

Metodika
publikace otevřených dat
veřejné správy ČR

verze 1.0

Praha, listopad 2012

Zpracovali:

Dušan Chlapek	Fakulta informatiky a statistiky, Vysoká škola ekonomická v Praze
Jan Kučera	Fakulta informatiky a statistiky, Vysoká škola ekonomická v Praze
Martin Nečaský	Matematicko-fyzikální fakulta, Univerzita Karlova v Praze

Obsah

OBSAH.....	3
SHRNUTÍ METODIKY.....	4
1 ÚVOD.....	6
2 VYMEZENÍ ZÁKLADNÍCH POJMŮ.....	6
2.1 OTEVŘENÁ DATA.....	6
2.2 DATOVÝ KATALOG.....	7
2.3 ROLE PŘI PUBLIKACI A KATALOGIZACI OTEVŘENÝCH DAT.....	7
2.3.1 Poskytovatel dat.....	7
2.3.2 Kurátor dat.....	8
2.3.3 Správce Datového katalogu.....	8
2.3.4 Provozovatel Datového katalogu.....	8
2.3.5 Redaktor.....	9
2.3.6 Koncový uživatel.....	9
3 PROCES PUBLIKACE OTEVŘENÝCH DAT.....	10
4 AKTIVITY PROCESU PUBLIKACE OTEVŘENÝCH DAT.....	11
4.1 ANALÝZA A VÝBĚR DAT K UVEŘEJNĚNÍ.....	11
4.1.1 Analýza dat.....	11
4.1.2 Výběr dat.....	11
4.1.3 Volba licence.....	12
4.1.4 Jmenování kurátora dat.....	12
4.2 VÝBĚR VHODNÉHO FORMÁTU ZVEŘEJŇOVANÝCH DAT.....	13
4.3 NÁVRH ZPŮSOBU PŘÍSTUPU K DATŮM.....	13
4.4 EXPORT DAT DO NAVRŽENÉHO FORMÁTU.....	13
4.5 PUBLIKACE DAT.....	13
4.6 KATALOGIZACE DAT.....	14
4.7 DALŠÍ ZPŮSOBY, JAK DÁT O DATECH NA VĚDOMÍ.....	14
5 DATOVÉ FORMÁTY.....	14
5.1 STROJOVÁ ČITELNOST DAT.....	14
5.2 KDE HLEDAT DATOVÉ STANDARDY.....	16
5.3 DOPORUČENÍ PRO VOLBU FORMÁTŮ.....	17
6 SLOVNÍK POJMŮ.....	18
7 POUŽITÉ ZKRATKY.....	22
8 ZDROJE.....	23
9 ODKAZOVANÉ PŘEDPISY.....	24
9.1 ČESKÉ PRÁVNÍ PŘEDPISY.....	24
9.2 EVROPSKÉ PŘEDPISY.....	24
PŘÍLOHA A: OTEVŘENÁ PROPOJITELNÁ DATA.....	25
FORMÁT RDF.....	25

Shrnutí metodiky

Metodika uvedená v tomto dokumentu představuje ucelenou sadu metodických doporučení pro publikování otevřených dat veřejné správy. Otevřená data lze obecně charakterizovat jako data zveřejněná na internetu způsobem, který neomezuje žádné uživatele ve způsobu jejich použití (technicky ani legislativně) a opravňuje všechny uživatele k jejich dalšímu šíření, pokud při tomto využití a šíření bude uveden autor dat a pokud i ostatní uživatelé budou mít stejná oprávnění s dále šířenými daty nakládat (tj. šířením nedojde k omezení těchto práv například tím, že by uživatel dále šířící otevřená data omezil jejich užití pouze na nekomerční účely).

Publikace otevřených dat probíhá v šesti krocích:

- 1 Analýza a výběr dat k uveřejnění – cílem tohoto kroku je analyzovat dostupná data, popsat jejich strukturu a zvolit data, která je možné a vhodné zveřejnit jako otevřená data.
- 2 Výběr vhodného formátu dat – krok je věnován výběru vhodného formátu dat. Preferovány jsou otevřené standardizované (nebo alespoň široce využívané) formáty.
- 3 Návrh způsobu přístupu k datům – rozhodnutí, zda mají být data zpřístupněna v podobě stažitelných souborů nebo pomocí webových služeb.
- 4 Export dat do navrženého formátu – technické zajištěním převodu dat do zvoleného formátu.
- 5 Publikace dat – určení vhodné webové prezentace dat a volba URL, na které budou data dostupná uživatelům.
- 6 Katalogizace dat – tvorba záznamu o zveřejněných otevřených datech v Datovém katalogu. Spuštění Datového katalogu je plánováno na rok 2013.

Doporučení pro publikaci otevřených dat lze shrnout do následujících bodů:

- Otevřená data by měla být zveřejněna v maximálním možném rozsahu a v podobě, v jaké byla původcem jako primární (původní) vytvořena.
- Je třeba určit podmínky užití dat/zvolit licenci dat. Licence by měla klást na využití otevřených dat minimum omezení (nejlépe by měla vyhovovat Definici otevřenosti¹). Licenci je třeba zveřejnit.
- Pro zveřejněná data je třeba jmenovat odpovědnou osobu, tzv. kurátora dat.
- Je vhodné vytvořit samostatnou stránku ve vaší webové prezentaci/portálu určenou pro zveřejňování otevřených dat.
- Na stránku s otevřenými daty je vhodné umístit odkaz z hlavní nabídky/domovské stránky.
- Po vybudování Katalogu otevřených dat veřejné správy ČR bude třeba do něj zanést záznam o zveřejněných otevřených datech. V současné době je vhodné alespoň upozornit na zveřejněná data prostřednictvím sociálních sítí, tiskové zprávy nebo zprávou do mailing listu okfn-cz@lists.okfn.org. Lze využít i kanálu pro odběr novinek na webu (RSS, Atom).

¹ <http://opendefinition.org/okd/>

Samostatná kapitola metodiky je věnována datovým formátům a doporučením pro jejich výběr, protože strojová čitelnost dat a využívání otevřených standardů, které umožňují uživatelům volit nástroje pro zpracování dat dle jejich potřeby, jsou významnými atributy otevřených dat.

Hlavní doporučení pro datové formáty jsou následující:

- Formát by měl být otevřený, tj. jeho specifikace by měla být zájemcům volně dostupná.
- Formát by měl být dobře strojově čitelný, tj. strukturovaný takovým způsobem, že pomocí programové aplikace lze z dat získat žádané (vybrané) údaje.
- Formát by měl být vyhlášeným standardem nebo by měl být všeobecně využíváný.
- Je vhodné volit formáty založené na XML².
- Je třeba zveřejnit dokumentaci formátu dat.
- Při tvorbě XML formátu je vhodné využít vyhlášené datové prvky z ISDP.

² Zájemci mohou využít formáty založené na RDF (Resource Description Framework), viz Příloha A.

1 Úvod

Metodika uvedená v tomto dokumentu představuje soubor doporučených přístupů a postupů pro publikování otevřených dat veřejné správy. Metodika je určena všem orgánům veřejné správy, které mají za úkol, nebo se rozhodly zpřístupnit některá z dat, která spravují, ve formě otevřených dat.

Tato metodika navazuje na Koncepti katalogizace otevřených dat VS ČR (dále jen Koncepte) a podrobně rozpracovává proces publikace otevřených dat od identifikace vhodných dat ke zveřejnění až po informování veřejnosti o dostupnosti zveřejněných dat.

Metodika je členěna následujícím způsobem. V první části jsou nejprve definovány základní pojmy a celý proces publikace otevřených dat je celkově představen. Druhá část se podrobně věnuje jednotlivým krokům procesu publikace otevřených dat. V třetí části jsou uvedena doporučení ohledně vhodných formátů dat. V závěru jsou uvedeny použité zdroje a slovník pojmů.

2 Vymezení základních pojmů

Před tím, než bude popsán proces publikace otevřených dat a doporučení pro jeho provádění, je třeba vymezit základní pojmy, které budou v dále v metodice používány.

2.1 Otevřená data

Obecně lze otevřená data v souladu s definicí neziskové organizace Open Knowledge Foundation [13] definovat jako data zveřejněná na internetu způsobem, který neomezuje žádné uživatele ve způsobu jejich použití (technicky ani legislativně) a opravňuje všechny uživatele k jejich dalšímu šíření, pokud při tomto využití a šíření bude uveden autor dat a pokud i ostatní uživatelé budou mít stejná oprávnění s dále šířenými daty nakládat (tj. šířením nedojde k omezení těchto práv například tím, že by uživatel dále šířící otevřená data omezil jejich užití pouze na nekomerční účely).

V Koncepti jsou pak na základě 10 principů formulovaných Sunlight Foundation [19] podrobně vymezeny podmínky, které musí data veřejné správy splňovat, aby je bylo možné považovat za otevřená. Data veřejné správy jsou otevřená, pokud jsou:

- 1) **úplná** - data jsou zveřejněna v maximálním možném rozsahu. Rozsah může být definován právním předpisem, usnesením vlády, příp. poskytovatelem dat. Například seznam všech nemovitostí s číslem popisným nebo evidenčním v obci XY, nebo seznam všech památkově chráněných objektů v obci XY.
- 2) **snadno dostupná** - data jsou dostupná na Internetu a dohledatelná běžnými ICT nástroji a prostředky,
- 3) **strojově čitelná** - data jsou ve formátu, který je strukturovaný takovým způsobem, že pomocí programové aplikace z nich lze získat žádané (vybrané) údaje,
- 4) **používající standardy s volně dostupnou specifikací (otevřené standardy)** - data musí být ve formátu, který je volně (bezplatně) dostupný pro libovolné použití nebo do takového formátu převoditelný volně (bezplatně) dostupnou aplikací,
- 5) **zpřístupněna za jasně definovaných podmínek užití dat (licence) s minimem omezení** - podmínky musí být jasně a zřetelně definovány a zveřejněny,
- 6) **dostupná uživatelům při vynaložení minima možných nákladů na jejich získání** - poskytovatelé jsou v souvislosti s poskytováním dat oprávněni žádat úhradu maximálně ve výši, která nesmí přesáhnout náklady spojené s jejich zpřístupněním uživateli; poskytovatel

dat může jednorázově vyžádat i úhradu za mimořádně náročné pořízení dat, pokud si uživatel zpřístupnění těchto dat vyžádá.

Dále je vhodné (nikoliv však nutné), aby otevřená data byla:

- 7) **primární** (původní) - data, která jsou zveřejněna původcem dat v podobě, v jaké byla původcem jako primární (původní) vytvořena. Za primární data se považují i
 - a) referenční údaje ze základních registrů,
 - b) data z registrů a rejstříků VS,
 - c) agregovaná data (např. výsledky voleb) pokud není možné zveřejnit data, z nichž byla provedena agregace,
 - d) agregovaná data - (např. statistiky nad jinými otevřenými daty) pokud je uveden způsob agregace a odkaz na zveřejněná primární data, z nichž byla agregace provedena.
- 8) **zveřejněná bez zbytečného odkladu** - zveřejnění dat není zdrženo činnostmi, které nesouvisí s jejich přípravou; činnosti nezbytné pro publikaci dat jsou provedeny v čase, který umožní jejich zveřejnění bez nepřiměřeně dlouhé prodlevy od okamžiku vzniku dat,
- 9) **neomezující přístup** - data dostupná způsobem, který nediskriminuje jednotlivce nebo skupinu osob,
- 10) **stále dostupná** - data jsou dostupná on-line po dobu uvedenou jejich poskytovatelem.

Pojmem otevřená data se v této metodice rozumí data veřejné správy, které vyhovují alespoň bodům 1-6 uvedeným výše.

2.2 Datový katalog

Katalog otevřených dat veřejné správy ČR (dále jen Datový katalog) je jednotným místem pro vyhledávání otevřených dat veřejné správy a pro získávání údajů o těchto datech (tj. získávání metadat). Datový katalog neslouží primárně k ukládání samotných otevřených dat, ale k ukládání záznamů o těchto datech. Na základě těchto záznamů jsou uživatelé Datového katalogu schopni vyhledat data, která potřebují z jednoho místa bez nutnosti prohledávat izolované webové portály jednotlivých orgánů veřejné správy. Záznamy Datového katalogu také obsahují popis dat a tím usnadňují uživatelům práci s těmito daty.

2.3 Role při publikaci a katalogizaci otevřených dat

Na publikaci a katalogizaci otevřených dat se podílí několik rolí. Jednotlivé role a jejich zodpovědnosti jsou uvedeny níže. Pro úplnost, je zde popsána i role koncového uživatele, který bude využívat otevřená data a k jejich vyhledávání bude využívat Datový katalog.

2.3.1 Poskytovatel dat

Poskytovatelem dat je pro oblast otevřených dat ve veřejné správě ČR jakýkoliv orgán veřejné správy, který má v rámci svých kompetencí právo či povinnost data zveřejňovat, a který rozhodne o zveřejnění dat ve formě otevřených dat. Pokud data pocházejí z některého z informačních systémů veřejné správy, je poskytovatelem těchto dat správce příslušného ISVS.

Poskytovatel dat bude zodpovídat za:

- zveřejnění dat ve formě otevřených dat,
- identifikaci potenciálních přínosů a rizik plynoucích z poskytování dat formou otevřených dat,

- určování podmínek, za jakých jsou data zveřejněna ve formě otevřených dat, a to zejména právních (licence)³, ekonomických, časových a technických,
- pověření kurátora správou dat.

2.3.2 Kurátor dat

Kurátorem je osoba, která zajišťuje zveřejnění otevřených dat a údržbu záznamů o datech v Datovém katalogu.

Kurátor bude zodpovídat za:

- zveřejnění otevřených dat,
- vytvoření a aktualizaci záznamu o poskytnutých otevřených datech v Datovém katalogu, včetně zařazení záznamů do příslušných klasifikačních struktur a vazby na ISDP a IS o ISVS,
- zneplatnění záznamů v Datovém katalogu.

2.3.3 Správce Datového katalogu

Správce Datového katalogu je subjekt odpovědný za Datový katalog. Subjekt plnící roli správce Datového katalogu může plnit i roli provozovatele Datového katalogu.

Předpokládá se, že roli správce Datového katalogu bude vykonávat MV ČR.

Správce Datového katalogu bude zodpovídat za:

- zajištění výběru nástroje,
- zajištění implementace vybraného nástroje,
- definici parametrů pro provoz Datového katalogu,
- sběr podnětů na zlepšení Datového katalogu a zadávání a zajištění implementace změnových požadavků,
- zajištění provozu Datového katalogu (vlastními silami nebo pověřením jiného subjektu rolí provozovatele Datového katalogu),
- organizaci práce s Datovým katalogem,
- určování redaktorů,
- zajišťování přístupových práv.

2.3.4 Provozovatel Datového katalogu

Provozovatel Datového katalogu provádí činnost spočívající zejména v zajištění běžného chodu a údržbě Datového katalogu a potřebného softwarového a hardwarového vybavení. Provozovatel odpovídá za zajištění bezpečnosti Datového katalogu.

Provozovatel Datového katalogu bude zodpovídat za:

- zajištění realizace a testování úprav parametrů a konfigurace prostředí pro provoz Datového katalogu podle pokynů správce Datového katalogu,
- provoz Datového katalogu v rámci dohodnuté úrovně poskytovaných služeb (dostupnost, doba odezvy, ...),
- zálohování a archivaci Datového katalogu,
- zpracování reportů o provozu Datového katalogu.

³ V mezích platné legislativy.

2.3.5 Redaktor

Redaktor je osoba, která provádí obsahovou kontrolu vložených záznamů o otevřených datech VS ČR.

Redaktor zodpovídá za:

- ověření korektnosti a klasifikace záznamu v Datovém katalogu, tj. za ověření:
 - vyplnění povinných atributů,
 - existenci odkazu na zdroj dat,
 - shodu popisu a obsahu dat,
 - zařazení datové množiny do klasifikačních struktur (např. EUROVOC),
 - provázání se záznamy v systémech ISDP a IS o ISVS,
- komunikaci s poskytovateli dat ohledně kvality záznamu v Datovém katalogu,
- označení záznamu ke zveřejnění.

Předpokládá se, že obsazování role redaktora bude spadat do působnosti MV ČR.

2.3.6 Koncový uživatel

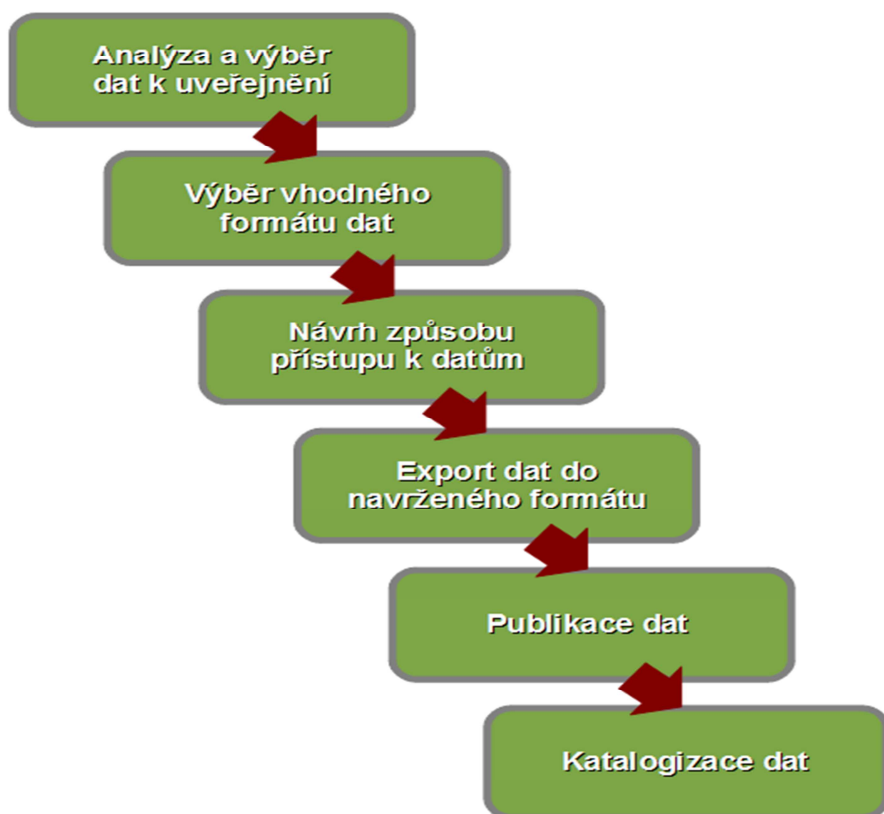
Koncový uživatel je osoba, která:

- vyhledává záznamy o otevřených datech VS ČR v Datovém katalogu,
- posílá správci Datového katalogu podněty na vytvoření nových záznamů, resp. úpravu stávajících záznamů v Datovém katalogu,
- využívá data v souladu s podmínkami jejich užití.

V roli koncového uživatele může obecně vystupovat kdokoli. Registrace není nutná a to ani pro využívání Datového katalogu.

3 Proces publikace otevřených dat

Proces publikace otevřených dat se skládá z několika kroků, které jsou schematicky zachyceny na následujícím obrázku.



Obrázek 1: Proces publikace otevřených dat

Prvním krokem je **analýza a výběr dat k uveřejnění**. Cílem tohoto kroku je analyzovat dostupná data, popsat jejich strukturu a zvolit data, která je možné a vhodné zveřejnit jako otevřená data.

Následující krok je věnován **výběru vhodného formátu dat**. Tento krok je zaměřen na výběr vhodného formátu dat z formátů, které jsou standardizované a obecně využívané. Využívání široce používaných formátů přispívá ke snadnějšímu využití zveřejněných dat.

Dalším krokem je **návrh způsobu přístupu k datům**, jehož hlavní náplní je rozhodnutí, zda mají být data zpřístupněna v podobě stažitelných souborů nebo pomocí webových služeb.

Ve čtvrtém kroku dochází k **exportu dat do navrženého formátu**. Krok je spojen technickým zajištěním tohoto exportu.

V následujícím kroku dochází k **publikaci dat**. V rámci kroku dochází k určení vhodné webové prezentace dat a volbě URL, na které budou data dostupná uživatelům.

Posledním krokem je **katalogizace dat**. Zde dochází k tvorbě záznamu o zveřejněných otevřených datech v Datovém katalogu, aby potenciální zájemci mohli data snadno vyhledat.

Podrobnosti k jednotlivým krokům jsou uvedeny v následující části metodiky.

4 Aktivity procesu publikace otevřených dat

V této kapitole jsou podrobněji popsány jednotlivé kroky publikace otevřených dat.

4.1 Analýza a výběr dat k uveřejnění

První krok publikace otevřených dat je zaměřen na identifikaci vhodných dat ke zveřejnění a jejich přípravu pro další kroky publikace. Aktivity v tomto kroku jsou zaměřeny zejména na body 1, 5 a 7 definice otevřených dat, tj. na úplnost, primárnost a právní podmínky otevřených dat. Hlavními činnostmi v rámci tohoto kroku jsou:

- analýza dat,
- výběr dat,
- volba licence,
- jmenování kurátora dat.

4.1.1 Analýza dat

Je vhodné začít tím, že provedete analýzu, jaká data váš orgán veřejné správy vytváří a případně již dnes publikuje. Dobrymi kandidáty na zveřejnění v podobě otevřených dat jsou data, která je třeba povinně zveřejňovat dle platné legislativy a dále pak data statistického charakteru, data z výkazů a přehledů, data z rejstříků, pokud jsou veřejnosti přístupná, nebo prostorová data (např. údaje o umístění určitých objektů, jako jsou např. lokality národních přírodních památek). Při určování, jaká data primárně poskytnout v podobě otevřených dat můžete v případě již dnes zveřejňovaných dat využít např. statistiku návštěvnosti vašeho webu (tj. zaměřit se na data, na která uživatelé nejčastěji přistupují), případně využít statistiky, o jaké údaje občané žádají v souladu se zákonem č. 106/1999 Sb., o svobodném přístupu k informacím.

Pokud jsou data ke zveřejnění určena právním předpisem či jinak (např. jako závazek v rámci plnění Akčního plánu České republiky „Partnerství pro otevřené vládnutí“ [1]), pak je potřeba analyzovat strukturu a sémantiku zvolených dat. Vhodné je i navrhnout konceptuální model dat, který zároveň může posloužit v kroku 2 i jako dokumentace.

Dále je u zvolených dat třeba provést právní analýzu, jejímž cílem je určit, jak můžete s daty nakládat a zda případně neexistují právní překážky zveřejnění dat.

4.1.2 Výběr dat

Pokud je to možné, měla by být otevřená data zveřejněna v maximálním možném rozsahu a v podobě, v jaké byla původcem⁴ jako primární (původní) vytvořena. Tj. pokud je to možné, data by měla být zveřejněna bez zásahu do jejich obsahu.

Nicméně ne vždy je možné takto postupovat, protože data mohou obsahovat například osobní údaje ve smyslu zákona č. 101/2000 Sb. I přes to mohou být pro potenciální zájemce užitečná anonymizovaná data. Proto definice otevřených dat připouští, aby za primární byla považována i data, která jsou anonymizovaná či jinak agregovaná, pokud data ve skutečně primární podobě zveřejnit nelze.

Pokud tedy zveřejnění primárních dat brání překážka, kterou lze odstranit tím, že z dat bude zveřejněna vybraná podmnožina nebo tím, že data budou zveřejněna v agregované podobě (např. nikoli jednotlivé záznamy šetření, ale jejich součty za okresy/kraje), pak je vhodné tento výběr/agregaci dat provést a popsat.

⁴ Za původce dat je považován orgán veřejné správy, který údaje vytvořil.

4.1.3 Volba licence

Jedním z atributů otevřených dat je jejich zpřístupnění za jasně definovaných podmínek. Proto je třeba tyto podmínky určit, respektive zvolit vhodnou licenci pro data, která hodláte zveřejnit. Kromě samotných dat je pak třeba zveřejnit i tyto podmínky/licenci, aby se všichni uživatelé dat s nimi mohli seznámit.

Dle bodu 5 definice otevřených dat by podmínky jejich užití měly klást na využití otevřených dat minimum omezení. Cílem je umožnit, aby uživatelé data mohli využívat k libovolným účelům, ať už komerčním či nekomerčním, aby mohli data spojovat a dávat do souvislostí s jinými daty a aby je mohli dle potřeby transformovat nebo zahrnout do svých databází či aplikací.

V souvislosti s licencováním otevřených dat se často uvádí tzv. Definice otevřenosti (Open Definition), která blíže definuje, jak chápat otevřenost dat (ale i např. uměleckých děl apod.) [12]. Lze tak doporučit, aby podmínky užití dat, respektive licence, byly voleny tak, aby byly v souladu s Definicí otevřenosti. Na webu věnovanému Definicí otevřenosti lze najít i příklady mezinárodně používaných licencí, které jsou s touto definicí v souladu⁵.

Jednou z těchto licencí je i licence Open Data Commons Public Domain Dedication and License⁶. Dle stanoviska Právnické fakulty Masarykovy univerzity vypracovaného pro Národní technickou knihovnu je tato licence v souladu s českým právním řádem [15]. Uvedená licence je tak jednou z možností, kterou můžete zvážit při volbě licence vašich dat.

Ve světě i v České republice je pro licencování obsahu využívána standardizovaná sada licencí Creative Commons⁷. V současné době je připravována aktualizace znění těchto licencí tak, aby bylo postiženo i tzv. zvláštní právo pořizovatele databáze (viz Směrnice 96/9/ES o právní ochraně databází, v českém právním řádu je toto právo upraveno v §§ 88 – 94 zák. č. 121/2000 Sb.) [14]. Do budoucna tak lze uvažovat i o využití těchto licencí, které jsou spojeny i se standardizovanou sadou piktogramů značících práva a povinnosti při užívání licencovaného obsahu (dat). Tyto piktogramy lze využít i v rámci webových stránek, na kterých jsou publikována otevřená data, a díky těmto piktogramům se uživatel může rychle orientovat v podmínkách užití zveřejněných dat.

Při určování podmínek užití dat/licence je ale třeba mít na paměti, že nelze uživatelům dat poskytnout oprávnění ve větším rozsahu, než jaká má k užití dat váš orgán veřejné správy. Toto je zejména relevantní v případě, že chcete zveřejnit data, která získáváte od jiného subjektu. Proto je vždy vhodné provést analýzu právních omezení dispozic s daty a zjistit tak, zda existují případná omezení, která by bránila zveřejnění dat v souladu s Definicí otevřenosti. Pokud tato omezení existují, ale nebrání Vám data pro určitý způsob užití zveřejnit, je třeba je zveřejnit za podmínek respektujících daná omezení. Zároveň je ale třeba uživatele upozornit na omezení, která při užívání dat musí respektovat.

4.1.4 Jmenování kurátora dat

Pro zveřejněná data je třeba jmenovat odpovědnou osobu, tzv. kurátora dat, který bude zodpovídat za jejich zveřejnění, průběžnou aktualizaci a v budoucnu i za vyplnění záznamu o těchto datech v Datovém katalogu.

Je možné, že tuto osobu již jmenovanou máte, pak je vše v pořádku. Pokud tomu tak není, je vhodné kurátorem dat jmenovat osobu, která je s obsahem dat a s procesem jejich pořízení obeznámena, případně za jejich pořízení již v současné době odpovídá. Pokud chcete jako

⁵ <http://opendefinition.org/licenses/>

⁶ <http://www.opendatacommons.org/licenses/pddl/1.0/>

⁷ <http://creativecommons.org/>

otevřená data zpřístupnit data, která již v současné době zveřejňujete, může být kandidátem na jejich kurátora osoba, která zodpovídá za jejich zveřejnění.

4.2 Výběr vhodného formátu zveřejňovaných dat

Při volbě vhodného formátu dat je vhodné postupovat tak, že je primárně využít existující otevřený standardizovaný (či alespoň obecně akceptovaný) XML formát. Zájemci mohou využít i RDF formát (Resource Description Framework), bližší informace o uplatnění tohoto formátu lze najít v příloze A. Otevřený standard je standard, jehož specifikace je volně dostupná. To dává možnost, aby nezávisle na sobě vznikaly aplikace, které umožní data v tomto formátu zpracovat. Díky použití otevřeného standardu tak uživatel není nucen pro práci s daty využívat jednu konkrétní aplikaci určitého výrobce.

Až v případě, že žádný vhodný formát nalezen nebyl, je nutné definovat vlastní. Vždy je nutné se nejprve zamýšlet nad možností kombinace a rozšíření existujících formátů, které danou problematiku pokrývají alespoň částečně. Vhodné součásti vašeho XML formátu byste např. měli vyhledávat také na portálu Informačního systému o datových prvcích⁸.

Volba správného datového formátu je velmi důležitá, a proto je datovým formátům věnována samostatná kapitola.

4.3 Návrh způsobu přístupu k datům

Je nutné rozhodnout, zda:

- 1 budou data zpřístupněna jako jeden celek v jednom či skupině souborů a v definovaných časových intervalech přepisovány (tj. vždy bude možné získat všechna data platná k poslednímu datu exportu),
- 2 budou zpřístupněny přírůstky v definovaných časových intervalech (např. každý měsíc budou zveřejněna nová data a změny),
- 3 bude umožněn online přístup do databáze pomocí vhodného dotazovacího rozhraní ve formě webové služby (tj. dotazem je specifikována požadovaná podmnožina dat, která je vyexportována do navrženého formátu a zaslána uživateli).

První dvě možnosti jsou výkonnostně méně náročné než (3) a jsou plně dostačující. Volbu mezi (1) a (2) je nutné provést na základě charakteru dat. Pokud se mohou měnit již existující data, je vhodnější spíše možnost (1). Pokud se existující data nemění a pouze přibývají přírůstky, potom je vhodnější varianta (2). Variantu (3) je nutné zvolit v případě, kdy se data velmi dynamicky mění a je nutné k nim zajistit aktuální online přístup.

4.4 Export dat do navrženého formátu

Poté, co je navržen vhodný XML (případně RDF) formát, je nutné vytvořit skripty nebo nakonfigurovat nástroje pro transformaci dat, které do něj exportují data z relační databáze. V případě XML je export přímočarý. V případě RDF je ještě nutné definovat a implementovat politiku tvorby HTTP URI pro objekty. Jednou přiřazené HTTP URI daného objektu by se již nemělo během následných exportů nikdy měnit.

4.5 Publikace dat

⁸ <http://www.sluzby-isvs.cz/ISDP>

Je nutné rozhodnout, na jakém URL budou data publikována. Toto URL by mělo být v rámci vaší vlastní domény. Z vašich webových stránek musí vést na zveřejněná data dobře viditelný odkaz (nejlépe z hlavní nabídky). Ideálním řešením je vytvořit samostatnou stránku ve vaší webové prezentaci/portálu určenou pro zveřejňování otevřených dat. Na tuto stránku umístíte odkazy na všechna zveřejňovaná otevřená data.

Na stránce určené pro zveřejňování otevřených dat je také třeba uvést podmínky jejich užití nebo zvolenou licenci. Je také možné postupovat tak, že na stránce pro zveřejňování otevřených dat bude umístěn zřetelný odkaz na plné znění licence či podmínek užití dat.

4.6 Katalogizace dat

Aby byla vaše data co nejlépe dohledatelná, je vhodné je katalogizovat. Katalog otevřených dat veřejné správy ČR je však teprve ve fázi koncepce a v současné době je plánováno jeho vybudování v roce 2013. Pro katalogizaci otevřených dat veřejné správy bude vytvořena samostatná metodika. V současné době je tak po zveřejnění otevřených dat vhodné dát o nich vědět alespoň některým z alternativních postupů uvedených níže.

4.7 Další způsoby, jak dát o datech na vědomí

Doplňující možností je upozornit na otevřená data prostřednictvím kanálu pro odběr novinek (RSS, Atom), pokud takovýto kanál na vašich stránkách/portálu provozujete. Případně můžete takovýto kanál na vašich webových stránkách zřídit. Stejně tak můžete pro upozornění na otevřená data využít i sociální sítě. Další doplňující možností je vydání tiskové zprávy, neboť zájem některých médií o otevřená data je velký, viz např. <http://data.blog.ihned.cz/>. Vhodné je také upozornit na nově otevřená data do mailing listu okfn-cz@lists.okfn.org, který sdružuje komunitu (sdružení, jednotlivci, akademici, byznys) zajímající se o otevřená data⁹.

5 Datové formáty

V této kapitole jsou uvedena doporučení pro volbu formátů dat a jsou zde uvedeny příklady vhodných formátů.

5.1 Strojová čitelnost dat

Jedním z významných atributů otevřených dat je jejich strojová čitelnost, která umožňuje jejich další automatizované zpracování. Následující tabulka hodnotí vhodnost jednotlivých formátů dat s přihlédnutím k různým charakteristikám strojové čitelnosti. Především je hodnocena existence standardizovaného prostředku pro zápis dat v jednotné strukturované podobě (např. možnost zápisu řádků tabulky ve struktuře, která je zpracovatelná v rámci různých SW aplikací). Také je hodnocena nezávislost formátu na aplikaci, tj. univerzálnost použití dat zapsaných v daném formátu na konkrétní aplikaci nebo typu aplikace (např. tabulkový editor). Dále je hodnocena existence standardizovaného prostředku, který umožňuje popsat strukturu pomocí datového schématu a případně i sémantiku (význam) dat ve strojově čitelné podobě. To je důležité pro zajištění správné práce s publikovanými daty. Dále je přihlédnuto k možnosti vytvářet strojově čitelná propojení mezi souvisejícími daty. Propojení umožňují při strojovém zpracování přímo kombinovat související data z různých zdrojů a zvyšovat tak jejich informační hodnotu.

⁹ Před odesláním zprávy do tohoto mailing listu je třeba se do něj registrovat.

Tabulka 1: Vlastnosti formátů dat

Formát	Nezávislost na aplikaci	Zápis ve strukturované podobě	Popis struktury dat	Popis sémantiky dat	Vytváření propojení	Vhodnost použití 1= nejlepší
PDF	Ne	Ne	Ne	Ne	Ne	5
DOC(X), RTF	Ne	Ne	Ne	Ne	Ne	5
TXT	Ano	Ne	Ne	Ne	Ne	5
HTML	Ano	Částečně	Ne	Ne	Ne	4
XLS(X)	Ne	Částečně	Ne	Ne	Ne	4
CSV	Ano	Ano	Částečně	Ne	Ne	3
JSON	Ano	Ano	Částečně	Ne	Ne	3
XML	Ano	Ano	Ano	Ne	Ne	2
OData	Ano	Ano	Ano	Částečně	Částečně	2
RDF	Ano	Ano	Ano	Ano	Ano	1

Tabulka 2: Vhodnost použití jednotlivých formátů dat

Formát	Komentář	Vhodnost použití 1=nejlepší
PDF	Je vhodný pouze pro publikaci dokumentů určených pro následné zpracování člověkem (tj. čtení). Není vhodný pro následné strojové zpracování, neboť není určen pro jednotný strukturovaný zápis dat. Je závislý na aplikaci (PDF prohlížeč či editor).	5
DOC(X), RTF	Z pohledu zveřejňování dat mají tyto formáty podobné vlastnosti jako formát PDF a nejsou tedy vhodné. Je závislý na aplikaci (prohlížeč či editor tzv. "rich text" dokumentů, tj. dokumentů s formátovaným textem a multimédií).	5
TXT	Má podobné vlastnosti jako předchozí. Není ale závislý na aplikaci a umožňuje jen zápis "čistých" textů bez formátování a multimédií.	5
HTML	Je často používán VS ČR pro zveřejňování dat na různých portálech v podobě čitelné pro člověka. Částečná strojová čitelnost je umožněna jen v případě, kdy jsou využity strukturované prvky formátu pro jednotný strukturovaný zápis dat (např. využití HTML tabulek). Z nich je možné "vytěžit" data pomocí specializovaného software. Často ale jednotná struktura chybí a data tak není možné strojově využít. V případě jednotné struktury ji není možné popsat pomocí standardizovaných prostředků. Je nezávislý na aplikaci.	4

Formát	Komentář	Vhodnost použití 1=nejlepší
XLS(X)	Je taktéž často používaný VS ČR pro zveřejňování dat na různých portálech. Umožňuje další zpracování dat člověkem v tabulkovém prohlížeči či editoru (typu MS Excel). Není ale obecně vhodný pro další strojové zpracování ¹⁰ . I když umožňuje strukturovaný zápis dat, jsou pomocí něj často vytvářeny složité tabulkové struktury, ve kterých jsou navíc využívány různé výpočty a makra. Ty nemají smysl mimo samotný tabulkový editor. Navíc není možné pomocí standardizovaných prostředků popsat strukturu a sémantiku dat a není možné související data propojovat.	4
CSV	Jedná se o standardizovaný formát pro zápis tabulkových dat ve strojově čitelné podobě. Je univerzálně použitelný nezávisle na tabulkovém editoru. Struktura dat je částečně popsána přímo uvnitř CSV souborů ve strojově zpracovatelné podobě, ale nebyl zatím vyvinut standardizovaný jazyk pro popis této struktury. Není možné popisovat sémantiku dat a neexistuje standardizovaný prostředek pro propojování souvisejících dat.	3
JSON	Jedná se o formát používaný především pro výměnu strukturovaných dat mezi serverovými a klientskými částmi webových aplikací. Struktura je popsána uvnitř JSON zdrojových kódů ve strojově zpracovatelné podobě, ale nebyl zatím vyvinut standardizovaný jazyk pro popis této struktury. Není možné popisovat sémantiku dat a neexistuje standardizovaný prostředek pro propojování souvisejících dat.	3
XML	Umožňuje zveřejňování strukturovaných dat (ne jenom tabulkových ale i hierarchických dat) ve strojově čitelné podobě. Strukturu je navíc možné popsat pomocí standardizovaného prostředku (např. jazyka XML Schema). Je často používán datovými API VS ČR. Je nezávislý na aplikaci, ve které je použit. Neumožňuje snadné propojování souvisejících dat a popis sémantiky.	2
OData	Jedná se o protokol (celý název je Open Data Protokol), který umožňuje standardizovaným způsobem webovým klientům číst a upravovat data na webových serverech. Kombinuje jednodušší protokol AtomPub s formátem JSON. Využívá standardního prostředku pro popis struktury vyměňovaných dat a částečně i jejich sémantiky. Umožňuje také propojovat související data, avšak pouze v rámci jednoho datového zdroje.	2
RDF	Umožňuje zápis grafových (síťových) dat. Umožňuje zachytit libovolná strukturovaná data ve strojově čitelné podobě. Nabízí standardizovaný prostředek pro popis struktury i sémantiky dat. Je založen na známých principech webu, s jejichž pomocí umožňuje i plné propojování souvisejících dat z různých zdrojů.	1

5.2 Kde hledat datové standardy

Při hledání vhodného datové standardu pro vaše data je vhodné navštívit webové stránky organizací, které se orientují na vydávání otevřených standardů. K významným organizacím patří např.:

- OASIS¹¹ (Organization for the Advancement of Structured Information Standards),
- W3C Consortium¹².

¹⁰ Tabulkové editory pracující s formáty XLS(X) zpravidla umožňují export dat z tohoto formátu do formátu CSV, který je lépe strojově zpracovatelný. U složitých tabulkových struktur využívajících výpočty nebo dokonce makra nemusí být možné exportovat úplný obsah souboru ve formátu XLS(X).

¹¹ <https://www.oasis-open.org>

Vhodné je také navštívit portál Joinup¹³, na kterém můžete nalézt formáty vytvořené v rámci programu Evropské unie Interoperability Solutions for the European Public Administration (ISA)¹⁴.

Pokud vytváříte XML formát pro vaše data, je vhodné využít Informační systém o datových prvcích (ISDP)¹⁵, na kterém můžete najít vyhlášené datové prvky, jenž můžete využít jako součást vašeho formátu.

5.3 Doporučení pro volbu formátů

Pro zveřejnění otevřených dat je možné zvolit jakýkoliv otevřený strojově čitelný formát splňující podmínky 3 a 4 otevřených dat. Nicméně aby množina používaných formátů byla co nejmenší, doporučujeme volit vhodný XML formát, neboť jazyk XML je dnes všeobecně akceptovaný a běžně veřejnou správou k publikaci dat využíváný. Je také důležité popsat strukturu zvoleného XML formátu a to pomocí jazyka XSD (XML Schema Definition language) včetně detailních komentářů a dokumentace vysvětlujících sémantiku jednotlivých značek tvořících XML formát (tj. XML elementů, příp. atributů).

V současné době se v oblasti otevřených dat začínají uplatňovat i technologie pro propojování dat z různých zdrojů a vznikají tak tzv. otevřená propojitelná data. S těmi je spojeno používání formátů v jazyku RDF (Resource Description Framework). Bližší informace jsou k dispozici v příloze A.

¹² <http://www.w3c.org>

¹³ <https://joinup.ec.europa.eu>

¹⁴ <http://ec.europa.eu/isa/>

¹⁵ <http://www.sluzby-isvs.cz/ISDP>

6 Slovník pojmů

Pojem	Vysvětlení	Zdroj
Application Programming Interface (API)	Specifikované rozhraní pro komunikaci mezi softwarovými komponentami, může obsahovat např. specifikaci procedur, datových struktur, tříd objektů nebo proměnných	[20]
Comma Separated Values (CSV)	Formát pro výměnu dat, kde každý záznam představuje jednu řádku souboru, a jednotlivé položky záznamu jsou odděleny čárkami; soubor může obsahovat záhlaví jako první řádek a jedná se o soubor prostého textu	[18]
číselník	Seznam přípustných hodnot datového prvku obvykle ve formě dvojic, to znamená kódovaného údaje a hodnoty jeho kódu	Zák. č. 365/2000 Sb.
data	Jakékoli fyzicky zaznamenané znalosti (vědomosti), poznatky, zkušenosti nebo výsledky pozorování procesů, projevů, činností a prvků popisovaného světa.	
data veřejné správy	Data, jejichž poskytovatelem je orgán veřejné správy	
datový prvek	Jednotka dat, která je v daném kontextu dále považována za nedělitelnou a je jednoznačně definován.	Zák. č. 365/2000 Sb.
dokument	Obsah na jakémkoli nosiči (psaný či tištěný na papíře či uložený v elektronické formě nebo jako zvuková, vizuální nebo audiovizuální nahrávka); jakákoli část takového obsahu	Směrnice 2003/98/E S
eXtensible Markup Language	Značkovací jazyk pro popis datových objektů označovaných jako XML dokumenty, značky jsou využity pro vyjádření logické struktury dat v XML dokumentu	[5]
Hypertext Markup Language (HTML)	Značkovací jazyk pro publikování obsahu na Webu (World Wide Web, WWW)	[17]
informační systémy veřejné správy (ISVS)	Informační systémy veřejné správy jsou souborem informačních systémů, které slouží pro výkon veřejné správy. Jsou jimi i informační systémy zajišťující činnosti podle zvláštních zákonů.	Zák. č. 365/2000 Sb.
JavaScript Object Notation (JSON)	Jednoduchý formát pro výměnu dat, který je založen na podmnožině jazyka JavaScript	[8]
licence	viz licenční smlouva	

Pojem	Vysvětlení	Zdroj
licenční smlouva	Licenční smlouvou autor poskytuje nabyvateli oprávnění k výkonu práva dílo užít (licenci) k jednotlivým způsobům nebo ke všem způsobům užití, v rozsahu omezeném nebo neomezeném, a nabyvatel se zavazuje, není-li podle § 49 odst. 2 písm. b) sjednáno jinak, poskytnout autorovi odměnu.	Zák. č. 121/2000 Sb.
Open Data Protocol (OData)	Protokol, který umožňuje standardizovaným způsobem webovým klientům číst a upravovat data na webových serverech. Kombinuje jednodušší protokol AtomPub s formátem JSON. Využívá standardního prostředku pro popis struktury vyměňovaných dat a částečně i jejich sémantiky. Umožňuje také propojovat související data, avšak pouze v rámci jednoho datového zdroje.	[10]
orgán veřejné správy (orgán VS)	Ministerstvo, jiný správní úřad nebo územní samosprávný celek	Zák. č. 365/2000 Sb. (podle §3)
otevřená data veřejné správy	Data veřejné správy jsou považována za otevřená, pokud splňují alespoň vlastnosti čísel 1, 4, 5, 7, 8 a 10 z následujících vlastností: 1 úplná - data jsou zveřejněna v maximálním možném rozsahu. Rozsah může být definován právním předpisem, usnesením vlády, příp. poskytovatelem dat. Například seznam všech nemovitostí s číslem popisným nebo evidenčním v obci XY, nebo seznam všech památkově chráněných objektů v obci XY. 2 primární (původní) - data, která jsou zveřejněna původcem dat v podobě, v jaké byla původcem jako primární (původní) vytvořena. Za primární data se považují i - a referenční údaje ze základních registrů, - b data z registrů a rejstříků VS, - c agregovaná data (např. výsledky voleb) pokud není možné zveřejnit data, z nichž byla provedena agregace, - d agregovaná data - (např. statistiky nad jinými otevřenými daty) pokud je uveden způsob agregace a odkaz na zveřejněná primární data, z nichž byla agregace provedena. 3 zveřejněná bez zbytečného odkladu - zveřejnění dat není zdrženo činnostmi, které nesouvisí s jejich přípravou; činnosti nezbytné pro publikaci dat jsou provedeny v čase, který umožní jejich zveřejnění bez nepřiměřeně dlouhé prodlevy od	upraveno dle [19]

Pojem	Vysvětlení	Zdroj
	okamžiku vzniku dat, 4 snadno dostupná - data jsou dostupná a dohledatelná běžnými ICT nástroji a prostředky, 5 strojově čitelná - data ve formátu, který je strukturovaný takovým způsobem, že pomocí programové aplikace lze z dat získat žádané (vybrané) údaje 6 neomezuující přístup - data dostupná způsobem, který nediskriminuje jednotlivce nebo skupinu osob, 7 používající standardy s volně dostupnou specifikací (otevřené standardy) - data musí být ve formátu, který je volně (bezplatně) dostupný pro libovolné použití nebo do takového formátu převoditelný volně (bezplatně) dostupnou aplikací, 8 zpřístupněna za jasně definovaných podmínek užití dat (licence) s minimem omezení - podmínky musí být jasně a zřetelně definovány a zveřejněny, 9 stále dostupná - data jsou dostupná on-line po dobu uvedenou jejich poskytovatelem, 10 dostupná uživatelům při vynaložení minima možných nákladů na jejich získání - poskytovatelé jsou v souvislosti s poskytováním dat oprávněni žádat úhradu maximálně ve výši, která nesmí přesáhnout náklady spojené s jejich zpřístupněním uživateli; poskytovatel dat může jednorázově vyžádat i úhradu za mimořádně náročné pořízení dat, pokud si uživatel zpřístupnění těchto dat vyžádá.	
otevřená propojitelná data	Otevřená data publikovaná podle zásad propojitelných dat (viz propojitelná data)	
otevřený standard	Standard, jehož specifikace je volně dostupná	
Portable Document Format (PDF)	Formát souboru pro reprezentaci obsahu dokumentů nezávislý na aplikačním softwaru, hardwaru a operačním systému; od roku 2008 otevřený standard ISO ISO 32000-1:2008	[21]

Pojem	Vysvětlení	Zdroj
propojitelná data	Data splňující sadu zásad pro publikování a propojování souvisejících dat na webu otevřeným a na standardech založeným způsobem. Konkrétními zásadami jsou: - pojmenování objektů na webu pomocí URI - použití HTTP URI, které umožňují je vyhledat v prostředí dnešního webu - při vyhledání URI jsou uživateli poskytnuta data o objektu reprezentovaná v modelu RDF - objekty jsou provázány pomocí odkazů mezi HTTP URI, takže je možné objevovat související objekty	upraveno dle [2]
Resource Description Framework (RDF)	Model pro reprezentaci a výměnu dat prostřednictvím webu. Má vlastnosti, které umožňují spojování různorodých dat s různými schématy. Jedná se o jednoduchý aplikačně i platformově nezávislý model reprezentující data ve formě grafu.	[9]
Rich Text Format (RTF)	Formát pro výměnu textu a grafiky použitelný na různých výstupních zařízeních a operačních prostředích a systémech. Formát je vytvářen společností Microsoft Corporation a jeho specifikace je volně dostupná	[11]
SPARQL	Dotazovací jazyk a protokol pro RDF data	[6], [16]
správce číselníku	Právní subjekt odpovědný za tvorbu a distribuci číselníku	Zák. č. 365/2000 Sb.
správce datového prvku	Právní subjekt, který nové datové prvky předkládá, navrhuje jejich změnu nebo zrušení	Zák. č. 365/2000 Sb.
strojově čitelná data	Data ve formátu, který je strukturovaný takovým způsobem, že pomocí programové aplikace lze z dat získat žádané (vybrané) údaje.	upraveno dle [7]
subjekt veřejného sektoru	„Subjektem veřejného sektoru“ (se) rozumí státní, regionální nebo místní orgány, veřejnoprávní subjekty a sdružení vytvořená jedním nebo několika takovými orgány nebo jedním nebo několika takovými veřejnoprávními subjekty	Směrnice 2003/98/E S
Uniform Resource Identifier (URI)	Uniform Resource Identifier (URI) je posloupnost znaků identifikující abstraktní nebo fyzický zdroj	[3]
Uniform Resource Locator (URL)	Podmnožina identifikátorů URI, které kromě identifikace zdroje obsahují i určení i způsobu přístupu k tomuto zdroji	[3]

7 Použité zkratky

Zkratka	Význam
API	Application Programming Interface
HTML	Hypertext Markup Language
ISDP	Informační systém o datových prvcích
IS o ISVS	Informační systém o informačních systémech veřejné správy
ISVS	Informační systém veřejné správy
JSON	JavaScript Object Notation
OData	Open Data Protocol
orgán VS	orgán veřejné správy
PDF	Portable Document Format
RDF	Resource Description Framework
RTF	Rich Text Format
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VS ČR	Veřejná správa České republiky
XML	eXtensible Markup Language

8 Zdroje

1. Akční plán České republiky „Partnerství pro otevřené vládnutí“ (2012), http://dl.dropbox.com/u/3675785/OSF/5053_12%20-%20III%20-%20material%20-%20akcni%20plan%20OGP.doc
2. Berners-Lee, T.: Linked Data - Design Issues (2006), <http://www.w3.org/DesignIssues/LinkedData.html>
3. Berners-Lee, T., Fielding, R., Masinter, L.: Uniform Resource Identifier (URI): Generic Syntax (2005), <https://tools.ietf.org/html/rfc3986>
4. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. In: Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (2009)
5. Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., Yergeau, F.: Extensible Markup Language (XML) 1.0 (Fifth Edition) (2008), <http://www.w3.org/TR/2008/REC-xml-20081126/>
6. Clark, K. G., Feigenbaum, L., Torres, E.: SPARQL Protocol for RDF (2008), <http://www.w3.org/TR/rdf-sparql-protocol/>
7. Evropská komise: Návrh Směrnice Evropského parlamentu a Rady, kterou se mění směrnice 2003/98/ES o opakovaném použití informací veřejného sektoru (2011), http://ec.europa.eu/information_society/policy/psi/docs/pdfs/opendata2012/revision_of_PSI_Directive/proposal_directive_CS.pdf
8. Introducing JSON, <http://www.json.org/>
9. McBride, B.: Resource Description Framework (RDF): Concepts and Abstract Syntax (2004), <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>
10. Microsoft Corporation: Open Data Protocol (OData) Specification (2012), <http://www.odata.org/media/30001/%5Bms-odata%5D.pdf>
11. Microsoft Corporation: Rich Text Format (RTF) Specification Version 1.9.1 (2008), <http://download.microsoft.com/download/2/f/5/2f599e18-07ee-4ec5-a1e7-f4e6a9423592/Word2007RTFSpec9.docx>
12. Open Definition, <http://opendefinition.org/okd/>
13. Open Knowledge Foundation: The Open Data Manual (2011), <http://opendatahandbook.org/>
14. Peters, D.: Version 4.0 – License Draft Ready for Public Comment! (2012), <http://creativecommons.org/weblog/entry/32157>
15. Polčák, R., Myška, M., Šavelka, J.: Zpřístupnění bibliografických dat Katalogu NTK a Národní databáze ISSN za podmínek odpovídajících režimu Open Definition (2012), <http://www.techlib.cz/files/download/id/3157/open-bibliographic-data-ntk-studie-2012.pdf>
16. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF (2008), <http://www.w3.org/TR/rdf-sparql-query/>
17. Raggett, D., Le Hors, A., Jacobs, I.: HTML 4.01 Specification (1999), <http://www.w3.org/TR/1999/REC-html401-19991224/>
18. Shafranovich, Y.: Common Format and MIME Type for Comma-Separated Values (CSV) Files (2005), <http://tools.ietf.org/html/rfc4180>
19. Sunlight Foundation: Ten Principles for opening up government information (2010), <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>
20. Wikipedia: Application Programming Interface (2012), http://en.wikipedia.org/wiki/Application_programming_interface
21. Wikipedia: Portable Document Format (2012), http://en.wikipedia.org/wiki/Portable_Document_Format

9 Odkazované předpisy

9.1 České právní předpisy

Zák. č. 101/2000 Sb., o ochraně osobních údajů a o změně některých zákonů

Zák. č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů

Zák. č. 365/2000 Sb., o informačních systémech veřejné správy a o změně některých dalších zákonů

9.2 Evropské předpisy

Směrnice Evropského parlamentu a Rady 2003/98/ES o opakovaném použití informací veřejného sektoru

Příloha A: Otevřená propojitelná data

Pojem otevřených dat je vhodné rozšířit do tzv. otevřených propojitelných dat. Orgány veřejné správy totiž disponují velkým množstvím dat, která již částečně zveřejňují, a díky důsledné aplikaci principů otevřených dat ve veřejné správě bude toto množství v čase narůstat. Mezi daty různých poskytovatelů (a samozřejmě i mezi různými daty toho samého poskytovatele) jsou zajímavé souvislosti a pro použitelnost dat je nutné tyto souvislosti zachytit v podobě propojení. Je známo, že informační hodnota dat roste s množstvím propojení na jiná související data. Propojení jsou z technického hlediska další data a je proto potřeba je zveřejnit opět jako otevřená.

Pro vytváření a zveřejňování propojení je nutné je publikovat v takové podobě, která umožní propojení zaznamenat. Data v této podobě se nazývají propojitelná data. Vytvořením propojení mezi daty a jejich otevřeným zveřejněním potom vzniknou tzv. otevřená propojená data (v angl. se ustálil pojem Linked Open Data). Ta lze charakterizovat následujícími technologickými principy (upraveno dle [2]):

- Konkrétní i abstraktní objekty reálného světa mají přiřazena neměnná URI jako jednoznačné identifikátory. Objekty rozumíme např. konkrétní úřady, města, školy, smlouvy, veřejné zakázky, politické strany, apod.
- Jsou striktně používána pouze HTTP URI tak, aby webové prohlížeče a jiné aplikace mohly k URI přistoupit a získat užitečnou informaci o příslušném objektu. Např. město Semily může mít přiřazeno HTTP URI ve tvaru <http://ld.opendata.cz/resource/business-entity/00276111>.
- Při přístupu k HTTP URI daného objektu jsou poskytnuta data o objektu ve strojově čitelném formátu RDF. To umožní využití dat o objektu v mnoha softwarových aplikacích.
- Data o objektu obsahují i propojení na jiné objekty (opět ve formátu RDF), takže je možné dohledat souvislosti a doplňující údaje k danému objektu. Z HTTP URI města Semily je např. možné se dostat na objekty jako semilské veřejné zakázky, školy či rozpočet, nebo na obecné informace o Semilech zveřejněné na Wikipedii (která je také dostupná v podobě RDF).

Propojená data tvoří nadstavbu současného webu, který lze chápat jako propojené dokumenty. Současný běžný web je tak někdy nazýván *Web of Documents*, zatímco web propojených dat je nazýván *Web of Data* [4]. Unikátnost a neměnnost HTTP URI přiřazeného danému objektu zajistí, že kdokoli může publikovat na svém webu svá tvrzení o objektu. V těchto tvrzeních lze vyhledávat podobně, jako vyhledáváme v současných webových dokumentech, avšak daleko přesněji. Můžeme přitom rozlišovat zdroje tvrzení a filtrovat pouze ty, kterým důvěřujeme. Při nalezení konkrétního objektu se pomocí propojení můžeme navigovat na související objekty podobně, jako procházíme dokumenty na dnešním webu.

Formát RDF

Pro publikování otevřených propojitelných dat (odlišujte od pojmu otevřená data, jak je vysvětleno výše!) je nutné použít formát v jazyku RDF (Resource Description Framework). Ten přímo pracuje s HTTP URI zveřejňovaných objektů a sjednocuje způsob zveřejňování údajů o objektech a souvislostí (tj. propojení) mezi nimi. Datovým modelem RDF je graf, který je přirozenou datovou strukturou pro publikaci a propojování dat na webu. Uzly grafu reprezentují objekty (identifikované svými HTTP URI) a jejich údaje (texty, čísla, datumy, apod). Hrany

přiřazují údaje k objektům a také reprezentují propojení mezi nimi. Graf je potom popsán v podobě trojic

subjekt predikát objekt .

kde subjekt je zdrojový uzel hrany, objekt je cílový uzel hrany a predikát specifikuje typ hrany.

Např.

<<http://www.opendata.cz/necasky#me>> foaf:name "Martin Nečaský" .

<<http://www.opendata.cz/chlapek#me>> foaf:knows <<http://www.opendata.cz/chlapek#me>>.

popisuje graf, jehož uzly jsou dvě osoby identifikované svými HTTP URI <<http://www.opendata.cz/necasky#me>> a <<http://www.opendata.cz/chlapek#me>> a jedna datová hodnota "Martin Nečaský". Trojice specifikují dvě hrany. První hrana spojuje osobu s jejím jménem. Druhá hrana spojuje dvě osoby dohromady a říká, že první osoba zná druhou osobu. Typem hrany může být cokoliv a je nazýván predikát. Možné typy hran (a další koncepty jako např. třídy objektů) jsou definovány v tzv. slovníku nebo také ontologii. Ontologie popisuje nejenom strukturu dat (tj. predikáty a jejich zapouzdření do tříd), ale také sémantiku definovaných predikátů a tříd. Více o formátu RDF a ontologiích lze nalézt ve zdrojích konsorcia W3¹⁶.

¹⁶ <http://www.w3.org/TR/rdf-primer/>, <http://www.w3.org/TR/owl2-primer/>