

# Open data

Open data jsou svým způsobem převratem fungování veřejné správy. V zájmu zvyšování kvality služeb poskytovaných občanů zveřejňují úřady volně svá data. Kurz nabízí vhled do výhod a rizik open dat a uvádí jednoduchý návod jak krok za krokem open data zavést na jakýkoli úřad veřejné správy.

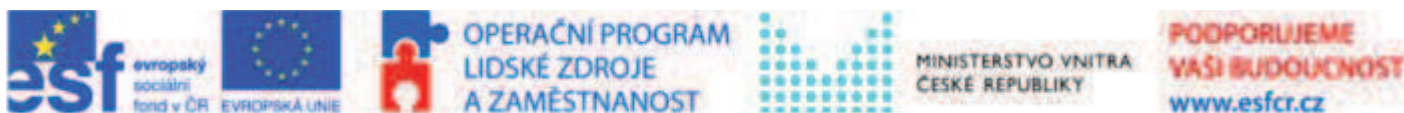
Cílem kurzu je vysvětlit účastníkům podstatu open dat, jejich přínos pro společnost, ale i možná rizika s nimi spojená. Naučit účastníky rozlišovat různé formy dat a open dat, pomoci účastníkům osvojit si odbornou terminologii související s problematikou open dat, předat účastníkům návod, jak krok za krokem open data zavést na úřadu veřejné správy, poradit účastníkům, kam (na koho) se obrátit s žádostí o pomoc při zavádění open do praxe úřadu, předat účastníkům informace o užitečných zdrojích informací k tématu kurzu (související legislativě, webových stránkách a odborné literatuře) a o již existujících zajímavých aplikacích open dat.

08:00:00

## Osnova modulu

- 1 | Informace ke studiu
  - 1.1 | Význam piktogramů
  - 1.2 | Definice použitých zkratk
- 2 | Teoretická část
  - 2.1 | Úvod do problematiky
  - 2.2 | Co jsou open data
  - 2.3 | Forma a kvalita open dat
  - 2.4 | Strojově zpracovatelný formát
    - 2.4.1 | Jedna hvězda
    - 2.4.2 | Dvě hvězdy
    - 2.4.3 | Tři hvězdy
    - 2.4.4 | Čtyři hvězdy
    - 2.4.5 | Pět hvězd
  - 2.5 | Přínosy open dat
  - 2.6 | Související legislativa
  - 2.7 | Slovníček pojmů
- 3 | Praktická část
  - 3.1 | Aplikace
  - 3.2 | Způsob zavedení do praxe
    - 3.2.1 | Výběr dat
    - 3.2.2 | Ochrana osobních údajů
    - 3.2.3 | Licence
    - 3.2.4 | Výběr formátu
    - 3.2.5 | Export dat
    - 3.2.6 | Popis dat
    - 3.2.7 | Katalogizace
    - 3.2.8 | Aktualizace dat
    - 3.2.9 | Úpravy dat
    - 3.2.10 | Zvyšování úrovně datasetů
- 4 | Kontrolní otázky
- 5 | Doporučené zdroje informací
- 6 | Shrnutí

# 1 | Informace ke studiu











Vzdělávání v oblasti základních registrů a dalších kmenových projektů eGovernmentu, registrační číslo projektu: CZ 1.04/ 4.1.00/A3.00004

Tento kurz byl vytvořen v rámci projektu financovaného z prostředků Evropského sociálního fondu ČR, operačního programu Lidské zdroje a zaměstnanost a je součástí souboru deseti eLearningových kurzů:

1. Použití základních registrů
2. Agendové informační systémy a Informační systémy veřejné správy
3. Služby soukromoprávního sektoru
4. Zabezpečení přístupu k datům
5. Zabezpečení připojení AIS
6. Programové období 2014 - 2020
7. CzechPOINT@office
8. Open data
9. Datové schránky
10. Doručování datovými schránkami

## 1.1 | Význam piktogramů

V kurzu se budete setkávat s piktogramy, které vám usnadní orientaci v textu, upozorní vás na důležité informace, právní předpisy, doporučenou literaturu apod. Piktogramy jsou společné pro všechny kurzy, je tedy možné, že s některými z nich se v tomto kurzu neseznáte. Přesto je vhodné se před zahájením studia se všemi seznámit.

	příklad z praxe		odkaz na paragraf
	dobry tip		důležitá informace
	odkaz na právní předpis		doporučená literatura
	kontrolní otázky		shrnutí učiva

## 1.2 | Definice použitých zkratk

ZKRATKA	VÝZNAM
PDF	Portable Document Format
CSV	Comma-separated values; hodnoty oddělené čárkami
URL	Uniform Resource Locator

## 2 | Teoretická část a legislativa

### 2.1 | Úvod do problematiky



Open data představují velmi specifický způsob zvýšení kvality výkonu veřejné správy na úrovni transparentnosti a otevřenosti. Obvykle nevyžadují žádné změny legislativy ani interního fungování státních úřadů a lze je zavést i bez nákupu jakýchkoli služeb či nástrojů.

Open data jsou poměrně mladou myšlenkou, která se dosud největšího rozvoje dočkala v zemích, jako jsou Velká Británie, USA, Estonsko a např. Keňa. Rozvoj open dat v posledních letech zaznamenávají také naši sousedé, konkrétně Německo, Rakousko a Slovensko. Zejména v Německu a Rakousku spolu začala mnohá města soutěžit, kdo zveřejní data první a kdo jich zveřejní více.

Samotná myšlenka open dat spočívá na velmi prostém principu a je i velmi snadno realizovatelná. Jedná se v zásadě o pouhé zveřejnění tabulek a databází, které jednotlivé úřady spravují volně na internetu. Aby se pak tato data mohla označovat za otevřená (open), stačí splnit několik málo podmínek, které budou rozebrány dále.

Jedná se tedy o to, že vedle tradičního zveřejňování statistických ročenek jsou občanům poskytnuta originální zdrojová data a díky tomu si kdokoli může vytvořit, jakékoli statistiky, které jej napadnou. Nad takto otevřenými daty nakonec soukromé subjekty (firmy a nadšenci) vytváří internetové a mobilní aplikace, které mohou kombinovat data různých úřadů a v přívětivé podobě je předkládají občanům.

V ČR se open data nacházejí zatím v začátcích, ale již se objevily první vlaštovky, které ke zveřejňování svých evidencí v podobě open dat přistoupili. Iniciativně se k myšlence přihlásila vláda v roce 2011, kdy přistoupila k Partnerství pro otevřené vládnutí (Open Government Partnership) a zavázala se do konce roku 2012 zveřejnit prvních 10 databází a do března 2013 vytvořit jednotný katalog otevřených dat. Z těchto plánů bylo naplněno pouze zveřejnění databáze výsledků voleb Českým statistickým úřadem. Ke zveřejňování open dat se však dobrovolně přihlásili jiní. Příkladem mohou být Česká obchodní inspekce (kontroly, pokuty, zabavené zboží ad.) a Český telekomunikační úřad (ceny volání, seznam veřejných telefonů, personalistika, kontroly a pokuty). Ze samospráv se jako první přihlásila v červenci 2013 Opava.

## 2.2 | Co jsou Open data



Open data, nebo také Otevřená data, jsou definována jako volně dostupná data bez omezení jejich dalšího využití zveřejněná ve strukturované, strojově zpracovatelné podobě.

Pro zjednodušení jsou otevřenými data každá, která jsou:

1. snadno dostupná (volně přístupná na internetu),
2. úplná (nejedná se o výsledné statistiky, ale o data ze kterých statistiky vznikají),
3. strukturovaná a strojově čitelná (nejčastěji jako tabulka s neměnnými sloupci nebo v jazyce XML),
4. poskytnutá pod otevřenou licenci (umožňují další zpracování dat).

Výhody otevřenosti dobře vystihl ve své přednášce na konferenci TED britský premiér David Cameron (dostupné na [http://www.ted.com/talks/david\\_cameron](http://www.ted.com/talks/david_cameron) i s českými titulky, 14 min.)

Zveřejnit open data tedy vyžaduje pouze:

- mít k dispozici data, která je možné zveřejnit,
- vystavit je na webovou stránku úřadu v jejich úplné podobě,
- napsat u dat, za jakých podmínek je možné jejich použití (licenci) a
- připojit alespoň základní popis těchto dat.

## 2.3 | Forma a kvalita open dat



Open data jsou zveřejňována v různých stupních kvality, přičemž čím kvalitnější forma open dat je zvolena, tím méně práce je zapotřebí pro uživatele dat, aby je mohli dále využít. Nejvyšší úroveň, na kterou je možné open data dostat je forma nazývaná Propojitelná otevřená data (Linked Open Data).

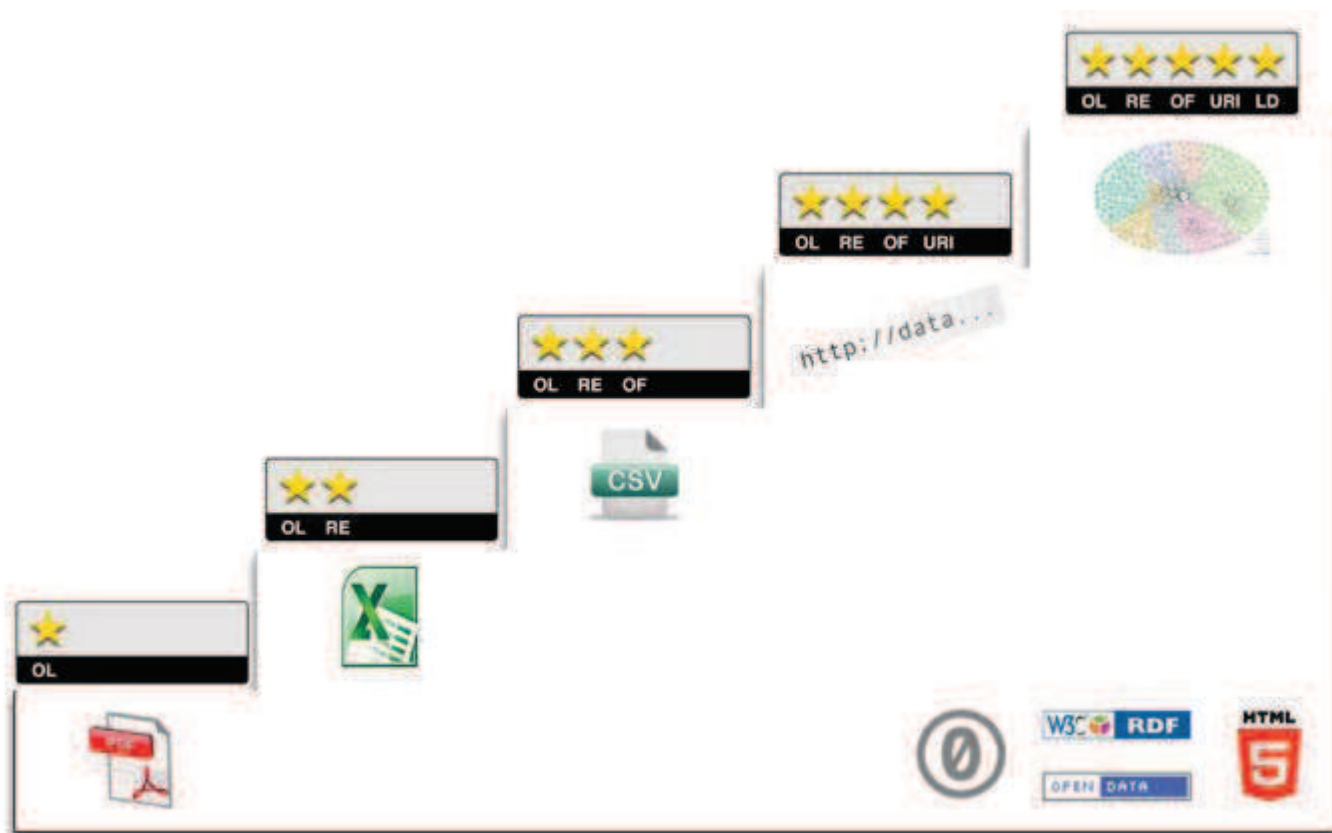
Za propojitelná je možné open data považovat v okamžiku, kdy má každý záznam v datech obsažený svůj jednoznačný identifikátor v podobě URL adresy (např. <http://linked.opendata.cz/resource/domain/coi.cz/sanction/68392> v případě Sankce České obchodní inspekce s ID číslem 68392) přičemž na této URL adrese jsou k dispozici data o tomto záznamu ve strojově zpracovatelné podobě, která není určena pro běžné uživatele. Princip spočívá v tom, že formou hypertextových odkazů jsou propojena data různých úřadů z různých zemí. Ten kdo zveřejňuje data tak nemusí sám zveřejňovat číselníky, které používá (třeba NUTS spravovaný statistickým úřadem) nebo název a sídlo společnosti, již se záznam dotýká (stačí odkaz na IČ v obchodním rejstříku). V této podobě je uživatel dat (programátor aplikace) schopen snadno zjistit další informace související, aniž by musel sám dohledávat, kdo a kde je zveřejnil. Jedná se tedy o stejný princip, jaký má současný web, kdy jsou webové stránky vzájemně provázány sítí hypertextových odkazů.

Forma propojitelných otevřených dat je sice nejvstřícnější variantou vůči možným uživatelům dat, ale zároveň je pro poskytovatele dat, tedy veřejnou správu, tou nejpracnější a nejdražší. To je důvodem, proč v ČR dosud žádný úřad nepublikuje svá data jako propojitelná. Nicméně experimentálně jsou do propojitelné podoby převáděna data na českých vysokých školách (seznam dat, zpracovaných Karlovou Univerzitou, je dostupný na <http://opendata.cz/cs/linked-data>).

Propojitelná otevřená data jsou náročným úkolem, a proto je dobré nenechat se jejich složitostí zastrašit. Vhodné je pro první zveřejnění dat použít jednodušší formu a na propojitelná se propracovat v průběhu let.

## 2.4 | Strojově zpracovatelný formát

Když se hovoří o zveřejňování nebo publikaci dat, tak se tím myslí uložení dat na internet, tak aby byla data volně přístupná. Data, která jsou zveřejňována, nemusí být jen ve formě open data. Forem, v jakých jsou data běžně zveřejňována, je celá řada. Z pohledu open dat je můžeme rozdělit například podle členění pěti hvězd. Toto členění uděluje datům 1 až 5 hvězd, přičemž čím více hvězd, tím lépe.



Zdroj: <http://5stardata.info>

Data ve formátech ze skupin s jednou až třemi hvězdami jsou již běžné i v ČR. Na úrovni 5 hvězd, tedy propojitelná data, není žádný dataset zveřejňovaný českými úřady veřejné správy. Úroveň 4 hvězd, tj. formát RDF/TTL nabízí k červnu 2014 pouze Česká obchodní inspekce. Obecně platí, že formát 1-3 zvládne bez zbytečných nákladů zavést kdokoli. Formáty 4-5 již vyžadují spolupráci s odborníky na open data z akademické (např. MFF UK a FIS VŠE - <http://opendata.cz>), neziskové (např. Fond Otakara Motejla - <http://otevrenadata.cz>) nebo soukromé sféry.

## 2.4.1 | Jedna hvězda

Jedna hvězda náleží datům v nestrukturované podobě. Tedy v PDF, v MS Word nebo jiném textovém editoru. Tyto formáty jsou vhodné pouze pro zpracování člověkem, tedy čtení. Automatické zpracování je prakticky vyloučeno. Přestože by byla tato data zveřejněna pod otevřenou licenci, není možné je označit za open data, protože je není možné snadno znovu použít. Může jít ovšem o velmi užitečný doplněk k open datům. Pokud například úřad zveřejní open data o správních rozhodnutích, kde jsou uvedeny základní údaje, může uvést odkaz na textový soubor s plným zněním rozhodnutí.

## 2.4.2 | Dvě hvězdy

Dvě hvězdy označují data, která jsou nejen zveřejněna pod otevřenou licenci, ale současně jsou i ve strojově zpracovatelném formátu jako je tabulka MS Excel. Na této úrovni je již možné začít hovořit o open datech, přestože podle některých autorů lze označovat za open data až ty na úrovni tří hvězd.

## 2.4.3 | Tři hvězdy

Tři hvězdy patří datům, která jsou na rozdíl od těch se dvěma hvězdami v otevřeném formátu. To znamená, že jsou zcela nezávislé na výrobci konkrétního programového vybavení. Nejtypičtějším příkladem je formát CSV, tedy „čistý“ textový formát s hodnotami oddělenými např. středníkem. Uložení do formátu CSV podporují prakticky všechny kancelářské balíky (MS Office, Open Office), tudíž je přechod ze dvou na tři hvězdy velice jednoduchý a nezabere více než pár minut, v závislosti na množství dat.

## 2.4.4 | Čtyři hvězdy

Čtyři hvězdy jsou označením pro mezi krok k propojitelným datům. Tento mezi krok spočívá ve zveřejnění ve formátu založeném na značkovacím jazyce XML, nejčastěji RDF nebo TTL. Data v tomto formátu již mají vlastní URL adresu pro každý záznam, ačkoli na této adrese nic není a v prostředí webu neexistuje. Obdobně jsou využity odkazy na jiná data. Například číselník zveřejněný ve vlastním datovém souboru (datasetu) s vlastní adresou nebo údaj o adrese zveřejněný v Registru územní identifikace. Zde se místo uvedení kódu existujícího v jiném datasetu (např. NUTS CZ0806) uvede URL adresa tohoto kódu, která je ve čtyřhvězdičkovém souboru správce tohoto číselníku (např. <http://data.czso.cz/resource/NUTS/CZ0806>).

## 2.4.5 | Pět hvězd

Pět hvězd patří formě propojitelných otevřených dat. Jedná o podobu, kdy na konkrétní webové adrese jsou informace o jednotlivých záznamech dat. Při využití vizualizačního nástroje může webová stránka zobrazovat například takto vypadající záznam o pokutě:

Attributes	Values	
rdf:type	coicz:Sanction	<i>Typ dat</i>
skos:notation	68392	<i>ID pokuty</i>
dcterms:valid	2013-05-25(xsd:date)	<i>Datum právní moci</i>
coicz:paragraf	§11	<i>Za</i>

	§ 5a	<i>porušení paragrafů</i>
coicz:zakon	Zák. 634/1992	<i>Za porušení zákona</i>
s:result	<a href="http://linked.opendata.cz/resource/domain/coi.cz/sanction/68392/price-specification">http://linked.opendata.cz/resource/domain/coi.cz/sanction/68392/price-specification</a>	<i>Odkaz na výši pokuty</i>
void:inDataset	<a href="#">Sankce ČOI</a>	<i>Odkaz na dataset pod který náleží</i>
is s:result of	<a href="http://linked.opendata.cz/resource/domain/coi.cz/check-action/221212100029901">http://linked.opendata.cz/resource/domain/coi.cz/check-action/221212100029901</a>	<i>Odkaz na související kontrolu</i>

## 2.5 | Přínosy Open dat

Vzhledem k tomu, že jednou z vlastností open dat je strojová zpracovatelnost, bývá obvyklé, že pro člověka je práce s open daty nepohodlná. To je přirozené, protože cílovou skupinou open dat nejsou koneční uživatelé z řad běžných občanů, ale spíše technicky zdatné osoby. Užití open dat lze rozdělit do dvou hlavních skupin:

- datová žurnalistika a
- webové a mobilní aplikace.



Vzhledem k malému množství open dat v ČR je o ně velký zájem mezi jejich uživateli. I přes tento zájem je třeba počítat, že po zveřejnění nových open dat může trvat jeden až dva roky, než budou skutečně využita. Je třeba počítat s tím, že si potenciální uživatelé musí všimnout nově dostupných dat, musí se ujistit, že budou dostupná dlouhodobě, pak vymyslet co vše by rádi s daty udělali a pak naprogramovat vše potřebné. Čím žádanější data, tím dříve budou využita. Například o výsledky voleb nebo jízdní řády je enormní zájem, naopak personální statistiky mohou ležet ladem delší dobu, než je zveřejní i jiné úřady aby je bylo s čím srovnávat.

Datová žurnalistika je dynamicky se rozvíjející obor, kde v dostupných datech hledají novináři a jiní nové závislosti a zajímavé informace, na jejichž hledání nemá úřad čas. V současnosti nejaktivnější datovou žurnalistiku v ČR má na svědomí redakce iHned.cz. Jejich výstupy jsou dostupné na adrese <http://data.blog.ihned.cz/>

Webové a mobilní aplikace vytváří soukromé subjekty, které vidí možnost zisku například na reklamě nebo na prodeji mobilních aplikací anebo neziskové subjekty v rámci akademické činnosti nebo z prostého nadšení pro věc. V obou případech ale platí, že dokud nejsou data k dispozici, nemůže žádná aplikace vzniknout, a proto musí být proces zveřejnění open dat odstartován na straně veřejné správy.

Z pohledu úřadu mohou open data přinést pokles počtu dotazů dle zákona 106/1999 Sb., o svobodném přístupu k informacím, a dotazů od novinářů, kteří si budou moci rychleji vše dohledat sami.

Z pohledu ekonomiky se jedná o vznik nových pracovních míst a nárůst HDP a tím vyšší odvod daní. Konkrétní čísla jsou dostupná zatím jen ze zahraničí:

- Přínos do britské ekonomiky byl již v roce 2010 vyčíslen na 2 miliardy liber ročně a v roce 2011 už mělo jít o 16 miliard liber.
- Zveřejnění dat o městské hromadné dopravě v Londýně vedlo ke vzniku 500 projektů a tím 5 tisících pracovních míst.
- Potenciál ročního obrátu nad open daty v celé Evropské unii se odhaduje na 3 500 000 000 000 Kč (3,5 bilionu).

## 2.6 | Související legislativa



V České republice nejsou k červnu 2014 open data v legislativě nijak zakotvena, ale zároveň nejsou ani omezena s výjimkou chráněných informací (bezpečnost, osobní údaje apod.) Nicméně připravuje se implementace Směrnice Evropského parlamentu a Rady 2013/37/ES, kterou se mění Směrnice Evropského parlamentu a Rady 2003/98/ES o opakovaném použití informací veřejného sektoru.

Implementace se pravděpodobně bude provádět skrze novelu zákona o informačních systémech veřejné správy a zákona o svobodném přístupu k informacím. Příprava těchto novel je v době psaní tohoto dokumentu v začátcích a tak není jasné, zda budou open data pro některé oblasti uložena jako povinnost nebo zůstanou zcela dobrovolná. Zároveň se také diskutuje odklad případných povinností pro samosprávy až o dva roky. Jistým je nyní pouze to, že bude vymezen pojem open data (možná v překladu otevřená data) a bude zřízen státní katalog open dat.

Diskutována je také možnost, že čerpání některých dotací bude podmíněno zveřejněním open dat.

Směrnice Evropského parlamentu a Rady 2003/98/ES o opakovaném použití informací veřejného sektoru ve znění pozdější úpravy Směrnicí Evropského parlamentu a Rady 2013/37/ES. Konsolidované znění je dostupné na <http://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:02003L0098-20130717&qid=1401533406548&from=CS>

- Zákon číslo 101/2000 Sb., o ochraně osobních údajů
- Zákon číslo 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským
- Zákon číslo 365/2000 Sb., o informačních systémech veřejné správy
- Zákon číslo 106/1999 Sb., o svobodném přístupu k informacím

## 3 | Praktická část

### 3.1 | Aplikace



Pokud se veřejná správa rozhodne, že nebude sama investovat peníze do vlastního vývoje aplikací, ale raději zveřejní open data a počká, co s nimi kdo udělá, tak sice nemá žádnou kontrolu nad tím co a za jak dlouho vznikne, ale zato vše vznikne zdarma a v konkurenčním prostředí. Platí totiž, že nejsme schopni domyslet, jakými způsoby by bylo možné data prezentovat veřejnosti. Konkurenční prvek vede k tomu, že všichni, kteří nad daty postaví nějakou aplikaci, jsou závislí na návštěvnosti a tak



motivování své služby stále zlepšovat, jinak se kdykoli může objevit někdo jiný, kdo stejná data využije zajímavějším způsobem.

Již nyní je k dispozici celá řada aplikací, které vznikly často i přes překážky, které původci dat vytvořili, když data neposkytli jako open data. Příklady zajímavých aplikací jsou například:

- [Rozklikávací rozpočet](#) – kde lze snadno zjistit kam plynou peníze ze státního rozpočtu
- [Rozpočet obce](#) – kde jsou k dispozici informace o hospodaření obcí
- [Všechny zakázky](#) – s přehledem veřejných zakázek na jednom místě
- [Mapa kriminality](#) – s přehledem o zjištěných trestných činech v různých regionech
- [Výsledky kontrol](#) – kde jsou k dispozici informace o kontrolách a pokutách (zatím jen ČOI)

V zahraničí vznikly aplikace jako:

- V USA mobilní aplikace s přehledem jak hlasují jednotliví kongresmani nebo mapa vztahů mezi politiky a podnikateli
- Ve Vídni mobilní aplikace s mapou veřejných záchodků
- V Británii aplikace pro výběr vhodné lokality, kde si koupit dům dle desítek statistik
- A mnohé další aplikace vznikají mimo jiné i v rámci soutěže Společně otevíráme data, do jehož prvního ročníku 2013 se přihlásilo 25 aplikací.

## 3.2 | Způsob zavedení do praxe



Pokud se rozhodnete zveřejnit některé své databáze a evidence jako open data, je to možné učinit s pracností jedné osoby v rozsahu 20-30 hodin. Stačí následovat tyto kroky, které vás provedou procesem zveřejnění open dat:

1. Příprava zveřejnění:
  - a. Výběr data vhodných ke zveřejnění
  - b. Pro vybraná data zhodnocení, zda jejich zveřejnění není v rozporu s ochranou osobních údajů
  - c. Příprava a výběr licence (pravidel použití)
2. Zveřejnění dat
  - a. Volba vhodného formátu dat
  - b. Export dat
  - c. Popis dat (metadata)
  - d. Katalogizace
3. Provoz a rozvoj
  - a. Pravidelné vystavování aktuálních dat
  - b. Úprava dat dle námětů od uživatelů
  - c. Zvyšování kvality dat

## 3.2.1 | Výběr dat



Zcela první krok pro zveřejnění open dat musí být vždy výběr dat, která lze zveřejnit. Je zapotřebí zmapovat, jaké evidence a databáze máte k dispozici a jste schopni je bez zbytečných nákladů převést do podoby open dat. Takovým zbytečným nákladem je například situace, kdy data nejsou evidována ve strukturované podobě, ale třeba v mnoha PDF dokumentech uložených ve spisové službě. V takové situaci, ale nemůže s daty pracovat ani sám úřad a je tedy ke zvážení, zda nezavést evidenci takových dat. Zbytečným nákladem ale zcela jistě není případ, kdy data obsahují velké množství chyb nebo data ze dvou souvisejících evidencí spolu nekorespondují nebo si dokonce odporují. Tehdy je nutné přistoupit k „čištění“ dat jejich opravami bez ohledu na to, zda nakonec budou zveřejněna nebo ne, protože databáze plná chyb je zcela bezcenná i pro interní fungování úřadu.

Jakmile zmapujete všechny možné zdroje dat a údaje, které obsahují, nastává okamžik rozhodnutí, která data zveřejnit. V první řadě musí jít o data v jednotné strukturované podobě, tedy nejlépe v tabulce nebo podobném formátu. V případě dat, která nevznikají z činnosti úřadu, ale jsou přijímána od soukromých subjektů, je nutné zohlednit vliv práva duševního vlastnictví. Tedy, zda nejsou data autorským dílem (prostě číselné údaje jím nejsou) a pokud jsou, tak zdali má úřad oprávnění tato data volně zveřejnit.

Nesmí být opomenuta ani ochrana osobních údajů rozebraná dále.

Výstupem tohoto prvního kroku by měl být seznam dat, která je úřad schopen zveřejnit.

## 3.2.2 | Ochrana osobních údajů



V souladu se zákonem 101/2000 Sb., o ochraně osobních údajů nesmí být na internetu zveřejněny žádné osobní údaje s výjimkou případů, kdy je toto zveřejnění důležitým veřejným zájmem. Veřejný zájem je v tomto případě příliš neurčitým pojmem, který může být vykládán všelijak. Stanovisko Úřadu pro ochranu osobních údajů (ÚOOÚ), který je oprávněn zveřejnění osobních údajů pokutovat až do výše 10 mil. Kč, lze za veřejný zájem považovat informace o ohrožení zdraví a života. Nicméně je sporné, zda by takovýto výklad obstál u soudu.

Za osobní údaj považuje ÚOOÚ jakoukoli kombinaci údajů, které mohou vést k jednoznačné identifikaci konkrétní osoby. To znamená, že mezi osobní údaje se řadí také:

Informace o platech zubařů dle okresů, protože může existovat okres, ve kterém je jediný zubař a tím by byl jednoznačně identifikovatelný.

- Osobním údajem může být překvapivě i IČ podnikající fyzické osoby (živnostníka), protože k IČ je možné dohledat konkrétního občana, který podniká na své jméno.
- Informace o osobách s uvedením jejich pohlaví, věku a obce trvalého bydliště, protože z těchto údajů může být v menších obcích identifikovatelná jediná osoba.

Jakmile je k osobnímu údaji uvedena informace, která může představovat zásah do soukromí, není možné takové informace zveřejnit. Samotné uvedení věkové struktury obyvatelstva obce zásahem do soukromí není. To znamená, že je možné uvést, že v obci žije jedna žena ve věku 57 let. Ale již je porušením ochrany osobních údajů informace, že v této obci trpí 100 % žen ve věku 57 let rakovinou. V případě živnostníků je v pořádku zveřejnit seznam podnikatelů v obci, ale již může být považováno za zásah do soukromí dle

stanoviska ÚOOÚ, pokud se zveřejnění, kteří živnostníci dostali pokutu za okrádání zákazníků. Výjimku z ochrany osobních údajů má specifická legislativa. Tedy pokud jiný zákon výslovně uvádí, že je údaj možné zveřejnit, má tato úprava přednost před zákonem o ochraně osobních údajů.

Možným pohledem na posouzení, zda je možné data zveřejnit, je optika skrze zákon 106/1999 Sb., o svobodném přístupu k informacím. Tento zákon umožňuje komukoli požádat o poskytnutí informací a dat, které má úřad k dispozici. Úřad pak musí posoudit, zda je možné požadované informace poskytnout a to i z pohledu ochrany osobních údajů. Je tedy možné posoudit data i podle toho, zda jsou již nyní volně poskytována na základě žádostí dle zákona 106/1999.

V případě, že si sami nedokážete s jistotou odpovědět na otázku, zda je možné data zveřejnit nebo spadají pod ochranu osobních údajů, je nejlepší obrátit se přímo na ÚOOÚ s žádostí o vydání stanoviska k vašim konkrétním datům. Stanovisko sice není závazné, ale není pravděpodobné, že by ÚOOÚ pokutoval jednání, které sám schválil. V žádosti o stanovisko uveďte, jak by z dat mohlo být dovozeno, o jakou konkrétní osobu se jedná a jaké všechny údaje hodláte zveřejnit a nejlépe i jaký je veřejný zájem pro zveřejnění těchto údajů. Žádost také musí obsahovat požadavek na jednoznačné schválení, zda je možné uvedená data poskytovat paušálně pro všechny možné záznamy, které mohou být obsaženy.

V případě, že data spadají pod ochranu osobních údajů, je třeba z nich osobní údaje vypustit, tedy tzv. anonymizovat. To se dá provést neuvedením těch specifických údajů, které umožňují jednoznačnou identifikaci osoby. Tedy neuvádět obec ale jen okres nebo neuvádět okres, ale jen kraj nebo neuvádět IČ fyzických osob.

### 3.2.3 | Licence



Jednou z hlavních součástí open dat je jejich poskytnutí pod otevřenou licenci. Důvodem je, že česká legislativa neumožňuje vzdání se práv k dílu. Proto je třeba specifikovat za jakých podmínek je možné s datysety pracovat a využívat je.

Licence by měla umožnit:

- kopírovat a dále šířit data,
- využívat data v jiných dílech a
- využívat data pro komerční účely.

Ze standardizovaných licencí se jako nejvhodnější pro české podmínky považuje:

- Open Data Commons Public Domain Dedication and Licence (ODC PDDL) v plném znění na <http://opendatacommons.org/licenses/pddl/1.0/>
- a nově také Creative Commons od verze 4.0, která by měla být v českém znění dostupná od září 2014 na <http://www.creativecommons.cz/>

Nevýhodou této licence PDDL je, že je v plném znění dlouhá na 6 stran a není oficiálně dostupná v českém jazyce. Licence CC 4.0 pak dosud není ve finální české verzi zatím dostupná. Zřejmě z těchto důvodů se úřady, které již open data publikují, přiklonily k vlastní licenci. Došlo zde k jisté neformální standardizaci licence, původně použité [Českým statistickým úřadem](#). ČOI i ČTÚ si pak licenci mírně uzpůsobili svým podmínkám, ale v obecné podobě stačí jako otevřenou licenci použít takovéto znění podmínek užití uvedené na webových stránkách:

*„Data zveřejněná na webových stránkách poskytovatele, smí uživatel:*

- šířit – kopírovat, distribuovat a sdělovat veřejnosti,
- užívat a citovat,
- užívat komerčně,

a to při dodržení následujících podmínek:

- jako zdroj dat musí být uveden poskytovatel (nikoli však takovým způsobem, aby vznikl dojem, že poskytovatel podporuje uživatele dat nebo způsob, jakým jsou open data využita),
- open data poskytovatele mohou být použita pouze takovým způsobem, který nijak nemění nebo neupravuje jejich význam/smysl.

Uživatel dále bere na vědomí, že:

- v případě dalšího šíření open dat poskytovatele je uživatel povinen uvést znění těchto podmínek užití, nejlépe přímým odkazem na tuto webovou stránku,
- tyto podmínky užití neplatí pro dokumenty, na něž se vztahují práva duševního vlastnictví třetích stran,
- poskytovatel neodpovídá za případné škody vzniklé dalším užitím a zpracováním open dat, jakož i za škody způsobené případnými chybami, které by open data mohla obsahovat,
- poskytovatel si vyhrazuje právo z provozních, odborných či jiných důvodů dočasně omezit přístup k zveřejňovaným datům, a to i bez předchozího upozornění. “

### 3.2.4 | Výběr formátu

V otázce formátu dat, nebo jinak řečeno, toho jak budou data zveřejněna, je třeba vyřešit několik otázek:

#### 1. Jaký datový formát?

Aby bylo možné data označovat za open data, je třeba, aby byla v rozsahu 3-5 hvězd. S ohledem na potřebu hlubších znalostí pro publikování data o 4 nebo 5 hvězdách se nedoporučuje začínat rovnou na této úrovni. Proto je pro prvotní zveřejnění open dat nejlepší poskytovat je ve formátu CSV. Toho je možno dosáhnout uložením dat např. do formátu excelové tabulky a z ní Uložit jako... formát CSV. Pokud je to možné, je vhodné doplnit každý záznam (řádek) o jeho unikátní identifikátor.

#### 2. V kolika souborech data zveřejnit?

Pokud zveřejňujete jednu excelovou tabulku, pravděpodobně stačí ji zveřejnit jen jako jeden soubor. Pokud ale budete zveřejňovat data z databáze je pravděpodobné, že bude nejlepším způsobem zveřejnění několika souborů, které se budou na sebe navzájem odvolávat. Podobně může být v samostatném souboru uložen číselník, který je používán v datech uložených v dalším souboru.

#### 3. Jakým způsobem aktualizovat data?

Dle povahy dat je možné data aktualizovat trojím způsobem:

- a. Jeden soubor – při každé aktualizaci je smazán původní soubor a nahrazen novým, který obsahuje původní i nová data. Tato varianta je vhodná zejména v případech, kdy může docházet ke změnám ve starých datech.
- b. Pro každé období samostatných soubor – při každé aktualizaci je zveřejněn nový soubor s daty za období navazující na to, které již zveřejněno je. Např. pokud jsou data zveřejňována jednou ročně, tak pro každý rok budou data v samostatném souboru, který ale bude mít stejnou strukturu, jako ten předchozí.
- c. On-line přístup k datům – použije se v případech, kdy k tomu existují technické prostředky anebo je povaha dat taková, že se mění velice často. Vyžaduje vytvořit podmínky pro přístup

do databáze skrze aplikační rozhraní prostřednictvím sítě internet, což vyžaduje investiční náklady a je vhodné jen pro úzký výběr dat.

#### 4. Jak často aktualizovat data?

Frekvence aktualizací dat je zcela na rozhodnutí jejich poskytovatele. Zvolená frekvence by měla zohledňovat to, jaký je zájem o nejnovější data a jak častou jsou tato data ve zdrojovém systému měněna a doplňována.

#### 5. Jak stará data ještě zveřejňovat?

Z pohledu uživatelů data je žádoucí, aby byla dostupná data za co nejdelší časové období. V případě, kdy jsou v datech obsaženy i osobní údaje, musí být zohledněno „právo na zapomenutí“, jinak řečeno po deseti letech už by neměly být úřadem poskytovány informace o méně závažných prohřešcích konkrétní osoby.

### 3.2.5 | Export dat

Konkrétní způsob exportu dat je zcela závislý na systému, z něž mají být data exportována. Pro samotné zveřejnění bývá obvykle zapotřebí připravit postupy či skripty pro převod dat do podoby vhodné pro zveřejnění. Jedná se například o nástroje pro převod dat do správného formátu, anonymizaci nebo jiné úpravy dat. V případě, že systém umožňuje exportovat data do formátu .xls (starý MS Excel), je také nutné mít na paměti, že omezení tohoto formátu je jen na 65.536 řádků a pro export většího množství dat je proto nezbytné použít jiný formát včetně nového MS Excel s příponou .xlsx.

### 3.2.6 | Popis dat

Open data jsou v základní podobě zveřejňována v datových souborech, např. v excelových tabulkách a CSV. Tyto soubory se nazývají datasety nebo také datové sady, či datové množiny. Aby bylo možné open data využívat i lidmi, kteří nejsou odborníky v odvětví činnosti úřadu, je nezbytné popsat datasety, aby bylo zřejmé, čemu se věnují a co obsahují.

Informace o datasetu by měly obsahovat minimálně:

- název datasetu,
- datum poslední aktualizace,
- periodicitu (frekvenci) aktualizace,
- popis, jaká data jsou v datasetu obsažena,
- rozklíčování, jakou informaci obsahují jednotlivé sloupce datasetu (vhodné umístit jako stažitelný soubor, na který je možné se odkazovat),
- název poskytovatele a původce (autora) datasetu,
- kontakt pro případné dotazy,
- odkaz na podmínky užití (licenci).

Čím podrobněji a přesněji je dataset popsán, tím snazší je práce s ním a předejde se tak chybné interpretaci dat nebo doplňujícím dotazům.

### 3.2.7 | Katalogizace

Protože jsou data zveřejněna na webu jejich poskytovatele, bylo by pro uživatele dat velmi náročné hledat, kde jsou open data dostupná, zvláště pokud ani neví, která instituce je za sběr takových dat zodpovědná. Z toho důvodu existují Katalogy dat. Katalogů existuje ve světě velké množství a většina si mezi sebou vzájemně vyměňuje informace, takže přidáním záznamu do jednoho katalogu se záznam časem objeví i v ostatních.

Katalog dat obvykle shromažďuje informace o adrese, kde jsou ke stažení open data, co tato data obsahují, jak často jsou aktualizována, kdo je spravuje a řadu dalších informací o datech. Oficiální státní katalog otevřených dat veřejné správy ČR by mohl být spuštěn již od začátku roku 2015, ale jeho spuštění bude pravděpodobně navázáno na chystané legislativní změny, což může termín spuštění odsunout. Prozatím je k dispozici například experimentální katalog na adrese <http://cz.ckan.net/>

Plánovaný státní katalog open data by měl být provozován na portálu veřejné správy gov.cz a vkládání nových záznamů a úprava existujících by měla být prováděna skrze datové schránky (obdobně jako je tomu u popisů životních situací). Za každý záznam by měl být zodpovědný ten úřad, který jej založil a ten bude moci záznam v katalogu upravovat, přičemž díky zaslání aktualizace prostřednictvím datových schránek potvrdí, že je skutečně oprávněn záznam upravovat.

### 3.2.8 | Aktualizace dat

Provoz open data v zásadě vyžaduje jen udržovat data stále dostupná na webových stránkách a pravidelně je aktualizovat. Způsob aktualizace dat se odvíjí od způsobu zvoleného v bodě 2a) odstavec 3. Dle nastavených procedur se jen provede export dat za zvolené období a data se zveřejní na webových stránkách. Adekvátně aktualizaci je třeba také upravit záznam o datech v katalogu dat. Zejména při aktualizaci dat přírůstkovým způsobem (nové období je v samostatném souboru) je třeba do katalogu přidat novou adresu, na které se soubor nachází.

### 3.2.9 | Úpravy dat



Ze strany občanů a uživatelů dat se mohou vracet připomínky k datům, které jsou velmi cenným zdrojem podnětů pro opravy a zlepšování dat. Připomínky mohou mířit na chyby v datech nebo jejich formátu, což umožní opravit zdrojová data nebo upravit způsob exportu dat.

Scházet se mohou také žádosti o rozšíření dat o další údaje nebo rovnou o další datové soubory. Je na místě analyzovat, zda úřad tato data má k dispozici, jestli existuje objektivní důvod, proč je nelze zveřejnit a pokud překážky nejsou, tak pro nový dataset provést kroky 1 a 2.

V případě jakýchkoli úprav existujících datasetů je nezbytné informovat o změnách, protože úpravy mohou způsobit problémy uživatelům dat, kteří na nich mají nastaveny automatické skripty, které je zpracovávají.

### 3.2.10 | Zvyšování úrovně datasetů



Zvyšování úrovně datových souborů do formátu s vyšším počtem hvězd může být náročným úkolem.

Přechod ze dvou na tři hvězdy lze provést snadno uložením tabulky do formátu CSV.

Přechod ze tří hvězd (CSV) na čtyři (RDF) již vyžaduje vymyslet a použít URL adresy nejlépe v podobě standardů a využití specializovaných nástrojů na převod do formátu RDF. Aby formát respektoval uznávané standardy, je vhodné spolupracovat na převodu dat do RDF s odborníky na open data z akademické (např. MFF UK a FIS VŠE - <http://opendata.cz>), neziskové (např. Fond Otakara Motejla - <http://otevrenadata.cz>) nebo soukromé sféry. Pro usnadnění převodů dat je například v rámci evropského projektu COMSODE vytvářen nástroj Open Data Node. Mimo to, ale již nyní existuje na trhu řada komerčních i bezplatných nástrojů pro převod dat do formátu RDF.

Přechod ze čtyř na pět hvězd pak vyžaduje provoz vlastního web serveru, na kterém bude každý záznam dostupný na vlastní URL adrese. Zde při použití dalších nástrojů může být rozlišováno, zda k datům přistupuje aplikace nebo uživatel a podle toho vrátí buď uživatelský přívětivější podobu dat nebo data v čistě strojovém formátu. Provozování propojitelných dat také vyžaduje využívání specializovaných nástrojů a je vhodné spolupracovat na přechodu s odborníky.

## 4 | Doporučená literatura



- Metodika publikace otevřených dat veřejné správy ČR - [http://www.korupce.cz/assets/partnerstvi-pro-otevrene-vladnuti/otevrena-data/Methodika\\_Publ\\_OpenData\\_verze\\_1\\_0.pdf](http://www.korupce.cz/assets/partnerstvi-pro-otevrene-vladnuti/otevrena-data/Methodika_Publ_OpenData_verze_1_0.pdf)
- Fond Otakara Motejla - <http://www.otevrenadata.cz/> a <http://www.nasstat.cz/>
- OTEVŘENÁ DATA, PŘÍLEŽITOST PRO ČESKOU REPUBLIKU - <http://www.nasstat.cz/res/data/003406.pdf>

## 5 | Souhrn



Open data jsou data zveřejněná na internetu ve strojově zpracovatelném formátu, pod otevřenou licencí. Aby byla data dobře dohledatelná, měla by být zaznamenána v jednotném katalogu, který je pro ČR připravován.

Open data mohou mít různou úroveň, ale pro začátek se doporučuje zveřejnit data ve formátu CSV, který odpovídá třem hvězdám z pěti. Zveřejnění dat v takové podobě lze zvládnout velmi rychle a snadno za zhruba 20-30 hodin práce. Ekonomické přínosy zveřejnění dat se dostaví se zpožděním až v době, kdy začnou být data využívána, ale pak přínosy mnohanásobně převýší náklady na zveřejnění a údržbu dat.

Cílem zveřejnění open dat je umožnit jejich opakované použití například pro datovou žurnalistiku nebo vznik webových a mobilních aplikací, které již poskytují nové služby občanům a zvyšují transparentnost fungování veřejné správy. Vznikající aplikace nestojí veřejnou správu zcela nic a vzájemně si konkurují, což vede ke zlepšování služeb poskytovaných občanům.

Pro zveřejnění open dat stačí:

1. vybrat data vhodná pro zveřejnění,
2. která nejsou chráněna zvláštním zákonem,
3. uvést u dat, že jsou poskytovány pro další užití včetně komerčního,
4. poskytnout data ve vhodném formátu (pro začátek nejlépe CSV),
5. připravit si postupy a skripty pro export dat,
6. zveřejnit srozumitelný popis dat,

7. zaznamenat data do katalogu dat, aby je bylo možné snáze najít,
8. pravidelně data aktualizovat,
9. déle je zlepšovat a rozšiřovat a
10. časem zvážit i zvyšování úrovně poskytování dat.

Momentálně v ČR působí několik subjektů, které nabízí bezplatnou pomoc se zveřejňováním open dat, k nimž se řadí:

- Fond Otakara Motejla,
- Iniciativa opendata.cz sdružující pracovníky vysokých škol
- a další.

Přestože česká legislativa v roce 2014 open data přímo nepodporuje, zároveň jim ani nebrání a v připravované implementaci evropské směrnice možná pro některé informace formu open data dokonce nařídí.



# TEST

## 1.1 | Co musí umožnit otevřená licence?

- Povolit komerční i nekomerční použití
- Citovat jen se souhlasem úřadu, který data zveřejnil
- Povolit stažení dat až po podpisu memoranda s Ministerstvem vnitra

## 1.2 | Co může být považováno za osobní údaj?

- Cokoli co identifikuje konkrétní právnickou osobu (společnost)
- Cokoli co identifikuje konkrétní orgán veřejné moci
- Cokoli co identifikuje konkrétní fyzickou osobu

## 1.3 | Co není výsledkem open dat?

- Datová žurnalistika
- Mobilní a webové aplikace
- Zrušení tiskového oddělení

## 1.4 | Co open data nemohou přinést?

- Příjem do rozpočtu z licencí na open data
- Vyšší informovanost občanů
- Podporu národní ekonomiky

## 1.5 | Data na úrovni dvou hvězd bývají zpravidla ve formě:

- Textového editoru
- Tabulky MS Excel
- Čistého textového formátu CSV

## 1.6 | Jaká data mohou být zveřejněna jako open data?

- Vše o co požádají novináři
- Vše co můžeme poskytnout v odpovědi na žádost dle zákona 106/1999 Sb., o svobodném přístupu k informacím
- Vše co máme na úřadě v tabulkových souborech

## 1.7 | Jaká data není možné zveřejnit bez anonymizace?

- Souhrnný počet obyvatel jednotlivých územních celků
- Hrubé měsíční příjmy jednotlivých domácností s adresami
- Data o příjmech a výdajích rozpočtu úřadu

## 1.8 | Jaký formát dat je z pohledu open dat nejlepší?

- CSV – data oddělená středníky
- PDF – naskenované tabulky
- RDF – data v jazyce XML s URL adresou pro každý záznam

## 1.9 | Jaký krok zveřejnění open dat musí být první?

- Zápis do národního katalogu open dat
- Analýza jaká data zveřejnit
- Popis dat

### 1.10 | Jaký nástroj pomůže zájemcům o data k jejich nalezení?

- Katalog open dat
- Anonymizace dat
- Napsání licence na míru

### 1.11 | Jaký zákon musí být vždy zohledněn při analýze, zda data zveřejnit?

- Zákon 183/2006 Sb., o územním plánování a stavebním řádu
- Zákon 634/1992 Sb., o ochraně spotřebitele
- Zákon 101/2000 Sb., o ochraně osobních údajů

### 1.12 | Ve kterém roce přistoupila vláda k Partnerství pro otevřené vládnutí?

- 2009
- 2011
- 2013

### 1.13 | Za propojitelná lze open data označit v okamžiku, kdy:

- jsou zcela nezávislá na výrobci konkrétního programového vybavení
- jsou veškerá data zveřejněna pod otevřenou licenci, ale současně jsou i ve strojově zpracovatelném formátu
- má každý záznam v datech obsažený svůj jednoznačný identifikátor v podobě URL adresy