prosazují nejen v EU
- Významnou roli hrají USA
 - Portál Data.gov
 - Open Data Policy (květen 2013)
 - Definuje pojmy včetně otevřených dat
 - Používání otevřených a strojově čitelných formátů, používání otevřených licencí atd.
- Brazílie
 - National Open Data Infrastructure
 - Součástí je i datový katalog
 - Stanovuje datové standardy v rámci veřejné správy
 - Podporuje orgány VS, které chtějí publikovat otevřená data
 - Podporuje zapojení veřejnosti
- Open Data Index (Open Knowledge Foundation)
- Open Data Certificate (Open Data Institut)

Zdroj USA: <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

Zdroj Brazílie: <https://www.undpegov.org/sites/undpegov.org/files/Brazil-OD-2013-05-29.pdf>

Datový katalog Evropské unie

The screenshot shows the European Union Open Data Portal interface. At the top, there is a navigation bar with the logo and the text "European Union Open Data Portal BETA". Below this is a breadcrumb trail: "Europa > Open Data Portal > Data". The main navigation menu includes "Home", "Data", "Applications", "Linked Data", and "About". A search bar is present with the placeholder text "Find datasets...". Below the search bar, there are radio buttons for "Show results with:" and options for "all of these words", "any of these words", and "the exact phrase". A message states "Data Portal contains 6285 datasets that you can browse, learn about and download." To the right of the search bar is a "Suggest a dataset" box with the text "Is there a dataset from the EU that you could not find in this portal? Please request the dataset >>".

Top Publishers [view all publishers >](#)

- Eurostat** (6053 datasets)
- European Environment Agency** (106 datasets)
- Joint Research Centre** (33 datasets)
- Fusion for Energy Joint Undertaking** (13 datasets)
- Publications Office** (13 datasets)
- Directorate-General for Health and Consumers** (12 datasets)
- Directorate-General for Competition** (6 datasets)
- European Asylum Support Office** (4 datasets)

Most viewed datasets [view all datasets >](#)

- Forest / Non-Forest Map 2000** (121 views)
- HICP - all items excluding energy, food, alcohol and tobacco** (10 views)

Recently updated datasets [view all datasets >](#)

- Enterprises using Internet for interaction with public authorities (NACE Rev. 2)** was updated on 06/11/13.
- Water abstracted for manufacturing industry** was updated on 06/11/13.
- Private expenditure in institutions (ISCED 5-6) as % of public and private funding** was updated on 06/11/13.
- Gross value added - NACE Rev. 1: L-P - current prices** was updated on 06/11/13.
- Psychiatric care beds in hospitals**

Zdroj: <http://open-data.europa.eu/>

publicdata.eu
Europe's Public Data **BETA**

Datasets Groups About Apps & Ideas Search

Search Europe's Public Data

Find datasets

Finance and Budgeting (436)
Social Questions (226) Environment (213)
Transportation (191)
Agriculture, Fisheries, Forestry (175)
Education and Communication (174)
Population (140)
Economy and Industry (113) Health (79)

View larger map >

Open Data Showcase

Featured Applications

ZNasichDani / From Our Taxes
ZNasichDani.sk uncovers who are influential pe...

OpenCorporates
OpenCorporates has taken one of the most impor...

Live London Underground tube map
It plots the current positions of all London U...

Zdroj: <http://publicdata.eu/>

Open Data Index

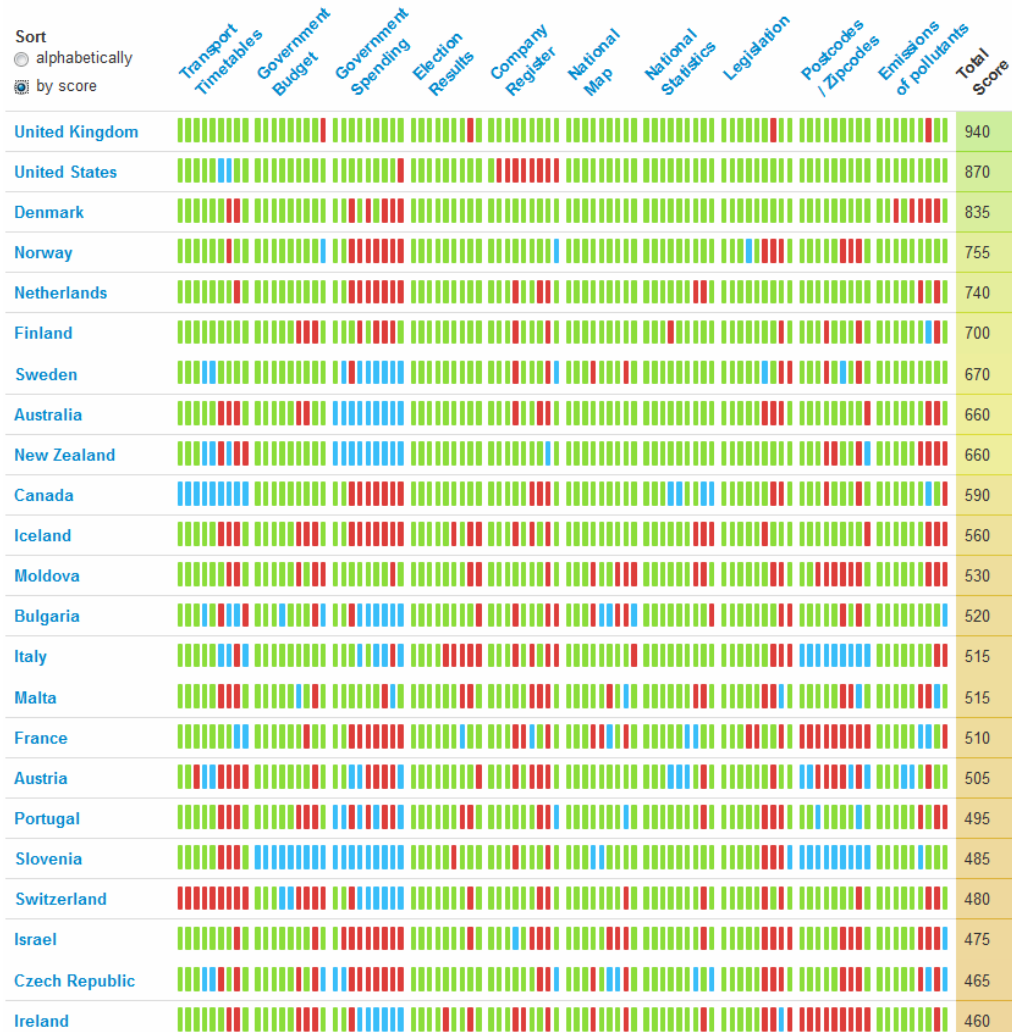
- vytváří organizace Open Knowledge Foundation (OKFN)
- monitoruje otevřenost vybraných skupin datových množin v jednotlivých zemích (i mimo EU)
 - např. jízdní řády, výsledky voleb, obchodní rejstřík, statistická data, emise, atd.
- nejedná se o garantovaný index, reporty za jednotlivé země mohou zasílat např. vybraní jedinci z neziskových organizací

Zdroj: <http://index.okfn.org>

Open Data Index

Key: ■ Yes ■ No ■ Unsure ■ No data

Sort
 alphabetically
 by score



Zdroj: <http://index.okfn.org>

Open Data Certificate

- služba provozovaná organizací Open Data Institute (ODI)
- certifikát je udělen datové množině na základě atributů její kvality
 - existence kurátora, existence podmínek užití, původnost a úplnost dat, strojová čitelnost, existence dokumentace (formátu, procesů zajištění kvality, ...), ...
- certifikát indikuje úroveň otevřenosti datové množiny
 - 4 úrovně: raw, pilot, standard, expert

Zdroj: <http://certificates.theodi.org>

Open Data Certificate – úrovně



Raw: A great start at the basics of publishing open data.



Pilot: Data users receive extra support from, and can provide feedback to the publisher.



Standard: Regularly published open data with robust support that people can rely on.



Expert: An exceptional example of information infrastructure.

Zdroj: <http://certificates.theodi.org>

Obsah

1. Co jsou otevřená data
2. Aktivity v oblasti otevřených dat v EU
3. Aktivity v oblasti otevřených dat v ČR
4. Výzvy využívání otevřených dat v ČR

Výzvy pro veřejnou správu ČR

1. Publikace prioritních datových množin
2. Sjednocení způsobu publikace
3. Analýzy potenciálu otevřených dat pro konkrétní orgány VS
4. Katalog otevřených dat ČR
5. Implementace novely směrnice 2003/98/ES
6. Komunikační strategie

Publikace prioritních datových množin

- Zkušenosti z jiných zemí ukazují, že otevřená data nejsou zcela zdarma a nekoordinované zveřejňování všech dat přináší řadu problémů (pracnost publikace, katalogizace a údržby, nepřehlednost, nejasná využitelnost, ...)
- **Možné řešení:** Vyhlásit např. formou usnesení vlády prioritní datové množiny a povinnosti jejich publikace odpovědnými orgány VS ČR k určitému datu
 - Usnesení doplnit Konceptí publikace otevřených dat VS ČR, která stanoví
 - postupy publikace a katalogizace,
 - požadavky na kvalitu a aktuálnost prioritních datových množin a
 - otevřené datové formáty, ve kterých mají být data publikována.
 - Nutnost koordinovat a kontrolovat odpovědné orgány
 - Seznam prioritních datových množin založit na odborných analýzách, pravidelně redigovat a rozšiřovat o nové datové množiny
 - v prvním kole jen několik málo datových množin tak, aby bylo možné dosáhnout požadované kvality a kontrolovat ji

Sjednocení způsobu publikace

- Řada orgánů VS ČR spravuje datové množiny stejného typu, které jsou vhodné k otevření. Pokud však každý takový orgán zveřejní svá data jiným způsobem, nebude možné data jednoduše integrovat a využívat dohromady.
 - Např. výsledky kontrol (ČOI již publikuje, ČTÚ plánuje publikovat, existuje řada orgánů, které by také mohly publikovat)
- **Možné řešení:** V rámci seznamu prioritních datových množin identifikovat ty, které jsou stejného typu (např. výsledky kontrol). U nich koordinovat použití společného datového formátu pro reprezentaci datových množin jednotlivých orgánů.

Analýzy potenciálu otevřených dat pro konkrétní orgány

- Orgány VS ČR by měly analyzovat, jaké datové množiny spravují a navrhnout, které z nich by měly být zveřejněny v otevřené podobě. Provést takovou analýzu je však poměrně náročné, je potřeba dostatečné know-how.
- **Možné řešení:**
 - Umožnit orgánům sdílet know-how; informovat o těch, které již takovou analýzu provedli; zveřejňovat analýzy na jednom místě
 - Již hotovou analýzu má Český telekomunikační úřad (ČTÚ)
 - Vytvořit šablonu (nebo osnovu) pro analýzy
 - specifikovat otázky, na které by analýzy měly odpovídat
 - lze vycházet z analýzy ČTÚ
 - Vybidnout vybrané orgány k vypracování analýz na základě šablony

Katalog otevřených dat ČR

- Aby otevřená data publikovaná na webových stránkách jednotlivých úřadů bylo možné využít, musí je být potenciální zájemce o data schopen jednoduše a rychle vyhledat.
- **Možná řešení:**
 - Centrální katalog otevřených dat VS ČR
 - Všechny orgány VS ČR katalogizují svá otevřená data v centrálním katalogu dat
 - Návrh centrálního katalogu popsán v Koncepti katalogizace otevřených dat VS ČR
 - Distribuovaný katalog otevřených dat VS ČR
 - Datové katalogy na webových stránkách jednotlivých orgánů VS
 - Standardizace formátu katalogizačního záznamu, používaných klasifikačních schémat, postupu katalogizace apod.
 - Možnost implementace standardu vytvářeného na úrovni EU (DCAT-AP)
 - Integrace datových katalogů a vyhledávání napříč datovými katalogy jednotlivých orgánů VS

Implementace novely směrnice 2003/98/ES

- Určit/vyložit, kdy je „možné a vhodné“ publikovat data v otevřeném a strojově čitelném formátu spolu s metadaty
- Zvolit vhodnou formu zajištění obecné zásady využitelnosti dat VS pro komerční i nekomerční účely
 - Vzorové podmínky užití
 - Implementace a doporučení některé z otevřených licencí používaných i ve světě
 - Zajištění zásady přímo ustanovením zákona
- Článek 9
 - „Členské státy učiní praktická opatření, která **usnadňují vyhledávání dokumentů dostupných pro opakované použití**, jako **seznamy fondů** hlavních dokumentů s příslušnými **metadaty**, pokud je to možné a vhodné, přístupné **on-line a ve strojově čitelném formátu**, a internetové portály, které jsou propojeny s těmito seznamy fondů. Kdykoli je to možné, vytvoří členské státy podmínky pro vícejazyčné vyhledávání dokumentů. “
 - **Možné řešení:** vybudování datového katalogu (výzva č. 4)

Komunikační strategie

- Lidé (ve VS i mimo ni) ne vždy správně nechápou otevřená data. Je třeba zajistit koordinaci aktivit otevřených dat ve VS. Je třeba zajistit, aby výsledky projektů, které jsou využitelné různými orgány VS, opravdu byly sdíleny a používány a aby jednotlivé orgány VS zbytečně nevynakládaly prostředky na řešení, která již byla vytvořena
- **Možná řešení:**
 - Komunikační strategie určující, jakými kanály o otevřených datech informovat a kde se o otevřených datech mohou zájemci dovědět
 - Vytvořit oficiální webové stránky VS ČR o otevřených datech, kde budou publikovány garantované informace, závazné pokyny a které mohou sloužit jako platforma pro spolupráci osob a orgánů napříč veřejnou správou