

Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Bc. Tomáš Kroupa

Softwarová architektura otevřené veřejné správy

Katedra softwarového inženýrství

Vedoucí diplomové práce: Mgr. Martin Nečaský, Ph.D.

Studijní program: Informatika

Studijní obor: I2 Softwarové systémy

Praha 2012

Poděkování

Mé poděkování patří Mgr. Martinu Nečaskému, Ph.D. za cenné rady, proaktivní přístup, inspiraci a především za jeho celkovou angažovanost v dané oblasti. Děkuji také pracovníkům Ministerstva vnitra, zejména pak odboru Hlavního architekta eGovernment za poskytnutí maximální součinnosti a předání cenných zkušeností z praxe. V neposlední řadě děkuji také Petře Karbanové (Pupí) za její velkou trpělivost, podporu a tvorbu potřebného zázemí.

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V Praze dne 25. července 2012

.....

Název práce: Softwarová architektura otevřené veřejné správy

Autor: Bc. Tomáš Kroupa

Katedra (ústav): Katedra softwarového inženýrství

Vedoucí diplomové práce: Mgr. Martin Nečaský, Ph.D., Katedra softwarového inženýrství

Abstrakt:

Veřejná správa disponuje významným množstvím informací, jejíž hodnota je dosud nevyužita. Aplikace principů OpenData a LinkedData umožní tyto informace nejenom efektivně publikovat ale především využít jejich hodnotu. Cílem této práce je analyzovat současný stav, zhodnotit a diskutovat překážky a navrhnout řešení k aplikaci principů OpenData a LinkedData na veřejnou správu ČR.

Klíčová slova:

otevřená data, linked data, otevřená veřejná správa, softwarová architektura

Title: Software architecture of open government

Author: Bc. Tomáš Kroupa

Department: Department of Software Engineering

Supervisor: Mgr. Martin Nečaský, Ph.D., Department of Software Engineering

Abstract:

Public administration owns a large amount of information, whose value is not utilized yet. An application of Open Data and Linked Data principles could enable not only to effectively publish this information, but also to exploit the value. The aim of this thesis is to analyse contemporary situation, assess and debate the barriers and also suggest the solutions for application of the principles in the Public Administration of The Czech Republic.

Keywords:

open data, linked data, open government, software architecture

Obsah

| | | |
|----------|--|-----------|
| 1 | ÚVOD | 1 |
| 1.1 | MOTIVACE | 1 |
| 1.2 | CÍLE PRÁCE | 1 |
| 1.3 | ZPŮSOB DOSAŽENÍ CÍLE | 2 |
| 1.4 | STRUKTURA PRÁCE | 2 |
| 1.5 | INSPIRACE | 2 |
| 1.6 | VÝCHODISKA | 5 |
| 2 | VYMEZENÍ POJMŮ | 11 |
| 2.1 | UŽITÍ POJMŮ | 11 |
| 2.2 | KLASIFIKACE POJMU DATA | 11 |
| 2.3 | VEŘEJNÁ SPRÁVA | 17 |
| 2.4 | ONTOLOGIE | 19 |
| 2.5 | SÉMANTICKÝ WEB | 21 |
| 2.6 | ŽURNALISTIKA ŘÍZENÁ DATY | 26 |
| 2.7 | OSTATNÍ | 26 |
| 3 | POPIS SOUČASNÉHO STAVU ŘEŠENÉHO PROBLÉMU | 32 |
| 3.1 | DATA VE VEŘEJNÉ SPRÁVĚ ČESKÉ REPUBLIKY | 32 |
| 3.2 | SOUČASNÝ STAV PUBLIKOVÁNÍ DAT | 36 |
| 3.3 | OTEVŘENÁ DATA V ČESKÉ REPUBLICĚ | 39 |
| 3.4 | OTEVŘENÁ DATA VE SVĚTĚ | 45 |
| 3.5 | NÁSTROJE PRO KATALOGIZACI OTEVŘENÝCH DAT | 54 |
| 3.6 | EXISTUJÍCÍ ONTOLOGIE A SLOVNÍKY | 56 |
| 3.7 | EXISTUJÍCÍ LICENCE PRO PUBLIKACI DAT | 62 |
| 3.8 | ZHODNOCENÍ SOUČASNÝCH PRAKTIK PUBLIKACE DAT | 64 |
| 4 | POPIS ŘEŠENÍ | 66 |
| 4.1 | RÁMEC ŘEŠENÍ | 66 |
| 4.2 | HARMONOGRAM REALIZACE | 72 |
| 5 | KATALOG OTEVŘENÝCH DAT VEŘEJNÉ SPRÁVY | 74 |
| 5.2 | VARIANTY PROVEDENÍ DATOVÉHO KATALOGU | 74 |
| 5.3 | VARIANTY REPLIKACE DAT | 76 |
| 5.4 | VARIANTY SPRÁVY DATOVÉHO KATALOGU | 77 |
| 5.5 | VARIANTY PLNĚNÍ DATOVÉHO KATALOGU | 79 |
| 5.6 | VARIANTY KONZUMACE OBSAHU PROSTŘEDNICTVÍM KATALOGU | 79 |
| 5.7 | VARIANTY POŘÍZENÍ DATOVÉHO KATALOGU | 80 |
| 5.8 | VYUŽITÍ PORTÁLU | 80 |
| 5.9 | ZHODNOCENÍ A DOPORUČENÍ OPTIMÁLNÍ VARIANTY | 81 |
| 6 | ARCHITEKTURA OTEVŘENÉ VEŘEJNÉ SPRÁVY | 82 |
| 6.1 | ENTERPRISE ARCHITEKTURA | 84 |
| 6.2 | DATOVÝ MODEL OTEVŘENÝCH-PROPOJENÝCH DAT | 89 |
| 7 | PRAKTICKÉ ŘEŠENÍ | 91 |
| 7.1 | MAPA CZECH POINT | 91 |
| 7.2 | PŘÍPADOVÁ STUDIE – DUPLICITA DAT | 92 |
| 7.3 | PŘÍPADOVÁ STUDIE – ŽIVOTNÍ SITUACE | 95 |
| 8 | HODNOCENÍ A ZÁVĚR | 98 |

| | |
|---|------------|
| CITOVANÁ LITERATURA | 99 |
| SEZNAM POUŽITÝCH TABULEK..... | 101 |
| SEZNAM POUŽITÝCH DIAGRAMŮ..... | 102 |
| PŘÍLOHY..... | 104 |
| PRÁVNÍ ÚPRAVA INFORMAČNÍCH SYSTÉMŮ VEŘEJNÉ SPRÁVY ČR..... | 104 |
| SEZNAM NAVRŽENÝCH ZDROJŮ K OTEVŘENÍ..... | 104 |
| APLIKACE VYUŽÍVAJÍCÍ PRINCIPY OTEVŘENÝCH DAT | 105 |
| ČSU – VEŘEJNÁ DATABÁZE – TEMATICKÉ OKRUHY..... | 106 |
| PORTÁLY OTEVŘENÝCH DAT VE SVĚTĚ..... | 107 |
| ORGANIZAČNÍ STRUKTURA MĚSTA A ÚTVARU MINISTERSTVA..... | 108 |

1 Úvod

To reach for new heights and reveal the unknown so that we do and learn will benefit all humankind¹.

The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect.²

1.1 Motivace

Položili jste si někdy otázku, kolik stojí provoz veřejného osvětlení v místě vašeho bydliště, jak vysoká je v něm kriminalita a jaké je srovnání s jinými obvody? Případně kde najdete nejbližší úřad schopný vykonat požadovanou agendu? Technologie dnešní doby umožňují na tyto a mnohé další otázky automatizovaně odpovědět. Nicméně pro relevantní odpověď jsou zapotřebí relevantní data, která jsou mnohdy nedostupná, nebo je jejich zpracování komplikováno.

Veřejná správa disponuje značným množstvím informací, ať už se jedná o informace sbírané od občanů nebo firem přes informace zpracované na základě statistik či analýz, až po informace generované činností veřejné správy. Vždy se jedná o informace „zaplacené“ z daní občanů země.

Jsou-li tyto informace volně přístupné komukoliv, začínáme hovořit o tzv. otevřené veřejné správě³. Otevřená veřejná správa je politickým cílem současné vládní reprezentace a řady nevládních organizací jak tuzemských, tak zahraničních.

Zveřejnění informací je krok správným směrem, avšak aby vedl ke skutečné otevřené veřejné správě, je zapotřebí splnit ještě několik podmínek. Ta nejdůležitější se týká vhodné datové infrastruktury zveřejňování informací a dále její vhodné softwarové architektury. Bez ní je vytěžování obrovského množství informací nepředstavitelné.

Zpřístupnění dat je velmi pozitivní aspekt, skutečná hodnota informací však spočívá nejenom v jejich „existenci“, ale, a to především, v jejich vztahu k okolí. Teprve když jednotlivé „fragmenty“ informací propojíme mezi sebou, využijeme jejich plný potenciál.

Pokud tyto dva základní přístupy – otevřenost a propojenost dokážeme naplnit a zároveň umožníme snadný mechanismus dohledatelnosti, zavedeme nový efektivní systém svobodného přístupu k informacím. Občané od veřejné správy očekávají stále lepší a kvalitnější služby a veřejná správa má možnost jim je dát.

1.2 Cíle práce

Cílem diplomové práce na téma „Softwarová architektura otevřené veřejné správy“ je provést analýzu přístupu k publikaci dat veřejnou správou s důrazem na principy OpenData⁴ a LinkedData⁵. Především se pak jedná o prozkoumání

¹ NASA Open Government Plan version 2.0. <http://open.nasa.gov/plan/introduction>

² Tim Berners-Lee, ředitel konsorcia World Wide Web

³ Zřídka také o otevřené vládě (vládnutí), v cizině Open governancy

⁴ OpenData je obecně publikační model založený na svobodném a volném přístupu k informacím bez technických omezení.

současného stavu veřejné správy ČR, analýzy dosavadních přístupů v této oblasti u významných zemí světa a v neposlední řadě identifikace překážek a návrh vhodné architektury pro publikaci a katalogizaci dat v ČR.

1.3 Způsob dosažení cíle

Při realizaci této diplomové práce budeme postupovat metodou analýzy zahraničních přístupů a studiem elektronických zdrojů, které se problematikou zabývají. Dále podrobně prozkoumáme současný stav a možnosti publikace v oblasti veřejné správy ČR. Poznatky získané metodou analýzy jednotlivých zdrojů podrobíme rozboru a následně z nich vytvořím e tuto práci.

Jelikož sledovaná oblast je velmi rozsáhlá a přístupy jednotlivých subjektů rozdílné, a zvláště v literatuře dosud spíše nezpracované, bude mou snahou ho čtenáři přiblížit a osvětlit, proč je právě efektivní publikace dat veřejné správy tak zásadní a důležitá.

Při citování informačních zdrojů vycházím z mezinárodních norem ISO 690 a ISO 690-2, v textu používám citování podle odkazu do seznamu použité literatury, stanoveného normou ČSN ISO 690 a doporučovaného mezinárodní normou ČSN ISO 7144:1986. Odkazy na zdroje mimo literaturu jsou uvedeny v poznámkách pod čarou.

1.4 Struktura práce

Diplomová práce je rozdělena sedmi hlavními kapitolami. V první úvodní kapitole jsou umíněna východiska a motivace pro otevřená-propojená data ve veřejné správě. Druhá kapitola se zabývá vymezením nejdůležitějších pojmů použitých v této práci. Třetí kapitola popisuje současný stav řešeného problému a to jak z pohledu používaných technologií tak z pohledu praxe v některých zemích. Čtvrtá kapitola řeší současnou situaci v ČR a navrhuje postup zavedení otevřených-propojených dat u nás. Pátá kapitola se zabývá jednou z nejproblematictějších oblastí práce a tou je katalogizace dat a volba vhodného katalogu. Šestá kapitola navrhuje globální architekturu s ohledem na vytýčené cíle. A konečně sedmá kapitola ukazuje použití otevřených-propojených dat na praktických příkladech.

1.5 Inspirace

Nadace Open Society Fund Praha ve své publikaci „Otevřená data ve státní správě: Nová éra rozhodování“ (1) uvádí následující skutečnosti, které podporují myšlenku otevřených dat:

V dubnu 2010 odhalil kanadský konzultant v soukromé firmě daňový únik v hodnotě 3,2 miliardy kanadských dolarů, jeden z největších v historii. Přišel na něj při běžné analýze dokumentů, regulující charitativní sbírky v Torontu⁶.

Policejní oddělení v americkém Houstonu přišlo o 25 milionů USD na pokutách za dopravní přestupky, protože se vyšetřovatel vůbec nedostavil k soudu⁷. Zjistila to organizace Texas Watchdog při zkoumání statistik policejního oddělení.

⁵ LinkedData je obecně publikační model založený na propojení souvisejících dat mezi sebou a to za dodržení předepsaných principů a použití vhodných technologií.

⁶ Zdroj: <http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/>

⁷ Zdroj: <http://www.texaswatchdog.org/2010/11/houston-police-miss-hundreds-of-court-dates-tickets/1290042682.story>

Transparency International Slovensko spočítalo, že 43 % veřejných zakázek v roce 2009 mělo jediného uchazeče⁸. Na každém dalším uchazeči se přitom statisticky⁹ ušetří 5-8 % nákladů.

1.5.1 Otevřená data a veřejná správa

Otevřená data jsou tématem, o kterém se v poslední době hodně mluví. Především proto, že veřejnost začíná data požadovat a řada obcí, nevládních organizací a v neposlední řadě i úřadů začíná svá data poskytovat.¹⁰

Aktivní zpřístupňování informací a dat, která mají státní orgány a veřejné instituce k dispozici pro veřejnost, se stává standardem komunikace státní správy s občany. Na celém světě přibývá vlád, které automaticky zveřejňují data bez snahy vyhodnocovat, zda jsou pro občany užitečná. Zkušenosti z takových míst (USA, Velká Británie, Slovensko, Keňa ad.) potvrzují, že přístupná data podněcují aktivitu a kreativitu firem, neziskových organizací či jednotlivých občanů, kteří vytvářejí aplikace a inovativní služby pro ostatní (2).

Mezi přínosy patří:

- Přínosy pro veřejnou správu
 - o úspora prostředků - veřejná správa získá přehled, kde jsou sbírána či tvořena data a vytvoří si tak strategii pro tvorbu důležitých informačních systémů,
 - o efektivnější práce s daty - systematizuje se sběr a zveřejňování dat, lépe se odhalují zdroje duplicitních dat,
 - o data budou ucelený zdroj pro analýzy a následná kvalifikovaná rozhodnutí.
- Přínosy pro (odbornou) veřejnost, komerční a akademickou sféru
 - o podklady pro svobodnou obchodní, vědeckou a výzkumnou činnost,
 - o efektivnější kontrola fungování veřejné správy,
 - o podpora fenoménu datové žurnalistiky, která data umí interpretovat a zpřístupňovat je tak občanům,
 - o podklady pro tvorbu softwarových aplikací.

1.5.2 Google Translator jako efektivní konzument otevřených dat

Google Translator¹¹ je jednou z významných služeb společnosti Google Inc. Tento nástroj pracuje na bázi statického překladu. Statický překlad můžeme jednoduše popsat jako indexaci významově stejných textů v různých jazycích a jeho následné vyhledávání.

Samotný Google nástroj marketingově popisuje takto¹²:

Překladač Google při překládání vyhledává podobné texty ve stovkách miliónů dokumentů a určuje, který překlad je nejvhodnější. Překladač Google dokáže rozpoznat podobný text v dokumentech, které již byly dříve přeloženy překladateli – lidmi. Pomocí inteligentních algoritmů může odhadnout správný překlad. Proces vyhledávání podobností ve

⁸ Zdroj: <http://www.transparency.sk/o-43-velkych-verejnych-obstaravani-v-roku-2010-sutazil-len-jeden-kandidat>

⁹ Zdrojem těchto dat je Vestník veřejného obstarávání, kde instituce povinně sdílejí vypsané tendry.

¹⁰ Seminář OpenData, konaný dne 28. 2. 2012 v Poslanecké sněmovně parlamentu ČR pořádaný iniciativou OpenData.cz

¹¹ Dostupný na adrese <http://translate.google.com/>

¹² Zdroj: http://translate.google.cz/about/intl/cs_ALL/

velkém množství textu je označován jako statistický strojový překlad. Protože jsou překlady vytvářeny strojově, nemusí být vždy dokonalé. Čím více dokumentů v určitém jazyce přeložených překladateli – lidmi může Překladač Google analyzovat, tím vyšší bude kvalita překladu. Proto se někdy může přesnost překladů v různých jazycích lišit.

Úspěch statického překladu závisí na množství relevantních textů, které jsou zaneseny do databáze.

Podle platného evropského práva¹³ je většina oficiálních dokumentů EU překládána do všech úředních jazyků EU¹⁴. Infrastruktura zveřejňování je navržena pro efektivní práci a umožňuje i strojové zpracování. Důvody, proč Evropská unie používá 23 úředních jazyků, jsou nasnadě: „demokracie, transparentnost a právo na informace“ (3).

Dále celá řada mezinárodních organizací¹⁵ pracuje s více úředními jazyky a své výstupy zveřejňuje. Jedná se např. o:

| Organizace | Úřední jazyky |
|-------------|--|
| OSN | angličtina, francouzština, španělština, ruština, čínština, arabština |
| OECD | angličtina, francouzština |
| WTO | angličtina, francouzština, španělština |
| UNESCO | angličtina, francouzština, španělština, ruština, čínština, arabština |
| OBSE | francouzština, němčina, angličtina, španělština, italština, ruština |
| ZEU | angličtina, francouzština, němčina, nizozemština, španělština, portugalská, italština, řečtina |
| Rada Evropy | angličtina, francouzština |

Tabulka 1. Úřední jazyky vybraných mezinárodních organizací

Jelikož přístup k informacím není omezen, mají pokročilé překladače k dispozici stovky milionů stránek různorodého textu ve více než 20 jazycích¹⁶. Jen do českého jazyka překládá více než 200 překladatelů na plný úvazek a dalších 60 překladatelů externě¹⁷.

Provoz překladatelské a tlumočnické služby EU stojí ročně 1% rozpočtu EU, což přibližně odpovídá 2€ na občana EU (3). Ačkoliv nelze dopad této služby přímo měřit, je její přínos nezanedbatelný. To se opírá i o tvrzení (3):

¹³ Nařízení č. 1., o užívání jazyků v Evropském hospodářském společenství. Úř.věst. L 17, 6.10.1958, s. 385, ve znění pozdějších předpisů.

¹⁴ Doplnující informace dostupné online na: http://ec.europa.eu/languages/languages-of-europe/eu-languages_cs.htm

¹⁵ Zdroj: Webové prezentace jednotlivých organizací.

¹⁶ Stávající dokumenty se překládají podle potřeby, nově vzniklé dokumenty se překládají do všech úředních jazyků. K 1. 5. 2012 se jedná o 23 jazyků.

¹⁷ Zdroj: <http://www.jazyky.com/content/view/732/54/>

Využíváním aplikací dostupných na trhu a vyvíjením vlastních řešení se evropské útvary zajišťující jazykové služby staly ideální laboratoří a celosvětovým referenčním nástrojem. Výsledky jejich úsilí jsou poskytovány veřejnosti a často je využívají výzkumní pracovníci a soukromé společnosti.

1.5.3 Mapa duchů – počátky londýnské kanalizace

John Snow¹⁸, anglický lékař a průkopník zavádění anestezie, provedl v roce 1854 na tehdejší dobu revoluční analýzu, a to srovnání výskytu cholery se statistikou¹⁹ znečištění zdrojů pitné vody v části Londýna Soho²⁰. Výstupem této analýzy bylo zjištění souvislosti mezi nedostatečnou hygienou a výskytem cholery.

Tato studie iniciovala ustanovení Metropolitan Board of Works (MBW)²¹, která doporučila vybudování důkladného systému kanalizační sítě. Nejrozsáhlejší stavební akce 19. století představovala stavbu 2 100 km tunelů a potrubí pod povrchem Londýna, které odváděly splašky a rozváděly pitnou vodu. Poté co byl tento systém dokončen, bylo zaznamenáno výrazné snížení výskytu cholery a jiných nákaz. Tento systém je využíván do současné doby.

Pouze skutečnost, že údaje o výskytu cholery byly veřejné a na tehdejší dobu dostatečně přístupné²², umožnili provedení takto důležité analýzy, která dokázala zachránit mnoho lidských životů.

1.6 Východiska

V současné době existuje velké množství informací na webu, zpravidla ale zůstávají skryty v neveřejných databázích, nebo jsou publikovány takovým způsobem, který není vhodný pro strojové zpracování.

Pojem „otevřená data ve veřejné správě“ je velmi aktuální a zabývá se jím celá řada institucí na evropské i národní úrovni.

1.6.1 Situace v ČR a Open Government Partnership

Zavádění nástrojů elektronické komunikace a propojování databází správních orgánů je součástí dlouhodobých záměrů vlády České republiky v oblasti modernizace veřejné správy²³.

Partnerství pro otevřené vládnutí (Open Government Partnership - OGP) je mezinárodní iniciativou administrativy USA podporující otevřenost, transparentnost a boj proti korupci²⁴. Cílem je konkrétní závazek vlád podpořit rozpočtovou

¹⁸ John Snow (15. března 1813 York – 16. června 1858 Londýn) byl anglický lékař a průkopník zavádění anestezie a hygieny.

¹⁹ Přesněji provedl zkoumání společných atributů v místech se zvýšeným výskytem cholery a těmi byla právě nedostatečná hygiena a infikované zdroje pitné vody.

²⁰ Soho je londýnská čtvrť v obvodu Westminster. Hranice Soho tvoří na severu Oxford Street, na západě Regent Street, na jihu Piccadilly Circus a Leicester Square a na východě Charing Cross Road.

²¹ Metropolitan Board of Works byl hlavní správní institucí Londýna, zahájila svou činnost 22. prosince 1855 a jeho hlavním úkolem bylo zajistit infrastrukturu pro rychle se rozvíjející Londýn.

²² Ačkoliv nemůžeme přímo hovořit o zveřejnění v souladu s principy OpenData. Dle úrovně tehdejší technologie se jedná o velmi uživatelsky přívětivou publikaci.

²³ Zdroj: Věcný záměr IS Cizinecké policie, interní projekt Ministerstva vnitra

²⁴ Zdroj: <http://www.opengovpartnership.org>

transparentnost, zvýšit občanskou angažovanost, bojovat s korupcí a transformovat se ve více otevřené, efektivní a odpovědné instituce²⁵.

O připojení k mezinárodní iniciativě Partnerství pro otevřené vládnutí rozhodla vláda České republiky svým usnesením ze dne 14. září 2011 č. 691 a pověřila místopředsedkyni vlády, předsedkyni Legislativní rady vlády a předsedkyni Vládního výboru pro koordinaci boje s korupcí přípravou akčního plánu k dosažení otevřeného vládnutí a naplňování zásad Open Government Partnership²⁶.

Úvodní teze projektu jsou k dispozici v (4).

Dne 4. 4. 2012 došlo ke schválení akčního plánu²⁷. V plánu se mimo jiné uvádí²⁸:

Třemi hlavními okruhy, které byly na základě veřejné diskuze do Akčního plánu zařazeny, jsou:

- přijetí zákona o úřednících veřejné správy zajišťující odpolitizování, profesionalizaci a stabilizaci veřejné správy,
- zefektivnění systému svobodného přístupu k informacím,
- zpřístupnění dat a informací.

Způsob plnění výše uvedených bodů je ve čtyřech po sobě navazujících krocích:

1. Identifikace a odstranění překážek
 - a. Právní otevřenost, tj. zveřejnění dat pod otevřenou licenci
 - b. Technická otevřenost, tj. zveřejnění dat ve standardním strojově čitelném formátu
2. Vytvoření infrastruktury otevřených dat ČR
3. Otevření nejdůležitějších datových zdrojů
4. Vytvoření katalogu dat ČR

Body 1. a 2. jsou předmětem této práce, vycházejí z akčního plánu a doplňují ho o vlastní přístup k otevřeným datům.

Akční plán dále uvádí seznam doporučených zdrojů pro první fázi otevření, seznam je uveden v příloze.

Harmonogram realizace akčního plánu:

| Aktivita | Termín |
|---|--------------|
| 1. Identifikace a odstranění překážek | 30. 9. 2012 |
| 2. Vytvoření infrastruktury otevřených dat ČR a jejich pravidel v | 31. 12. 2012 |

²⁵ Zdroj: <http://www.opengov.cz>

²⁶ Zdroj: <http://www.vlada.cz/assets/clenove-vlady/pri-uradu-vlady/karolina-peake/tiskove-zpravy/Akcni-plan-OGP.pdf>

²⁷ Zdroj: <http://www.vlada.cz/cz/clenove-vlady/pri-uradu-vlady/karolina-peake/tiskove-zpravy/vlada-dnes-schvalila-akcni-plan-ceske-republiky-partnerstvi-pro-otevrene-vladnuti-94353/>

²⁸ Zdroj: <http://www.vlada.cz/assets/clenove-vlady/pri-uradu-vlady/karolina-peake/tiskove-zpravy/Akcni-plan-OGP.pdf>, kapitola III. Závazky ČR, str. 3.

| | |
|--|--------------|
| oblasti veřejných zakázek | |
| 3. Otevření nejdůležitějších datových zdrojů | 31. 12. 2012 |
| 4. Vytvoření katalogu dat veřejné správy, průběžné plnění a čištění katalogu | 31. 3. 2013 |

Tabulka 2. Harmonogram realizace akčního plánu vlády

Národní ekonomická rada vlády²⁹ (NERV) - poradní orgán Vlády ČR, který pomáhá vládě nalézt nejvhodnější podobu reforem a ekonomických opatření v jednom ze svých doporučení uvádí:

„Ukládání veškerého obsahu vyprodukovaného orgány veřejné moci (vč. výběrových řízení, hlasování orgánů, ...) do digitálních knihoven a vybudování úložiště (nebo sítě úložišť) dokumentů veřejné moci v elektronické podobě.“

Jan Farský, poslanec Parlamentu ČR k otázce otevřených dat uvádí:

„Je nezbytné, aby i veřejné instituce pochopily změny, jež přinesla informační revoluce. Jde o úplně nový rozměr fungování státní správy, která musí poskytovat maximum informací v uživatelské podobě. Kdo drží informace, má moc, a té se stát nerad vzdává. Nyní ale přišla doba, kdy je nezbytné informace poskytnout, vrátit je lidem. Což je nutně spojeno s tím, že se my, politici, vzdáme části moci, což nikdy nebude bezbolestné. Proto nebude jednoduché data otevřít. Přesto se o to musíme zasadit.“³⁰

1.6.2 Východiska vyplývající z rozhodnutí orgánů Evropské unie

Akční plán Evropské unie na období 2011 – 2015 nabádá k přechodu ze současné elektronické správy na novou generaci otevřených, flexibilních a společných služeb, které budou hladce fungovat jak na místní či regionální, tak na vnitrostátní a celoevropské úrovni. Jak uvádí Evropská komise, elektronizace vede na jedné straně k efektivnějšímu využívání zdrojů, na straně druhé ke snižování výdajů. Sdílení veřejných i soukromých zdrojů pak znamená větší efektivitu při vývoji.

V akčním plánu se přímo uvádí

Veřejný sektor je „zlatým dolem“ informací. Mnohé z údajů, které orgány veřejné správy shromažďují, se nepoužívají nebo slouží jen omezeným účelům. Uvolnění neosobních veřejných údajů (geografických, demografických, statistických, údajů o životním prostředí atd.) zejména ve strojově čitelném formátu umožní občanům a podnikům nalézt způsoby jejich využití a vytvářet inovativní produkty a služby.

Myšlenka je formálně ukotvena ve směrnici Evropské komise pro opakované použití informací veřejného sektoru³¹ a rozhodnutím komise o opakovaném použití

²⁹ Zdroj: <http://www.vlada.cz/cz/ppov/ekonomicka-rada/narodni-ekonomicka-rada-vlady-51371/>

³⁰ Zdroj: Tisková zpráva, Seminář OpenData, konaný dne 28. 2. 2012 v Poslanecké sněmovně parlamentu ČR pořádaný iniciativou OpenData.cz. Dostupné online na: http://janfarsky.cz/soubory/otevrena_data_verejna_sprava.pdf

³¹ Směrnice 2003/98/ES o opakovaném použití informací veřejného sektoru, 17. 11. 2003, Úř. věst. L 345, 31. 12. 2003, s. 90 -96.

informací Komise³². Částečně je podpořena v rozhodnutí Komise o opakovaném použití dokumentů Komise³³.

Směrnice o opakovaném použití informací Komise ve svém úvodu zmiňuje:

1. Nové informační a komunikační technologie vytvořily nebývalé možnosti pro slučování a kombinování obsahu z různých zdrojů.
2. Informace veřejného sektoru jsou nevyužitým zdrojem, který může být základem pro nové produkty a služby informační společnosti s přidanou hodnotou. Komise zdůraznila hospodářský potenciál informací veřejného sektoru ve svém sdělení ze dne 23. října 2001³⁴ nazvaném „eEurope 2002: Vytváření rámce EU pro využití informací veřejného sektoru“.
3. Komise a ostatní orgány mají samy v držení mnoho dokumentů různých druhů, které by mohly být opakovaně použity v informačních produktech a službách s přidanou hodnotou, a které by mohly poskytnout užitečný zdroj obsahu pro společnosti i občany.

1.6.3 Využívání elektronických služeb a informačních zdrojů

Limitem pro využívání otevřených-propojených dat je technologická stránka. Pouze společnost, kde je dostatečná technologická gramotnost a technologie, jako je například připojení k internetu apod., může plně využít potenciál otevřených-propojených dat.

Níže uvedené případy mohou být uvedeny jako názorný příklad využití elektronických služeb a informačních zdrojů.

Připojení organizačních složek státu a obcí k síti Internet³⁵:

| | 2005 | 2010 |
|--------------------------|------|------|
| Obecní úřady | 41 % | 88 % |
| Organizační složka státu | 86 % | 97 % |

Tabulka 3. Připojení organizačních složek státu a obcí k síti Internet

V případě obcí nad 5000 obyvatel a krajských orgánů se v roce 2010 jedná již o 100 %.

Využívání služeb na bázi e-banking ve světě³⁶ je znázorněno na dalším diagramu.

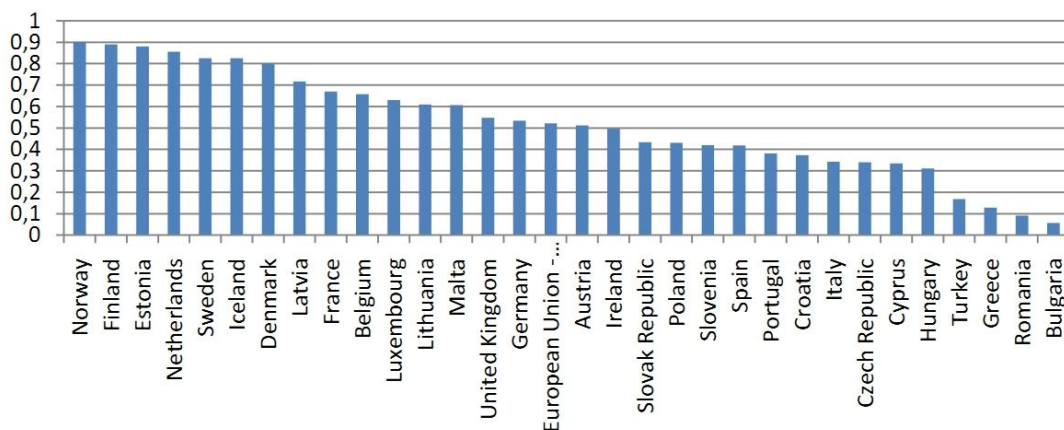
³² Rozhodnutí komise 2006/291/ES o opakovaném použití informací Komise, 7. 4. 2006, Úř. věst. L 107/38, 20. 4. 2006

³³ Rozhodnutí komise 2011/833/EU o opakovaném použití dokumentů Komise, 12. 12. 2011, Úř. věst. L 330, 14. 12. 2011, s. 39-42.

³⁴ KOM(2001) 607 v konečném znění

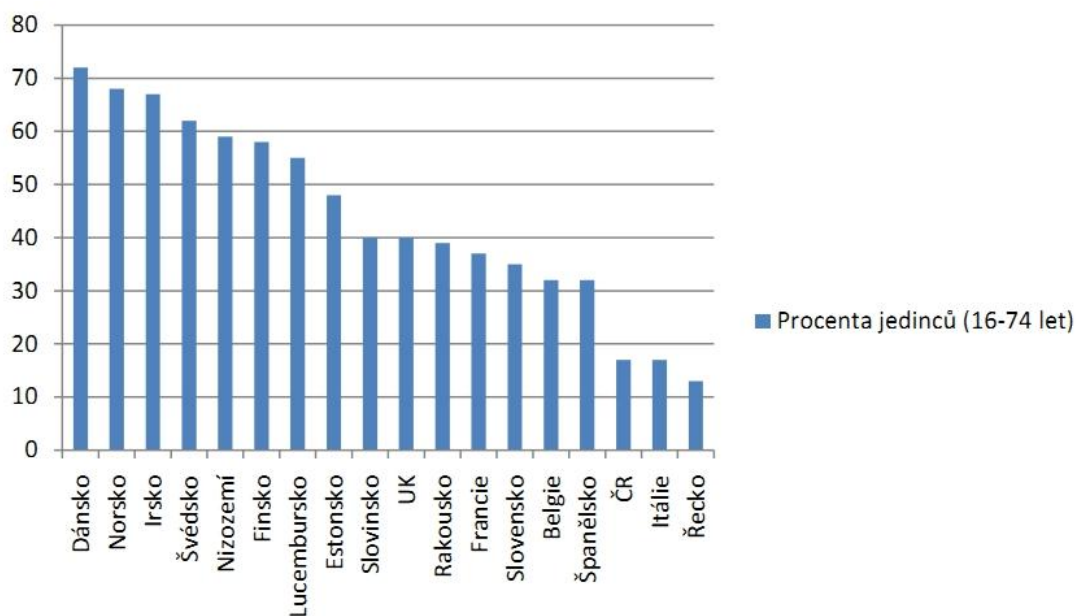
³⁵ Časopis Veřejná správa 10/2012. Dálkově přístupné informační systémy. Doc. Judr. Pavel Matas, CSc.

³⁶ Zdroj: Towards a Trusted and Sustainable European Federated eID system – final report. Evropská komise 2010. Dostupné online na http://ec.europa.eu/information_society/activities/egovernment/studies/completed_studies/completed_studies_2010/index_en.htm



Obrázek 1. Využívání služeb na bázi e-banking ve světě

Následující graf zobrazuje jedince, kteří využívají internet ke komunikaci s veřejnou správou v roce 2010³⁷:

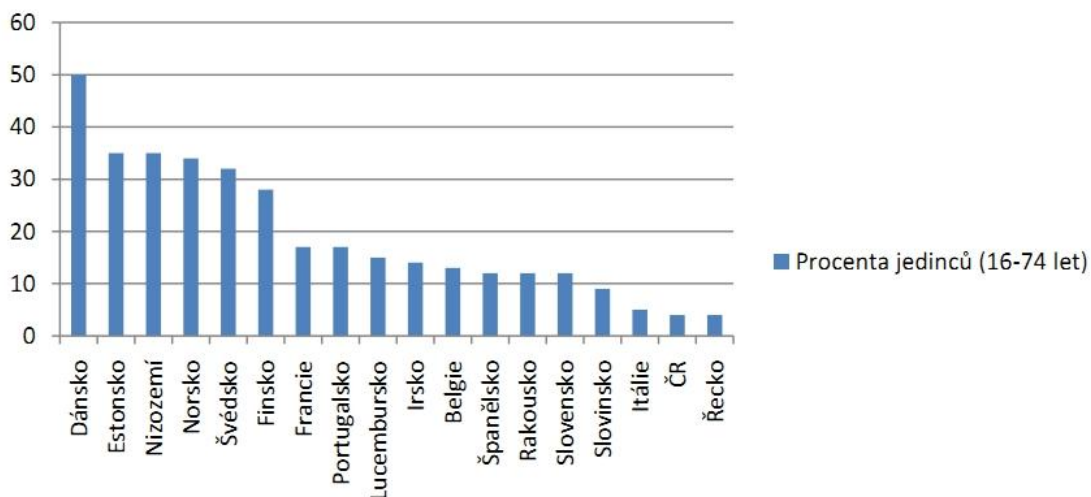


Obrázek 2. Jedinci, kteří využívají internet ke komunikaci s veřejnou správou v roce 2010

Následující graf zobrazuje jedince, kteří používají internetu k podání vyplněných formulářů veřejné správy v roce 2010³⁸.

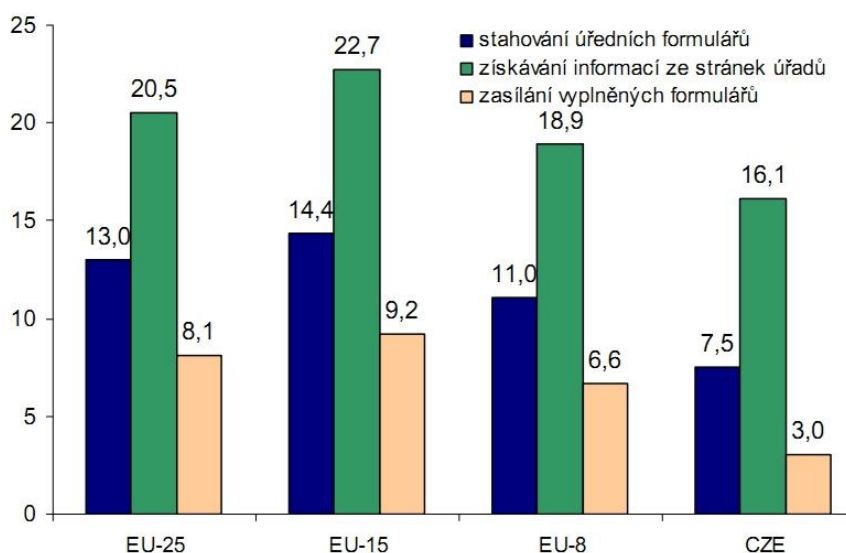
³⁷ Zdroj: Eurostat

³⁸ Zdroj: Eurostat



Obrázek 3. Jedinci, kteří používají internetu k podání vyplněných formulářů veřejné správy v roce 2010

Využívání informací z webových prezentací úřadu³⁹ je uvedeno na posledním grafu:



Obrázek 4. Využívání informací z webových prezentací úřadu

³⁹ Zdroj: Strategie SmartAdministration 2007-2015 + EUROSTAT 2007. Dostupné online na http://www.smartadministration.cz/file/9_1_1/download/

2 Vymezení pojmů

Tato kapitola se zabývá vymezením nejdůležitějších pojmů používaných v této práci a to v rozsahu potřebném pro pochopení kontextu.

2.1 Užití pojmů

Vymezení řady pojmů je nejednotné a v různých zdrojích vnímáno odlišně. Především vymezení pojmu „informace“ je značně komplikované, protože odborné zdroje jej považují za „zpracovaná data“ zatímco česká legislativa považuje pojem „informace“ za prakticky totožný pojmu „data“.

Pro účely této práce, a z hlediska jejího zadání, se přidržím české právní formy a nebudu, pokud není řečeno jinak, rozlišovat mezi pojmem „informace“ a pojmem „data“.

Pojem „otevřená data“ v sobě zahrnuje jak data veřejná (veřejné správy), tak data produkovaná soukromou a neziskovou sférou. Tato práce se zabývá pouze veřejnou správou, a proto, pokud není uvedeno jinak, je pojmem „otevřená data“ myšleno vždy „otevřená data veřejné správy“.

Některé zdroje pracují s pojmem „otevřená vládní data“. Toto názvosloví je v příslušných citacích zachováno, i zde se však jedná o „otevřená data veřejné správy“.

V odborné literatuře je pojem „taxonomie“ přísně vymezen a formulován jako základ pro vědecké třídění. V námi sledované oblasti však zpravidla dochází k zjemnění pojmu na „lineární slovník“ případně „hierarchii pojmů“. V této práci, pokud není uvedeno jinak, pracujeme s pojmem taxonomie v jeho zjednodušené podobě jako „hierarchie pojmů“, a pokud není uvedeno jinak, bude pojem totožný s pojmem „slovník“.

2.2 Klasifikace pojmu data

2.2.1 Data – informace – znalosti

Data jsou jakékoliv vyjádření (reprezentace) skutečnosti, schopné přenosu, interpretace či zpracování. Účelem dat je přenášet a dále zpracovávat odraz skutečnosti. Jsou to jakékoliv zaznamenané poznatky či fakta (5).

Informace je definována pomocí dat a znalostí. Jsou to data, která mají smysl (význam). Jsou to sdělitelné (komunikovatelné) znalosti. Je to údaj, ke kterému si člověk přiřadí význam. (5) Slovo informace pochází z latinského „informare“, neboli uváděti ve tvar, dodati tvar, podobu formovat, tvořit, zobrazovat, představovat, vytvářet představu, pojem (6).

Znalost je informace, která prošla uspořádáním a analýzou, aby se stala srozumitelnou a použitelnou k řešení problému nebo rozhodování. (7)

Znalost je tedy to, co jednotlivec ví po osvojení dat a informace a po jejich začlenění do souvislostí. Účelem znalosti je schopnost porozumět skutečnosti. Jako media pro uchovávání znalostí slouží lidská paměť, organizace, dokument nebo počítač. (5)

Vlastníme-li data, ale nechápeme jejich smysl, ztrácí pro nás význam, jsou nepoužitelná. Pokud informaci správně pochopíme, stává se pro nás znalostí, kterou můžeme zpracovat opět do dat. (5)

Zákon o svobodném přístupu k informacím⁴⁰ vymezuje pojem informace takto:

Informace je jakýkoliv obsah nebo jeho část v jakékoli podobě, zaznamenaný na jakémkoliv nosiči, zejména obsah písemného záznamu na listině, záznamu uloženého v elektronické podobě nebo záznamu zvukového, obrazového nebo audiovizuálního.

Zveřejněnou informací je taková informace, která může být vždy znovu vyhledána a získána, zejména vydaná tiskem nebo na jiném nosiči dat umožňujícím zápis a uchování informace, vystavená na úřední desce, s možností dálkového přístupu nebo umístěná v knihovně poskytující veřejné knihovnické a informační služby⁴¹.

Kvalita dat

Kvalita je jedním z rozhodujících faktorů zda data použijeme nebo ne. Samotná skutečnost že data máme k dispozici ještě nemusí znamenat, že nám budou prospěšná. Profesor Jaroslav Král na konferenci Datakon 2005 uvedl:

Je nutné zajistit nejenom ochranu dat, ale také jejich kvalitu a zavést procedury jak jednat, není-li kvalita dat ideální, ale musí se používat.

Zároveň zmínil hlavní kategorie a jejich dimenze pro měření kvality dat (pro oblast e-governmentu). Jedná se o:

| Kategorie | Dimenze |
|------------------------|--|
| Vnitřní, internistická | <ul style="list-style-type: none"> - Přesnost - Objektivnost - Důvěryhodnost - Reputace |
| Dostupnost | <ul style="list-style-type: none"> - Dostupnost - Bezpečnost |
| Kontextuální | <ul style="list-style-type: none"> - Relevantnost - Přínos - Včasnost - Úplnost - Rozsah |
| Reprezentační | <ul style="list-style-type: none"> - Interoperabilita - Srozumitelnost - Výstižná a stručná reprezentace - Konzistentní reprezentace |

Tabulka 4. Atributy kvality dat

⁴⁰ § 3 odst. 3) zákona 106/1999 Sb., o svobodném přístupu k informacím.

⁴¹ § 3 odst. 5) zákona 106/1999 Sb., o svobodném přístupu k informacím.

Dodržování datové kvality je v některých zemích i legislativně upraveno, pro příklad v USA je od roku 2001 v platnosti Data Quality Act⁴². Česká republika nemá legislativně upravenou oblast datové kvality.

Datovou kvalitu můžeme (částečně) zajistit dodržáním těchto pravidel:

Taxonomie: Použití standardizované taxonomie nebo ontologie pro publikaci dat a přidání metadat výrazně zvýší vypovídající hodnotu a použitelnost, jako je např. vyhledávání, třídění a porovnávání publikovaných dat.

Definice dat: Každá položka nebo sloupec v datech by měl mít definován formát, taxonomii, ontologii nebo slovník, který používá⁴³. Tím dosáhneme sjednocení a možnosti data z různých zdrojů mezi sebou porovnávat, třídít a klasifikovat.

Standardní formát souboru: Umožňuje snadnější integraci dat v rámci datových souborů. Podporuje myšlenku svobodného přístupu k datům bez nutnosti využívat uzavřené technologie.

Proces opravy chyb: Měl by existovat způsob jak informovat producenta dat o možném nesouladu dat s realitou a zajištění zpětné vazby konzumentům dat, aby se o opravě nebo aktualizaci dozvěděli.

Důvěrnost: Data by měla mít uveden zdroj a metodu zpracování. Pokud data vycházejí z jiných datových sad, mělo by to být rovněž uvedeno.

Datový prvek

Pojem datový prvek je definován různými zdroji podobně ale s drobnými rozdíly, nejčastěji používanými vymezeními je:

- Datový prvek je jednotka dat, která je v daném kontextu považována za nedělitelnou⁴⁴.
- Datový prvek je jednotka dat, pro kterou jsou stanoveny identifikace, popis a formát hodnoty⁴⁵.
- Datový prvek je jednotka dat, pro kterou se definice, identifikace, zobrazení a přípustné hodnoty specifikují podle souboru atributů⁴⁶.
- Datový prvek je jednotka dat, která je v daném kontextu dále považována za nedělitelnou a je jednoznačně definována⁴⁷.

Databáze

Speciálním typem dat jsou data jako celek – databáze. Databáze je definována jako soubor nezávislých děl, údajů nebo jiných prvků, systematicky nebo metodicky uspořádaných a individuálně přístupných elektronickými nebo jinými prostředky, bez ohledu na formu jejich vyjádření⁴⁸. V širším pojetí je databází míněno jak samotná data, tak jejich struktura.

⁴² Dostupné z <http://www.gpo.gov/fdsys/pkg/PLAW-106publ554/html/PLAW-106publ554.htm>

⁴³ ISVS např. povinně využívají datové definice a slovníky definované v ISDP.

⁴⁴ Podle ČSN ISO/IEC 2382-4

⁴⁵ Podle ČSN ISO 9735

⁴⁶ Podle ČSN ISO/IEC 11179-2

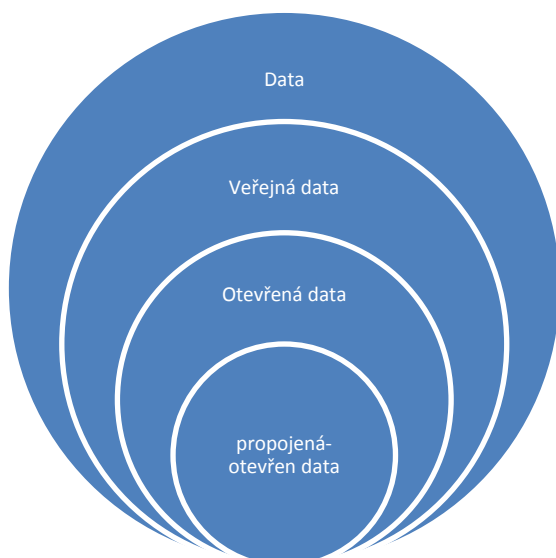
⁴⁷ Podle § 2 písm. f) zákona 365/2000 Sb.

⁴⁸ § 88 zákona č. 121/2000 Sb., zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, ve znění pozdějších předpisů.

2.2.2 Veřejná data

Veřejná data jsou všechna data veřejné správy, ke kterým není veřejnosti odepřen přístup. Data mohou být v jakékoliv podobě, mohou být zpoplatněna⁴⁹, případně může být omezeno jejich další použití. Specifickým případem veřejných dat jsou otevřená veřejná data v cizině zpravidla Open Government Data.

Strukturu dat ve smyslu přístupnosti zobrazuje následující diagram.



Obrázek 5. Diagram vztahu mezi typy dat

2.2.3 Strojově čitelná data

Strojově čitelná data jsou speciální formou dat, pro která platí, že jejich zpracování nebrání technické omezení.

Připravovaná novela zákona o svobodném přístupu k informacím⁵⁰ zavádí pojem „Strojově čitelný formát“ který je definován takto: „Strojově čitelným formátem se rozumí formát datového souboru, který je strukturovaný takovým způsobem, že s pomocí programové aplikace lze z tohoto souboru získat žádané (vybrané) údaje.“

- Strojově čitelná data = data přístupná.
- Dobře strojově čitelná data = data popsaná a přístupná.

2.2.4 Otevřená data

Otevřená data jsou „surová“ data, která jsou volně dostupná na internetu komukoliv, kdykoliv a pro libovolný účel. Otevřená data státní správy jsou jakékoliv informace shromažďované ve státní správě pro vlastní účely, která jsou zveřejněna tak, aby je i ostatní mohli použít pro vlastní účely.

Přímo pro otevřená data veřejné správy formulovala nadace Sunlight Foundation deset principů pro posuzování, zda jsou data veřejné správy otevřená a vhodně zpřístupněná veřejnosti. Podle těchto principů jsou data veřejné správy považována za otevřená, pokud jsou:

⁴⁹ Pouze do výše materiálových nákladů, podle § 17 zákona 106/1999 Sb., o svobodném přístupu k informacím.

⁵⁰ Novela je v době psaní tohoto textu je v meziresortním připomínkování.

1. úplná,
2. primární,
3. zveřejněná bez zbytečného odkladu,
4. snadno dostupná (fyzicky či elektronicky),
5. strojově čitelná,
6. neomezující nikoho ani žádný ze způsobů přístupu,
7. používající standardy s volně dostupnou specifikací (otevřené standardy),
8. jasně licencovaná za podmínek umožňujících jejich využití s minimem možných omezení,
9. stále dostupná,
dostupná uživatelům při vynaložení minima možných nákladů na jejich získání.

Výhody užití otevřených dat

Podle Tim Berners-Lee můžeme kvalitu otevřených dat hodnotit v pěti stupních, jedná se o:

1. Data jsou publikována pod otevřenou licencí.
2. Data jsou strojově čitelná.
3. Data nemají uzavřený formát.
4. Data jsou publikována pomocí RDF.
5. Data jsou publikována jako provázaná.

Klasifikace uvádí pro jednotlivé stupně tyto výhody při použití⁵¹:

| Stupeň | Výhoda pro konzumenta | Výhoda a podmínky pro producenta |
|--------|--|---|
| 1. | <ul style="list-style-type: none"> - Data můžete prohlížet a tisknout - Data můžete ručně zadat do jiného systému - Data můžete uložit a použít lokálně | <ul style="list-style-type: none"> - Jednoduchá publikace |
| 2. | <ul style="list-style-type: none"> - Všechny výhody předchozího stupně - Data můžete exportovat do jiného strukturovaného formátu - Data můžete přímo zpracovávat pomocí proprietárního SW např. pro výpočty nebo vizualizace | <ul style="list-style-type: none"> - Jednoduchá publikace |
| 3. | <ul style="list-style-type: none"> - Data můžete zpracovávat pomocí SW, který daný formát podporuje, nikoliv pouze s výchozím proprietárním SW | <ul style="list-style-type: none"> - Data bude možná nutné konvertovat do otevřeného datového formátu - Stále se však jedná o jednoduchou publikaci |
| 4. | <ul style="list-style-type: none"> - Na data je možné odkazovat z jiného místa - Data je možné tagovat | <ul style="list-style-type: none"> - Je nutné přiřadit URI k datovým položkám a přemýšlet o tom, jak reprezentovat data - Získáte kontrolu nad datovými |

⁵¹ Zdroj: lab.linkeddata.deri.ie/2010/star-scheme-by-example/.

| | | |
|----|--|--|
| | - Je možné znovupoužít část dat | položkami, díky které můžete optimalizovat jejich přístup (load balancing, cache) |
| 5. | - Můžete procházet další související data (i publikovaná jinde) - Znáte datové schéma | - Budete muset investovat prostředky k propojení dat s jinými daty na webu - Vaše data budou zjistitelná - Můžete zvýšit hodnotu svých dat |

Otevřená datová infrastruktura⁵²

Otevřená datová infrastruktura je vybudována na technologických principech Linked Data. Principy popisují, jak mají být data publikována na webu. Konkrétními principy jsou následující:

- Věci reálného světa i abstraktní koncepty mají přiřazeny jednoznačné webové adresy (tzv. URI).
- Data o věcech jsou publikována ve strojově čitelném formátu (RDF) v jedné či více datových sadách na webu.
- Datové sady jsou vzájemně provázány, takže data o jedné věci či konceptu v jedné sadě mohou být obohacena o další data z jiných sad.

2.2.5 Linked data – Otevřená provázaná data

Rozšiřují otevřená data o odkazy na jiná otevřená data. Těmito odkazy zvyšují svou vypovídající hodnotu, protože se uvádějí do širšího kontextu a zároveň zvyšují hodnotu dat, na které odkazují. Pojem nabývá na popularitě zejména s rostoucím nástupem sémantického webu.

Pojmem propojená data nebo otevřená-propojená data rozumíme sadu principů LinkedData

LinkedData představuje publikační model pro zveřejňování strukturovaných dat na webu. Je vymezen principy a doporučeními, která stanovují, jakým způsobem mají být data na web publikována. Postupy LinkedData mohou sloužit jako prostředek k integraci dat a zároveň vedou k uskutečnění vize sémantického webu. Cílem LinkedData je volné zpřístupňování dat ve formě, která je jednoduše automatizovaně zpracovatelná.

Tento publikační model je založen na webových standardech, jako HTTP a URI, a technologiích sémantického webu, jako je datový model RDF a ontologické jazyky (RDFS, OWL a SKOS).

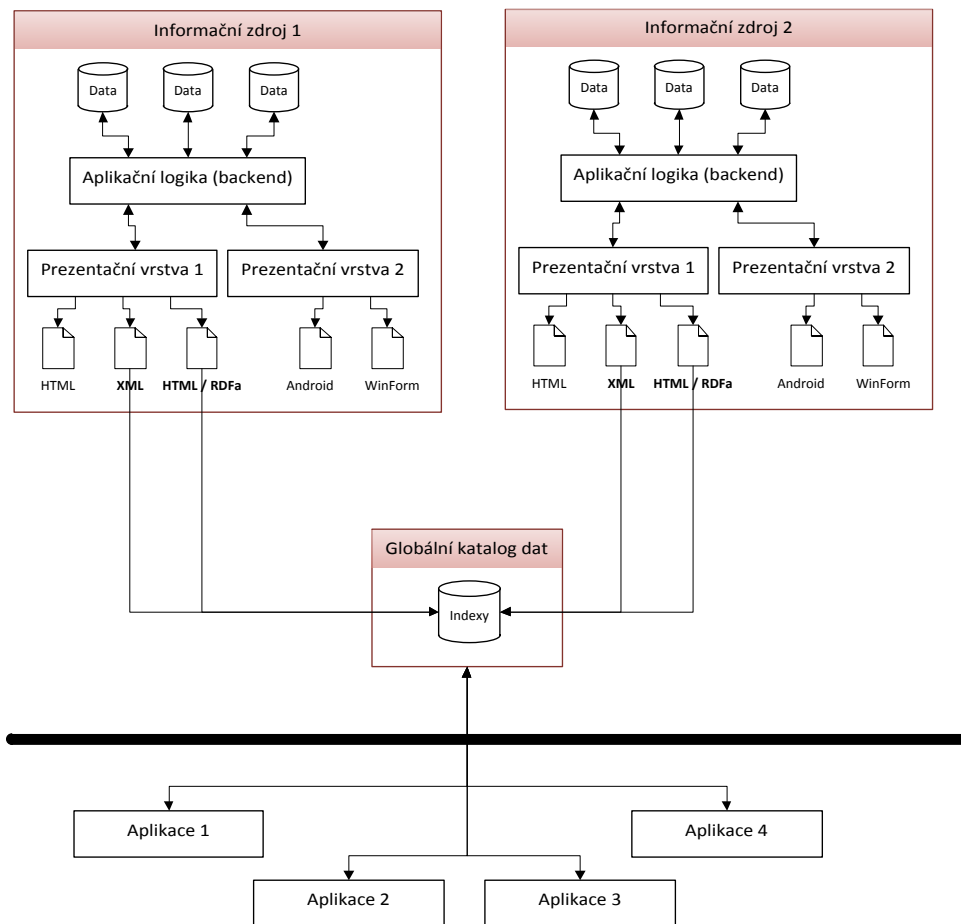
LinkedData je označení souhrnné pro

- metodu publikování strukturovaných dat na webu,
- instance dat, která jsou na webu publikována v souladu s touto metodou.

Ukázka možné publikace a propojení dat prostřednictvím katalogu naznačuje následující schéma. Schéma zobrazuje informační zdroje (portály, IS, apod.), kde každý má svou architekturu. Prezentační vrstvy jednotlivých informačních zdrojů publikují data pomocí různých prostředků přístupů a v různých formátech. Data, která nesou sémantickou informaci (např. prostřednictvím RDF) případně je možné je

⁵² Zdroj: Iniciativa pro otevřená data v ČR, www.opendata.cz

automatizovaně převést mohou být ukládána případně indexována ve společném uložišti. Toto uložišti se tak stává zdrojem informací z různých oblastí a je možné nad ním stavět funkčně bohaté aplikace. V tomto případě je globálním katalogem míněno nejenom dohledání existence dané informace v daném zdroji ale také uložení jeho sémantického obsahu.



Obrázek 6. Architektura publikace otevřených-propojených dat

Termín „*LinkedData*“ zpravidla nebývá překládán a používá se v původním anglickém znění. V této práci je místy nahrazen termínem „*otevřená-propojena data*“, a to z důvodu požadavku na kompletní Českou dokumentaci ze strany úřadů státní správy.

Detailně se můžeme s pojmem *LinkedData* seznámit např. na <http://www.LinkedException.org> nebo na <http://www.w3c.org/standards/semanticweb/data>.

2.3 Veřejná správa

Veřejná správa je základním pojmem správního práva, její legální definici však v žádném právním předpisu nenajdeme, i když řada právních předpisů jej používá⁵³. (8)

⁵³ Např. zákon č. 150/2002 Sb., soudní řád správní, ve znění pozdějších předpisů nebo zákon č. 365/2000 Sb., o informačních systémech veřejné správy a další zákony, ve znění pozdějších předpisů.

Např. P. Průcha veřejnou správu definuje jako: „Správu veřejných záležitostí ve veřejném zájmu a subjekty které ji vykonávají, ji realizují jako právem uloženou povinností, a to z titulu svého postavení jako veřejnoprávních subjektů.“ (9)

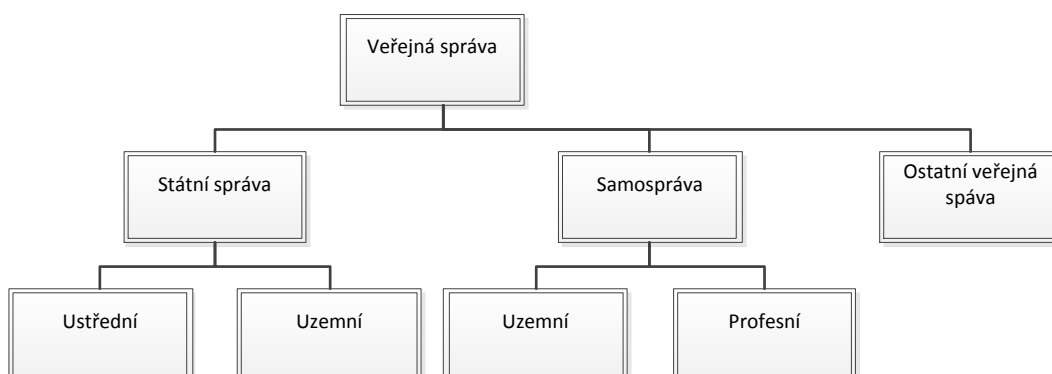
Slovník nejčastěji používaných pojmů ve veřejné správě⁵⁴ spravovaný Institutem pro veřejnou správu Praha obsahuje tuto definici⁵⁵:

Veřejnou správou rozumíme:

1. určitý druh činnosti (spravování veřejných záležitostí) a
2. instituce (organizace, úřad), které veřejnou správu vykonávají. V materiálním (funkčním) pojetí je veřejná správa činností státních nebo jiných veřejných institucí, která svým obsahem není ani činností zákonodárnou, ani soudní. Ve formálním (institucionálním, organizačním) pojetí je veřejná správa definována jako činnost orgánů označených jako správní úřady. Pojem veřejná správa je termínem společným (nadřazeným) pro pojem státní správa (která je vykonávána především státními orgány), samospráva (která je vykonávána orgány územních samosprávných celků či orgány zájmové/profesionální samosprávy) a ostatní veřejná správa (která je vykonávána zejména institucemi s právní subjektivitou – např. VZP, ČNB, ČT, ČTK).

Tato definice je rovněž součástí materiálu (10).

Veřejná správa je organizačně značně komplikovaná, její strukturu na nejvyšší úrovni zachycuje Obrázek 7, strukturu na základní úrovni pak Obrázek 8.



Obrázek 7. Organizační vymezení veřejné správy

⁵⁴ Zdroj: <http://www.mvcr.cz/clanek/slovník-nejcasteji-pouzivanych-pojmu-ve-verejne-sprave.aspx>

⁵⁵ Zdroj: <http://svs.institutpraha.cz/index.php?page=slovník&id=552>



Obrázek 8. Organizační struktura veřejné správy na místní úrovni (obec)

Konkrétní organizační struktura veřejné správy na základní úrovni, v našem případě města Poděbrady je naznačena v příloze.

2.4 Ontologie

Pojem ontologie je relativně široký. V této práci nepracujeme s filozofickým vymezením jako „nauka o bytí“ ale poněkud specializovanějším pro informatiku vhodným vymezením, které stanovil Tom Gruber ze Stanford University „explicitní specifikace konceptualizace“⁵⁶. Ontologie popisuje problematiku - doménu pomocí prvků, jako jsou instance, třída, atribut a vazba. Ontologie v tomto směru funguje jak slovník umožňující uchování a předání znalosti v dané doméně.

Účelem ontologií je podpora porozumění mezi lidmi, podpora komunikace mezi počítačovými systémy a podpora návrhu znalostně orientovaných systémů⁵⁷.



Obrázek 9. Znárodnění ontologie

⁵⁶ GRUBER, Tom. Knowledge Systems, AI Laboratory [online]. 1992 [cit. 2010-10-21]. What is an Ontology?. Dostupné z WWW: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>.

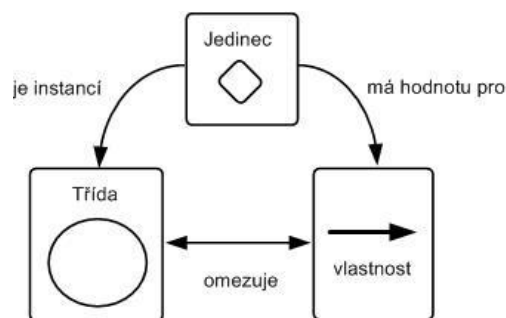
⁵⁷ HANYÁŠ, Petr. Sémantický web [online]. 2007-04-15 [cit. 2010-10-24]. Ontologie. Dostupné z WWW: <<http://www.hanyas.net/seweb/tutorial.php?article=109>>.

Instance (jedinec) je objekt našeho zájmu v popisované doméně, je buď konkrétní (kraj, město, člověk), nebo abstraktní (číslo, událost). Na diagramu jsou znázorněny jako uzly grafu „Moravskoslezský kraj“ a „Ostrava“. Některé ontologie nepodporují modelování instancí. Schopnost klasifikace instancí je jedním z podstatných posláních ontologie.

Třída (koncept, kategorie, rámeček) je základní stavební jednotkou ontologie. Obecně třídy v ontologii označují množiny konkrétních objektů, ale na rozdíl od objektově orientovaných modelů nezahrnují procedurální metody. Třída odpovídá unární relaci na dané doméně. V ukázkových příkladech jsou třídami „územně správní celek“, který je nadtřídou pro třídy „kraj“ a „obec“, které jsou tím pádem jeho podtřídami.

Vlastnosti (relace, vazba, vztah) specifikuje vztahy mezi instancemi nebo třídami, nejčastěji ve formě binárních relací. Na vlastnosti je možné klást omezení. V ukázce označeny jako (pojmenované) cesty mezi uzly grafu.

Vztahy mezi jednotlivými (zde zmíněnými prvky) naznačuje následující schéma: Schéma je součástí materiálu (11).



Obrázek 10. Vztahy mezi instancí, třídou a vlastností v ontologii

Doplňující informace včetně seznamu používaných ontologií a obecně o problematice sémantiky nalezneme na webovém portálu Semantic web⁵⁸.

S pojmem ontologie úzce souvisí pojem taxonomie. V oblasti standardizace a katalogizace se pojem z větší míry shoduje s pojmem klasifikace, tedy zařazení entity do určité skupiny, třídy. Třídy jsou zpravidla organizovány do hierarchie a vytvářejí strukturu. Díky tomu můžeme podle potřeby specializovat nebo generalizovat náš pohled na danou oblast.

Ontologii je možné zapisovat v mnoha jazycích, z nichž nejpoužívanější jsou jazyky OWL, RDFS a OIL. Jednotlivé jazyky mají různou vyjadřovací sílu a použití. Zápis jednoduché ontologie pomocí jazyka RDFS vypadá následovně:

⁵⁸ Dostupné na http://www.semanticweb.org/wiki/Main_page.

```

<rdfs:Class rdf:about="Kraj" rdfs:label="Kraj">
  <rdfs:subClassOf rdf:resource="UzemneSpravniCelek" />
</rdfs:Class>

<rdfs:Property rdf:about="RozlohaKM2">
  <rdfs:domain rdf:resource="UzemneSpravniCelek" />
</rdfs:Property>

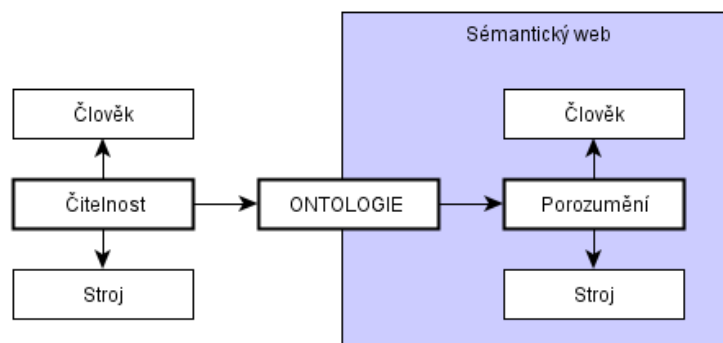
```

Obrázek 11. Ukázka použití jazyka RDFS

RDF Schéma (zkráceně RDFS) není klasickým ontologickým jazykem. Jedná se o „sémantické rozšíření formátu RDF. RDFS bylo vytvořeno v roce 1999 konsorciem W3C. Jde o nadstavbu, která doplňuje do struktury RDF hlavní konstrukce z rámcových (či objektových) systémů. Jedná se o třídy a binární sloty s možností stanovit definiční obor a obor hodnot. Kromě toho může být nad třídami i sloty definována hierarchie. Zdroje z RDF pak přiřazujeme třídám z RDFS jako jejich instance (pomocí atributu Type).

RDFS je intuitivní, což vyhovuje především webovým návrhářům při zachycení sémantiky obsahu stránek. Nevýhodou oproti tradičním ontologickým jazykům je neschopnost RDFS precizně specifikovat podmínky příslušnosti ke třídám (lokální omezení) a zcela postrádá datové typy.

Ontologie má významný vztah k sémantickému webu. Pomocí ní umožňujeme datům přidat sémantiku a zvýšit jejich vypovídající hodnotu a schopnost nejenom čitelnosti ale i porozumění. Vztah je znázorněn na následujícím diagramu:

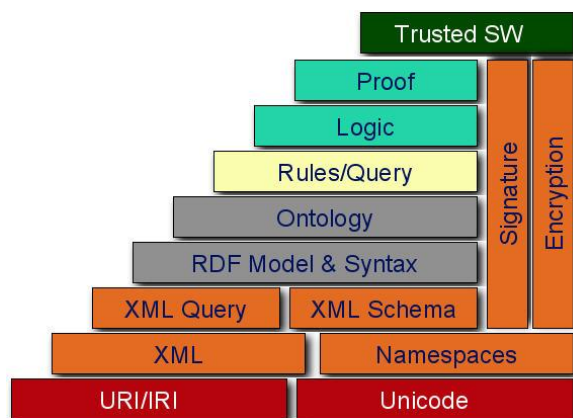


Obrázek 12. Vztah ontologie a sémantického webu

2.5 Sémantický web

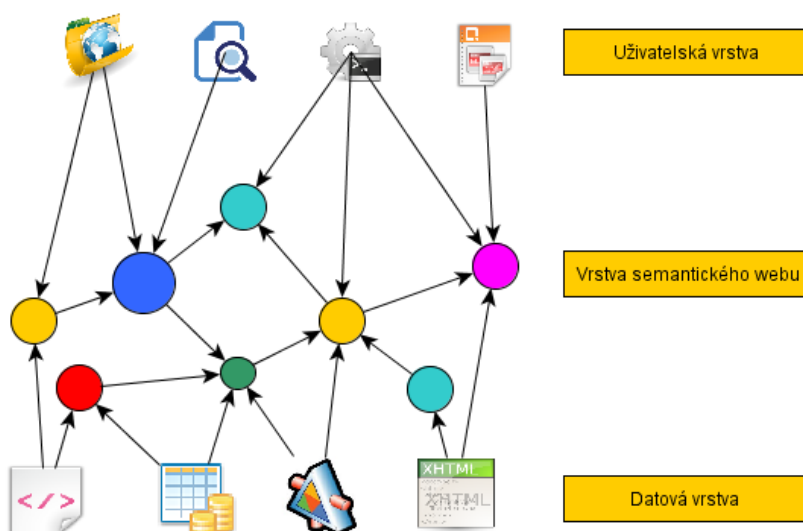
Sémantický web se má stát novým evolučním stupněm stávajícího webu. Jedná se o web, kde jsou informace strukturovány a uloženy podle standardizovaných pravidel, což usnadňuje jejich vyhledání a zpracování. Hlavním standardizovaným pravidlem je doplnění publikovaných informací o jejich sémantiku, tím se stávají srozumitelná nejenom pro tradiční (lidské), ale i strojové zpracování. Na specifikacích pro sémantický web pracuje World Wide Web Consortium⁵⁹. Architektura sémantického webu je znázorněna na diagramu.

⁵⁹ <http://w3c.org>



Obrázek 13. Architektura sémantického webu

Slovy zakladatele konsorcia W3C Tim Bearns-Lee sémantický web "není oddělený web, ale nástavba toho současného, ve kterém dostávají informace dobře definovaný význam, který umožňuje počítačům a lidem lépe spolupracovat"⁶⁰. Hlavním přínosem sémantického webu je zvýšení hodnoty publikovaných informací, přecházíme od webu dokumentů k webu dat.



Obrázek 14. Architektura sémantického webu

Jednou z mnoha aplikací, které staví na výhodách sémantického webu je služba WolframAlpha.

2.5.1 WolframAlpha

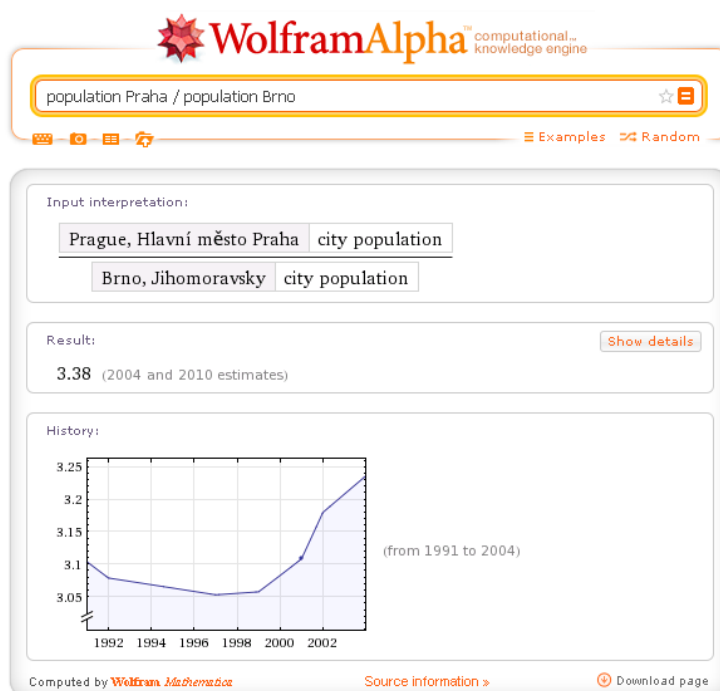
Wolfram alpha je služba provozovaná společností Wolfram Research a oficiálně byla spuštěna v květnu 2009. Zakladatelem je matematik a fyzik Stephen Wolfram.

⁶⁰ Zdroj: W3C Semantic Web Tutorial, dostupné z URL: <http://www.w3schools.com/semweb/default.asp> (květen 2010)

Wolfram Alpha je výpočetní znalostní stroj (computational knowledge engine), označovaný také jako znalostní vyhledávač, stroj na odpovědi (answer engine) či přirozeně vyhledávající stroj (natural search engines)⁶¹.

Wolfram Alpha indexuje strukturovaný obsah na internetu do své znalostní databáze, ta je následně doplňována a rozšiřována o redigovaný obsah vkládaný redaktory díky tomu je schopen podat relevantní odpovědi na dotazy z oblastí jako jsou medicína, matematika, geografie, socioekonomie, fyzika, chemie, historie, hudba apod.

Pro příklad při hledání odpovědi na dotaz „kolikrát je město Praha větší než město Brno. Dostaneme odpověď „3.38“.



Obrázek 15. Služba Wolfram Alpha

2.5.2 Technologie sémantického webu

Sémantický web využívá řadu technologií. Mezi nejdůležitější patří technologie pro identifikaci zdrojů URI, jazyky pro sémantický popis a jazyky pro popis ontologie.

Základní jazyky pro popis ontologie jsou Ontology Web Language a Resource Description Framework Schema. Základním jazykem pro sémantický popis je Resource Description Framework.

Implementace čistě sémanticky orientovaného webu je technologicky velmi náročná, a proto vznikají způsoby, jak do webu postaveného na stávajících technologiích přidat sémantickou informaci a zároveň zachovat jednoduchost původního webu. Jedná se např. o značkovací jazyky RDFa, Microdata, Mikroformáty.

⁶¹ Zdroj: <http://www.wolframalpha.com>

Pro efektivní vyhledávání ve strukturovaných datech se používá zpravidla jazyk SPRQL.

Jednotlivé technologie jsou popsány dále v textu.

Jazyk RDF

Jazyk RDF je obecný rámec pro popis, výměnu a znovupoužití metadat. Poskytuje jednoduchý model pro popis zdrojů, který není závislý na konkrétní implementaci. Datový model RDF umožňuje specifikovat trojici (zdroj, vlastnost, hodnota vlastnosti, tím se dosáhne přiřazení sémantiky zdroji a zvýšení jeho vypovídající hodnoty.

Značkovací jazyk

Značkovací jazyk je obecně prostředek pro doplnění dodatečné informace k jiným informacím – nejčastěji o jejich významu, struktuře, způsobu zobrazení. Mezi nejpoužívanější značkovací jazyky v oblasti sémantického webu patří RDFa, MicroFormaty a MicroData.

Jazyk RDFa

Jazyk RDFa je značkovací jazyk, technologie pro přenos strukturovaných informací uvnitř webových stránek. RDFa je jeden ze způsobů zápisu (serializace) datového formátu RDF. K přenosu dat v RDF používá atributy (X)HTML elementů webové stránky. Rozšiřuje způsoby použití atributů přítomných ve specifikaci (X)HTML (např. rel nebo href) a zavádí nové atributy (např. content). Doplnující informace k RDFa včetně rozsáhlého tutoriálu a mnoha praktických příkladů nalezneme na webu organizace W3C: <http://www.w3.org/TR/rdfa-lite/>.

Praktickou ukázkou využití RDFa předvedeme na příkladu portálu veřejné správy a jeho sekce kontaktních informací úřadů ČR.

```
<div>
  <h1>
    Hlavní město Praha</h1>
  <span>
    Adresa: Praha Mariánské náměstí 2/2</span>
  <span>
    Úřední hodiny podatelny</span>
  <ul>
    <li>Pondělí 8:00 - 12:00, 13:30 - 16:00</li>
    <li>Středa 8:00 - 12:00, 13:30 - 16:00</li>
    <li>Pátek 8:00 - 12:00, 13:30 - 16:00</li>
  </ul>
</div>
```

Obrázek 16. Současný stav publikace dat v HTML

```
<div xmlns:ont="http://ontology.data.gov.cz/1.0/" about="data.gov/ovm/00064581">
  <h1 property="ont:FullName">
    Hlavní město Praha</h1>
  <span property="ont:Address" content="data.gov.cz/adresa/123456">
    Adresa: Praha Mariánské náměstí 2/2</span>
</div>
```

Obrázek 17. Očekávaný stav publikace dat pomocí HTML / RDFa

Díky přidání značek do původního obsahu můžeme strojově rozpoznat význam jednotlivých údajů. Rozlišíme co je plný název entity a jaká je její adresa. Adresa je v našem případě reprezentována odkazem do základního registru RUIAN, který jednoznačně a trvale identifikuje všechny uzemní prvky v České republice.

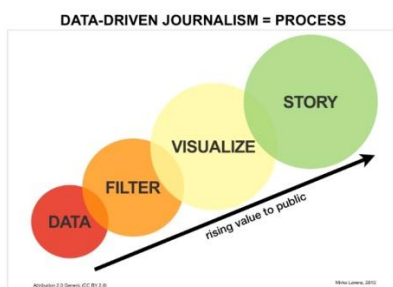
Použití RDFa pro zvýšení použitelnosti webových stránek postupně narůstá, nicméně v roce 2010 bylo pomocí RDFa označeno méně než 5 % všech veřejně dostupných stránek na webu⁶². Mezi uživatele této technologie patří např. BBC, kongresová knihovna US, stránky vlády USA a Velké Británie, LinkedIn a mnohé další.

⁶² Zdroj: Google, tisková zpráva 2010.

2.6 Žurnalistika řízená daty

Žurnalistika řízená daty je novinářský proces založený na analýze a filtrování velkých datových sad za účelem vytvoření nového příběhu. Žurnalistika řízená daty se snaží dosáhnout nové úrovně služeb pro veřejnost, pomoci spotřebitelům, manažerům a politikům pochopit strukturu a umožnit rozhodování na základě zjištění⁶³.

*Data driven journalism is a workflow that consists of the following elements: digging deep into data by scraping, cleansing and structuring it, filtering by mining for specific information, visualizing it and making a story.*⁶⁴



Obrázek 18. Proces Žurnalistiky řízené daty

Příkladem pro novinový článek postavený nad principy žurnalistiky řízené daty je článek pojednávající o kvalitě života v jednotlivých českých městech. Obsahem článku jsou např. srovnání míry kriminality, cenové mapy, dostupnosti služeb apod. Tato „surová“ data projdou analýzou a interpretací, výstupem je „objektivní“ zhodnocení.

Mezi nejznámější portály žurnalistiky založené na datech je britský deník The Guardian⁶⁵.

2.7 Ostatní

Dále jsou v práci používány tyto pojmy a termíny.

| Pojem | Popis |
|------------|--|
| Akční plán | Odsouhlasený a závazný postup vyhodnotitelných činností, který zpravidla sleduje splnění některé klíčové priority organizace. Akční plán může být zpracovaný v různé míře podrobnosti. Většinou je akční plán založen na měřitelných informacích, které poskytuje organizace a je cyklicky hodnocený ⁶⁶ . |
| CzechPOINT | Projekt Ministerstva vnitra ČR (zkratka pro Český Podací Ověřovací Informační Národní Terminál) kladoucí si za cíl vytvořit univerzální podatelnu, ověřovací místo a informační centrum, kde by bylo možné na jednom místě (úřadě) získat |

⁶³ Zdroj: <http://www.datadrivenjournalism.net>

⁶⁴ Mirko Lorenz, Information architect and multimedia journalist

⁶⁵ <http://www.guardian.co.uk/data>

⁶⁶ Zdroj: Institut pro veřejnou správu Praha, <http://svs.institutpraha.cz>.

| | |
|---------------------------|---|
| | veškeré údaje, opisy a výpisy, které jsou vedeny v centrálních veřejných evidencích a registrech, jakož i v centrálních neveřejných evidencích a registrech ke své osobě, věcem a právům. Místo, kde je dále možné ověřit dokumenty, listiny, podpisy a také elektronickou podobu dokumentů, učinit podání ke kterémukoli úřadu veřejné správy, a konečně získat informace o průběhu řízení ve všech věcech, které stát k jeho osobě vede ⁶⁷ . |
| CzechPOINT@home | Záměr Ministerstva vnitra ČR, ve kterém budou moci občané vykonávat vybrané agendy vůči veřejné správě samostatně. |
| CzechPOINT@office | Je neveřejné pracoviště úřadu, kde úředník samostatně čerpá informace, ověřuje a předkládá podání v rámci k eGovernmentu. Je určený pro úředníky orgánů veřejné moci, kteří ze zákona přistupují k rejstříkům nebo provádějí autorizovanou konverzi dokumentů z moci úřední ⁶⁸ . |
| Datové schránky | Datové schránky jsou informační systém veřejné správy zřizovaný podle zákona 300/2008 Sb. Datová schránka slouží pro komunikaci v oblasti veřejné správy. Jejím prostřednictvím lze činit podání kterémukoliv úřadu. Úřady prostřednictvím datové schránky doručují své písemnosti příslušným adresátům (fyzickým nebo právnickým osobám), stejně jako komunikují s jinými orgány veřejné správy. Veškerým úkonům, které jsou prostřednictvím elektronické datové schránky, resp. přepážky činěny, je přiznána ekvivalence k úkonům činěným písemně. ⁶⁹ |
| E-government | Se zabývá elektronizací výkonů veřejné správy. Transformace vnitřních a vnějších vztahů veřejné správy pomocí informačních a komunikačních technologií s cílem optimalizovat interní procesy ve vazbě na zákonné role jednotlivých účastníků ⁷⁰ . |
| Identifikační číslo osoby | Je v České republice unikátní osmimístné identifikační číslo právnické osoby, podnikající fyzické osoby nebo organizační složky státu. Starší kratší čísla jsou zepředu doplněna nulami. Přidělené IČO nesmí být přiděleno žádnému dalšímu subjektu, a to ani v případě že původní nositel již zanikl. Výjimkou je případ přeměny ekonomického subjektu ze zákona, nebo změna právní formy společnosti či družstva, kdy je subjektu ponecháno původní IČO. Způsob tvorby IČO stanoví podle zákona ⁷¹ Český statistický úřad, který je také pověřen vedením veřejné evidence. IČO je |

⁶⁷ Zdroj: Institut pro veřejnou správu Praha, <http://svs.institutpraha.cz>.

⁶⁸ Zdroj: Institut pro veřejnou správu Praha, <http://svs.institutpraha.cz>.

⁶⁹ Zdroj: <http://www.szrcr.cz/slovník-pojmu/datova-schranka-ds>

⁷⁰ PhDr. Robert Ledvinka, Konference ISSS 2012 Hradec Králové.

⁷¹ Zákon č. 89/2005 Sb., o státní statistické službě

| | |
|-----------------------------------|---|
| | <p>bezvýznamový identifikátor, kde poslední číslice je kontrolní. Algoritmus výpočtu je uveden např. zde⁷².</p> <p>IČO je díky svým vlastnostem vhodným kandidátem na jednoznačný identifikátor pro orgány veřejné moci a mimo rámec této práce i pro ostatní podnikající subjekty.</p> |
| Informační systém | Funkční celek, nebo jeho část, zabezpečující cílevědomou a systematickou informační činnost. Každý informační systém zahrnuje data, která jsou uspořádána tak, aby bylo možné jejich zpracování a zpřístupnění, a dále nástroje umožňující výkon informačních činností ⁷³ . |
| Informační systémy veřejné správy | Jsou souborem informačních systémů, které slouží pro výkon veřejné správy. Jsou jimi i informační systémy zajišťující činnosti podle zvláštních zákonů ⁷⁴ . |
| Konzument dat | Je v našem pojetí libovolná aplikace, služba nebo obecně proces, která pro svoji činnost využívá data. |
| Metadata | Data, popisující jiná data. Typickým příkladem jsou data XML schémat, popisující data instancí XML dokumentů. |
| Metainformační systém | Informační systém, který uchovává mimo jiné i popisy datových a funkčních rozhraní všech aplikací a služeb dostupných prostřednictvím referenčního rozhraní. |
| Orgán veřejné moci | <p>Je orgán, který je oprávněn autoritativně rozhodovat o právech a povinnostech fyzických a právnických osob, vůči nimž jedná vrchnostensky⁷⁵.</p> <p>V ČR může jako OVM vystupovat:</p> <ul style="list-style-type: none"> - Státní orgán (ministerstvo, soud, Policie ČR, správní úřad) - Orgány územní samosprávy - Orgány zájmové samosprávy |
| Orgán veřejné správy | Viz Orgán veřejné moci. Jedná se o starší terminologii zachovanou v některých zákonech. Některé materiály uvádějí rozdílnou definici pojmu OVM a OVS, s touto definicí nebudeme pracovat. |
| Producent dat | Je instituce nebo jednotlivec, který svou činností vytváří a publikuje data. |
| Provozní informační systém | Informační systém zajišťující informační činnosti nutné pro vnitřní provoz příslušného orgánu, například účetnictví, |

⁷² Zdroj: časopis Kriminální č. 3/1999, dostupné online:

aplikace.mvcr.cz/archiv2008/časopisy/kriminalistika/1999/9903/rak.html

⁷³ § 2 písm. b) zákona 365/2000 Sb., o informačních systémech veřejné správy.

⁷⁴ § 3 odst. 1. zákona 365/2000 Sb., o informačních systémech veřejné správy.

⁷⁵ Usnesení Ústavního soudu ze dne 25. listopadu 1993, sp. zn. II. ÚS 75/93.

| | |
|-----------------------------------|---|
| | správu majetku, nesouvisející bezprostředně s výkonem veřejné správy ⁷⁶ . |
| Representational State Transfer | sada architektonických principů pro návrh webových služeb, kde jednotlivé funkce těchto služeb jsou přímo přístupné pomocí protokolu http. |
| Servisně orientovaná architektura | Je obecně koncept realizace architektury informačních systémů, a to jak po stránce aplikační, tak po stránce technologické. Základem tohoto konceptu je rozdělení informačního systému na samostatné části, které poskytují služby. Architektura, která podporuje servisní orientaci a skládá se ze služeb. (12) |
| Služba | Je obecně schopnost nebo vlastnost vykonat nebo poskytnout něco, co někdo jiný potřebuje. |
| SPARQL | Simple Protocol And RDF Query Language je klíčovou technologií pro sémantický web, a převážně dotazovací jazyk, který je určen k manipulaci s RDF databázemi a k dotazování nad RDF grafy. |
| Uniform Resource Identifier | Je řetězec znaků s přesně definovanou strukturou, který identifikuje abstraktní nebo fyzický zdroj pomocí rozšiřitelných prostředků. Slouží k přesné specifikaci zdroje informací (ve smyslu dokument nebo služba), hlavně za účelem jejich použití pomocí počítačové sítě, zejména Internetu. URI je popsáno mezinárodní normou IETF RFC 2396 – Uniform Resource Identifier (URI): Generic Syntax ⁷⁷ . |
| Veřejný informační systém | Informační systém vedený správcem – orgány veřejné moci nebo jiný informační systém poskytující služby veřejnosti, který má vazby na informační systémy veřejné správy ⁷⁸ . |

Tabulka 5. Použité pojmy

2.7.1 Použité zkratky

V práci se používají následující zkratky:

| Zkratka | Vysvětlení |
|---------|----------------------------|
| AIS | Agendový informační systém |
| CzP | CzechPOINT |
| ČSU | Český statistický úřad |

⁷⁶ § 2 písm. b) zákona 365/2000 Sb., o informačních systémech veřejné správy.

⁷⁷ Dostupné online na <http://www.ietf.org/rfc/rfc2396.txt>

⁷⁸ § 2 písm. s) zákona 365/2000 Sb., o informačních systémech veřejné správy.

| | |
|------------------------------------|---|
| IČO (dříve také IČ ⁷⁹) | Identifikační číslo osoby |
| IOP | Integrovaný operační program (dotační program EU) |
| IS ISVS | Informační systém o informačních systémech veřejné správy |
| ISDP | Informační systém datových prvků |
| ISDS | Informační systém datových schránek |
| ISVS | Informační systém(y) veřejné správy |
| ISZR | Informační systém základních registrů |
| JIP | Jednotný identitní prostor |
| KAAS | Katalog autentizačních a autorizačních služeb |
| LD | LinkedData, otevřená-propojená data (veřejné správy) Pro účely této práce je zkratka totožná s LoD |
| LoD | Linked Open Data, otevřená-propojená data (veřejné správy) |
| OECD | Organizace pro hospodářskou spolupráci a rozvoj |
| OSN | Organizace spojených národů |
| OVM | Orgán veřejné moci |
| OVS | Orgán veřejné správy |
| RDF | Resource Description Framework |
| RDFa | Resource Description Framework in attributes |
| REST | Representational State Transfer |
| ROB | Registr obyvatel (fyzických subjektů) |
| ROS | Registr osob (právnických subjektů) |
| RPP | Registr práv a povinností |
| RUIAN | Registr územní identifikace, adres a nemovitostí |
| SOA | Servisně orientovaná architektura |
| ÚOHS | Úřad pro ochranu hospodářské soutěže |
| ÚOOÚ | Úřad pro ochranu osobních údajů |

⁷⁹ Zákon 111/2009 Sb., o základních registrech stanovil od 1. června 2010 oficiální zkratku IČO, dříve nebyla zkratka pevně ukotvena.

| | |
|-----|---|
| URI | Uniform Resource Identifier |
| VDP | Veřejný dálkový přístup (k datům RUIAN) |
| VS | Veřejná správa |
| WS | Web services, webová služba |

Tabulka 6. Použité zkratky

3 Popis současného stavu řešeného problému

3.1 Data ve veřejné správě České republiky

Stát nemá politiku jak s daty nakládat.⁸⁰

Veřejná správa České republiky je významným producentem dat. Kvalifikovaný odhad stanovuje, že přibližně 30 000 úřadů spravuje cca 1000 agend⁸¹, prakticky každá agenda zveřejňuje část svých dat.

Mezi nejvýznamnější veřejně dostupné datové zdroje můžeme zařadit především:

- Portál veřejné správy České republiky
- Portál veřejných zakázek
- Portály regionů, ústředních orgánů veřejné moci a institucí zřizovaných státem
- Obchodní rejstřík
- Insolvenční rejstřík
- Výsledky voleb
- A další významné zdroje uvedené v příloze

U žádného ze sledovaných systémů nedochází k (významné) publikaci dat v souladu s principy otevřená-propojená data.

... V posledních letech došlo k výraznému nárůstu činností, jejichž cílem je získávání dat pro výkaznictví, zejména statistického, ekonomického a účetního charakteru. Potřeba získání dat není koordinována, data jsou pořizována multiplicitně a nejsou sdíleně využívána a aktualizována. Tímto způsobem vzniká velká byrokratická zátěž, včetně nárůstu potřeby finančních prostředků na její zajištění. (13)

3.1.1 Portál veřejné správy České republiky

Portál veřejné správy je elektronická brána do veřejné správy a vznikl na základě zákona⁸². Hlavním smyslem portálu je usnadnit občanům a firmám orientaci (informační část) a komunikaci (transakční část) s úřady veřejné správy.

Portál veřejné správy významným způsobem přispívá k potřebě kvalitních služeb při poskytování důvěryhodných a garantovaných informací širokému spektru občanů ČR, včetně poskytování relevantních informací cizincům a k zjednodušení komunikace s úřady. Cílem je přispět k modernizaci veřejné správy také prostřednictvím informačních a komunikačních technologií a tím postupně naplňovat ústřední motto „Efektivní veřejná správa a přátelské veřejné služby“⁸³.

⁸⁰ Ing. Vladimír Weis, odbor hlavního architekta eGovernmentu MVČR, pracovní jednání řešitelského týmu Katalogizace otevřených dat veřejné správy ČR, 25. 5. 2012.

⁸¹ Jedná se o kvalifikovaný odhad uvedený v dokumentu Ministerstva vnitra ČR nazvaném Základní registry – Operační strategie.

⁸² Zákona č. 365/2000 Sb., o informačních systémech veřejné správy.

⁸³ Zdroj: <http://www.mvcr.cz/clanek/portal-verejne-spravy.aspx>.

Portál veřejné správy je předurčen, v souladu s programovým cílem vlády, stát se místem, které bude „integrovat a zpřístupňovat všechny zveřejňované a veřejně přístupné informace veřejné správy“, včetně možné komunikace s úřady⁸⁴.

Portál veřejné správy je svým zaměřením určen pro širokou veřejnost, státní správu a samosprávu, státní i soukromé organizace včetně podnikatelů, živnostníků a cizinců.

Portál veřejné správy poskytuje následující kategorie informací:

Informační část (poskytování garantovaných informací):

- Adresář úřadů ČR
 - o Úřady veřejné správy
 - podle regionů
 - podle druhu
 - o Činnosti úřadů
 - o Katalog informačních zdrojů veřejné správy
- Postupy pro řešení více než 450 životních situací napříč veřejnou správou
 - o Garantovány a aktualizovány subjekty veřejné správy
 - o Propojeny s databází Zákony, s Adresářem úřadů, s Agendami úřadů se službou Elektronická podání
 - o Jsou přístupné dálkově prostřednictvím Internetu na jednom místě
- Mapové služby – tematické mapové úlohy
- Zákony
- Obchodní věstník
- Nestátní neziskové organizace
- Evidence elektronických podatelů
- Povinně zveřejňované informace a novinky z veřejné správy
- Online noviny Veřejné Správy
- Základní informace pro občana, podnikatele a cizince formou rozcestníku
- Informace pro cizince v anglickém jazyce
- Slovník nejčastěji používaných pojmů ve veřejné správě

Transakční část (komunikace s úřady)

Bývalá transakční část Portálu veřejné správy je od 1. ledna 2012 nahrazena informačním systémem datových schránek. Tento krok byl učiněn v důsledku finančních úspor a je v souladu s platnou legislativou.

Detailní informace o portálu veřejné správy nalezneme na webu provozovatele - Ministerstva Vnitřní věci⁸⁵.

3.1.2 Portály regionů a ústředních orgánů státní správy

Jednotlivé OVM vedou vlastní webové prezentace, které jsou zdrojem u krajské samosprávy místně příslušných a u ústředních věcně příslušných informací. Odkazy na jednotlivé portály nalezneme na portálu veřejné správy <http://www.portal.gov.cz>

3.1.3 Portál o veřejných zakázkách a koncesích

Oblast zadávání veřejných zakázek upravuje zákon⁸⁶. Zadávání veřejných zakázek je významnou součástí národního hospodářství a tvoří přibližně 17 % HDP⁸⁷

⁸⁴ Zdroj: <http://www.mvcr.cz/clanek/portal-verejne-spravy.aspx>.

⁸⁵ Dostupné na <http://www.mvcr.cz/clanek/portal-verejne-spravy.aspx>.

v ČR, což představuje v absolutní výši více než 600 miliard korun ročně⁸⁸. V ČR je elektronicky uveřejňováno cca 55 % celkového finančního objemu veřejných zakázek, ostatní představují veřejné zakázky malého rozsahu nebo zakázky dle výjimek z působnosti zákona⁸⁹.

NIPEZ – Národní infrastruktura pro elektronické zadávání veřejných zakázek. NIPEZ je modulárně členěná soustava informačních systémů a komunikačního rozhraní, jehož cílem je podpořit procesy elektronizace zadávání veřejných zakázek. Hlavním cílem projektu NIPEZ je dosáhnout finančních úspor elektronizací zadávání veřejných zakázek v ČR, a to snížením cen nakupovaných komodit a snížením transakčních nákladů spojených s procesy zadávání veřejných zakázek na straně zadavatelů i dodavatelů⁹⁰.

Povinnost užívat systém NIPEZ vznikne všem subjektům veřejné správy od 1. 7. 2012⁹¹. Podrobné informace o NIPEZ a modulech e-tržišť naleznete na portálu VZ www.portal-vz.cz a na webu Ministerstva pro místní rozvoj www.mmr.cz. Dále je v provozu informační systém o veřejných zakázkách, který je přístupný na adrese www.isvzus.cz.

3.1.4 Systém základních registrů

Základní registry nepopisují stav faktický nýbrž stav legální⁹².

Projekty systému základních registrů jsou součástí konceptu „Základní cíle Strategie efektivní veřejná správa a přátelské veřejné služby⁹³ (Smart Administration) v období 2007 – 2015“, který vláda ČR projednala spolu s IOP dne 28. února 2007⁹⁴, a „Strategie Efektivní veřejná správa a přátelské veřejné služby v období 2007 – 2015“, která dále rozpracovává materiál „Základní cíle Strategie Efektivní veřejná správa a přátelské veřejné služby v období 2007 – 2015“⁹⁵. Legislativní oporou realizace systému základních registrů je zákon č. 111/2009 Sb., o základních registrech.

Základní registry se stanou ke dni 1. 7. 2012 patřící informačnímu systému státní správy⁹⁶.

Systém základních registrů je dle zákona⁹⁷ koncipován tak, že budou existovat čtyři základní registry, a podpůrné systémy:

- Registr osob,
- Registr občanů,
- Registr územních identifikací, adresa a nemovitostí,
- Registr práv a povinností,
- Informační systém základních registrů,
- Převodník identifikátorů fyzických osob.

⁸⁶ Zákon č. 137/2006 Sb., o veřejných zakázkách.

⁸⁷ Hrubý domácí produkt – Celková peněžní hodnota statků a služeb vytvořena za dané období (zpravidla jeden rok) na určitém území.

⁸⁸ Zdroj: Ministerstvo pro místní rozvoj, statistika pro rok 2010.

⁸⁹ § 18 a § 19 zákona č. 137/2006 Sb., o veřejných zakázkách.

⁹⁰ Zdroj: Portál veřejných zakázek, manažerské shrnutí

⁹¹ Zdroj: <http://www.portal-vz.cz/nipez/co-je-nipez/>

⁹² Ing. Ondřej Felix, CSc., Hlavní architekt eGovernmentu České republiky.

⁹³ Dostupné na adrese www.smartadministration.cz

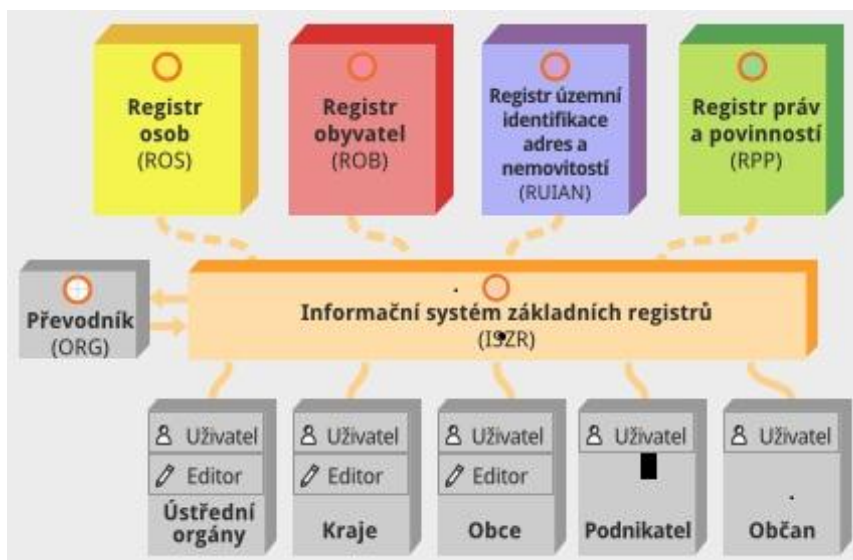
⁹⁴ Usnesení vlády č. 197/2007

⁹⁵ Usnesení vlády č. 757/2007

⁹⁶ Podle zákona 111/2009 Sb., o základních registrech.

⁹⁷ Zákona č. 111/2009 Sb., o základních registrech.

Konceptuální rozložení systému základních registrů je následující⁹⁸:



Obrázek 19. Architektura systému základních registrů

Jednotlivé registry jsou zdrojem referenčních dat. Referenčními daty v terminologii základních registrů jsou míněna data správná, úplná, aktuální a závazná⁹⁹.

Dle zákona 111/2009 Sb. o základních registrech mohou k údajům vedených v základních registrech¹⁰⁰ přistupovat pouze orgány veřejné moci, a to prostřednictvím agendového informačního systému¹⁰¹ v rozsahu jejich agendy¹⁰².

Výjimku z toho má RUIAN, popis je uveden v kapitole 3.3.2.

Jediná možnost přímého zpřístupnění těchto údajů, byť v agregované podobě, je možná pouze parlamentní změnou zákona. Tato cesta je i vzhledem k dlouho trvající přípravě zákona náročná a zdlouhavá, v tuto chvíli se jí nemá smysl zabývat.

Údaje mohou být zveřejněny také nepřímo, a to prostřednictvím speciálního agendového informačního systému. Tento AIS získá v rámci své agendy údaje ze ZR a nad těmito daty provede agregaci.

⁹⁸ Zdroj: Správa základních registrů. Dostupné na <http://www.szrcr.cz/>.

⁹⁹ Rozjezd projektů základních registrů, s. 39. Ondřej Felix, Hlavní architekt eGovernmentu ČR.

¹⁰⁰ Informační systém veřejné správy uvedený v § 3 zákona 111/2009 Sb.

¹⁰¹ Informační systém veřejné správy, který slouží k výkonu agendy. [Zákon 111/2009 Sb.]

¹⁰² Souhrn činností spočívajících ve výkonu vymezeného okruhu vzájemně souvisejících činností v rámci působnosti orgánu veřejné moci. [Zákon 111/2009 Sb.]

3.1.5 Ostatní producenti dat

Seznam významných producentů dat a nejčastější způsob publikace jejich dat je uveden v tabulce (14).

| Portál | Popis dat | Formát |
|--|--|-----------|
| ESF ČR | Poskytuje data o projektech financovaných strukturálním fondem ESF ČR. Data jsou poskytována ve formátu HTML v jednotné struktuře, která umožňuje vytěžování. | HTML |
| Informační systém o veřejných zakázkách | Poskytuje povinně zveřejňované údaje o nadlimitních a podlimitních veřejných zakázkách VS ČR. Data jsou uveřejňovány v jednotné HTML struktuře, kterou lze po poměrně časově náročném programování strojově vytěžovat. Obsahuje často nevalidní data (neexistující IČ, nesmyslné částky, ...). | HTML |
| Administrativní registr ekonomických subjektů (ARES) | Jedná se o XML datové API umožňující přístup k datům v několika zdrojových registrech (např. v obchodní, rejstříku). XML struktura reflektuje logickou strukturu zveřejňovaných dat, a proto je strojové zpracování poměrně snadné. Data ale nejsou zcela aktuální a úplná. | XML |
| Veřejné rozhraní k Centrálnímu systému účetních informací státu (ÚFIS) | Jedná se o XML datové API umožňující přístup k vybraným údajům účetních a finančních informací státu. XML struktura reflektuje logickou strukturu zveřejňovaných dat. | XML |
| Informační systém výzkumu, experimentálního vývoje a inovací (VaVal) | Informační systém poskytuje data o výzkumných aktivitách, projektech, záměrech, výsledcích a veřejných soutěžích. Data umožňuje exportovat do formátu XLS | HTML, XLS |

3.2 Současný stav publikování dat

VS publikuje svá data celou řadou přístupů a v celé škále formátů. Přístupy jsou nejčastěji:

- Statické soubory
- Dynamické soubory (služby na bázi REST)
- Rozhraní (webservices, ...)

Výstupem je zpravidla soubor určitého formátu. Zhodnocení vhodnosti jednotlivých formátů pro efektivní publikaci je zpracován v tabulce (14).

| Formát | Nezávislost na aplikaci | Zápis ve strukturované podobě | Popis struktury dat | Popis sémantiky dat | Vytváření propojení |
|-------------|-------------------------|-------------------------------|---------------------|---------------------|---------------------|
| Obrázek | -- | Ne | Ne | Ne | Ne |
| PDF | Ne | Ne | Ne | Ne | Ne |
| DOC(X), RTF | Ne | Ne | Ne | Ne | Ne |
| TXT | Ano | Ne | Ne | Ne | Ne |
| HTML | Ano | Částečně | Ne | Ne | Ne |
| XLS(X) | Ne | Částečně | Ne | Ne | Ne |
| CSV | Ano | Ano | Částečně | Ne | Ne |
| XML | Ano | Ano | Ano | Ne | Ne |
| OData | Ano | Ano | Ano | Částečně | Částečně |
| RDF | Ano | Ano | Ano | Ano | Ano |

Tabulka 7. Vlastnosti formátů dat

Popis a celkové zhodnocení jednotlivých dnes používaných formátů je uvedeno v tabulce (14).

| Formát | Komentář | Vhodnost použití ¹⁰³ |
|-------------|---|---------------------------------|
| Obrázek | Je vhodný pouze pro publikaci obrazových informací, jeho použití pro ostatní typy dokumentu a zvláště strukturovaných informací ve formě naskenovaného dokumentu je velmi nevhodné. | 5 |
| 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 |

¹⁰³ Stupeň 1-nejlepší, stupeň 5-nejhorší.

| | | |
|--------|---|---|
| HTML | Je často používaný 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 SW. 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 |
| 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. 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. | 3 |
| XML | Umožňuje zveřejňování strukturovaných dat (nejenom tabulkových ale i hierarchických dat) ve strojově čitelné podobě. Strukturu je navíc možné popsat pomocí standardizovaného prostředku (např. jazyka XML Schema). Je často používaný 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 | Má podobné vlastnosti jako formát XML. Navíc umožňuje propojování souvisejících dat z různých zdrojů a pomocí specializovaných jazyků (OWL) umožňuje i popis sémantiky. | 1 |

Tabulka 8. Popis a zhodnocení formátů dat

3.2.1 Právní úprava informačních systému veřejné správy

Informační systémy veřejné správy představují významný nástroj výkonu veřejné správy na všech úrovních. S ohledem na potřebu zajistit transparentnost, vzájemnou kompatibilitu, komunikaci a určitý stupeň jednotnosti všech ISVS, je kladen velký důraz na řízení této oblasti (15).

Budování, rozvoj a provoz IS VS na řadě míst podléhá zákonné úpravě. Seznam zákonů a vyhlášek, které významně ovlivňují IS je uveden v příloze. Proces budování IS VS je ovlivněn zejména zákonem 365/2000 Sb., o informačních systémech veřejné správy. Pro účely publikace dat jako otevřená resp. jako otevřená-propojená data jsou zajímavá zejména tato ustanovení:

Ministerstvo vnitra vyhledává, zpracovává, ukládá a vytváří nové informace, které jsou znalostní základnou pro kvalitní vytváření a rozvoj informačních systémů veřejné správy¹⁰⁴.

Ministerstvo vnitra stanoví pravidla pro sdílení dat a služby mezi jednotlivými informačními systémy veřejné správy prostřednictvím referenčního rozhraní a pravidla pro zápis datových prvků do informačního systému o datových prvcích. Postupy ministerstva a orgánů veřejné správy při vedení a zápisu datových prvků do informačního systému o datových prvcích, včetně postupů ministerstva při vyhlašování datových prvků, stanoví prováděcí právní předpis¹⁰⁵.

OVS jsou v rámci informačních systémů veřejné správy povinny uveřejňovat číselníky, pokud jsou správci těchto číselníků a není zákonem stanoveno jinak, a to i způsobem umožňujícím dálkový přístup¹⁰⁶ a předávat ministerstvu údaje do informačního systému o datových prvcích v elektronické podobě, ve formě a s technickými náležitostmi stanovenými prováděcím právním předpisem¹⁰⁷.

OVS jsou v rámci informačních systémů veřejné správy povinny zajistit, aby vazby jimi spravovaného informačního systému na informační systémy jiného správce byly uskutečňovány prostřednictvím referenčního rozhraní s využitím datových prvků vyhlášených ministerstvem a vedených v informačním systému o datových prvcích¹⁰⁸.

OVS jsou povinny zpřístupňovat ministerstvu vnitra v elektronické podobě, ve formě a s technickými náležitostmi stanovenými prováděcím právním předpisem, bez zbytečného odkladu informace o jimi spravovaném informačním systému a jím poskytovaných službách a používaných datových prvcích¹⁰⁹.

OVS vytvářejí a vydávají informační koncepci, uplatňují ji v praxi a vyhodnocují její dodržování. V informační koncepci orgány veřejné správy stanoví své dlouhodobé cíle v oblasti řízení kvality a bezpečnosti spravovaných informačních systémů veřejné správy a vymezí obecné principy pořizování, vytváření a provozování informačních systémů veřejné správy¹¹⁰.

3.3 Otevřená data v České republice

V současné době neexistuje v České republice standardní postup jak zveřejňovat data. Zákon řeší pouze malou část dat, produkovaných veřejnou správou – ani u těch není příliš specifický. Jedná se navíc pouze o zveřejňování dat v nepropojené podobě (1).

¹⁰⁴ § 4 odst. 1. písm. a) zákona 365 /2000 Sb., o informačních systémech veřejné správy.

¹⁰⁵ § 5 odst. 2. písm. c) zákona 365 /2000 Sb., o informačních systémech veřejné správy.

¹⁰⁶ Podle § 2 písm. n) zákona č. 365/2000 Sb., se jedná o přístup do informačního systému prostřednictvím sítě nebo služby elektronických komunikací (například s využitím internetu).

¹⁰⁷ § 5 odst. 2. písm. c) zákona 365 /2000 Sb., o informačních systémech veřejné správy.

¹⁰⁸ § 5 odst. 2. písm. d) zákona 365 /2000 Sb., o informačních systémech veřejné správy.

¹⁰⁹ § 5 odst. 2. písm. e) zákona 365 /2000 Sb., o informačních systémech veřejné správy.

¹¹⁰ § 5a odst. 1. zákona 365 /2000 Sb., o informačních systémech veřejné správy.

První zmínkou o vytvoření standardního postupu je akční plán vlády popsany v kapitole 1.6.1.

3.3.1 Český statistický úřad

Český statistický úřad¹¹¹ (ČSU) patří v České republice k největším producentům dat. Informace publikuje pro veřejnost, nemůžeme však zcela hovořit o otevřených datech, protože nejsou publikována pod otevřenou licenci¹¹².

ČSU je zřízen a pracuje na základě zákona¹¹³ a ve své činnosti se řídí Základním právním rámcem EU v oblasti statistiky¹¹⁴ a kodexem evropské statistiky¹¹⁵ (Code of Practice) přijatý výborem pro statistický systém¹¹⁶ dne 28. září 2011.

V preambuli kodexu se uvádí:

Evropský statistický systém bude předním celosvětovým poskytovatelem statistických informačních služeb a nejvýznamnějším poskytovatelem informací pro Evropskou unii a její členské státy. Evropský statistický systém, založený na vědeckých zásadách a metodách, nabídne a bude neustále zlepšovat program harmonizovaných evropských statistik, který je nezbytným základem pro demokratické procesy a pokrok společnosti.¹¹⁷

ČSU svojí činnost deklaruje jako:

Na základě získaných údajů vytvářet a poskytovat důvěryhodný, konsistentní obraz o stavu a vývoji společnosti v návaznosti na vyvíjející se potřeby všech uživatelů služeb státní statistiky v podmínkách měnícího se prostředí. Za tím účelem koordinovat státní statistickou službu v České republice.¹¹⁸

ČSU provozuje portál „Veřejná databáze ČSU“ dostupný na adrese <http://vdb.czso.cz/>.

Veřejná databáze ČSÚ (VDB) se buduje jako základní a jednotný datový zdroj pro prezentaci statistických údajů určených především pro veřejnost.¹¹⁹

VDB obsahuje pouze agregovaná statistická data, a to za všechny sledované úseky statistiky. Čerpá jednak z výsledků zpracování statistických úloh v ČSÚ, jednak využívá statistických údajů z externích a administrativních zdrojů, zejména z dalších pracovišť státní statistické služby. Neomezuje se jen na data za Českou republiku, k

¹¹¹ <http://www.czso.cz>.

¹¹² Data můžeme zpravidla bezplatně a volně použít při uvedení zdroje.

¹¹³ Zákon č. 89/1995 Sb., o státní statistické službě, ve znění pozdějších předpisů.

¹¹⁴ Článek 338 smlouvy o fungování Evropské unie. Dostupný online v češtině na <http://www.czso.cz/csu/>

redakce.nsf/i/zakladni_pravni_ramec_spolecenstvi_v_oblasti_statistiky.

¹¹⁵ Dostupný online na: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/10425-CS/CZ/10425-CS-CZ.PDF.

¹¹⁶ Zdroj: <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>.

¹¹⁷ Nařízení ES 223/2009, článek 4.

¹¹⁸ Poslání ČSU, dostupné online na

http://www.czso.cz/csu/redakce.nsf/i/mise_vize_strategicke_cile_a_podcile

¹¹⁹ Zdroj: vdb.czso.cz/vdb/vdb.jsp (O Veřejné databázi).

dispozici jsou i data za regiony republiky (kraje, okresy, obce a města apod.), ale také za údaje ze zahraničí.¹²⁰

VDB je provázaná se statistickým metainformačním systémem. Každý údaj je identifikován pomocí metainformací vymezující věcnou, časovou a prostorovou dimenzi. Ke každému údaji bude připojena informace o kvalitě údaje, důvěrnosti, zdroji, kontaktní osobě a bohatá škála dalších, doplňkových informací.¹²¹

VDB zajišťuje ochranu důvěrnosti prezentovaných dat. Je založena na principu jednoty čísla. Údaj se přitom může vyskytovat v různém kontextu v různých výstupních objektech (tabulkách, mapách, grafech apod.).¹²²

VDB je budována jako datové tržiště, které čerpá údaje z databází vznikajících v procesu zpracování statistických údajů. VDB tvoří:

- primární úroveň - obsahuje především databázi agregovaných statistických údajů
- sekundární úroveň - obsahuje výstupní objekty (statistické tabulky, mapy, grafy)
- rozhraní - aplikace zajišťující transformaci a vstup údajů do primární databáze (včetně univerzálního XML rozhraní) a aplikace umožňující přístup uživatelů k datům a výstupním objektům

Jádrem VDB jsou databáze se statistickými ukazateli a jejich hodnotami. Struktura těchto databází je úzce propojená na metainformační systém ČSÚ, především na soustavu číselníků a klasifikací.

ČESKÝ STATISTICKÝ ÚŘAD **Veřejná databáze** Napište nám | Web ČSÚ | English

Úvodní stránka | Navigační strom | Statistické ukazatele | O Veřejné databázi | Časté dotazy | Nápověda

1 Souborné informace
2 Životní prostředí, zemědělství
3 Práce, sociální statistiky
4 Obyvatelstvo, volby
5 Makroekonomika
6 Zahraniční obchod
7 Ceny
8 Průmysl, staveb., energ.
9 Služby

Hledej
 v kódech v textech
 Tabulky | Grafy | Mapy | Nastavit filtry

Nacházíte se: 5 - Makroekonomika / 51 - Finanční hospodaření

Objekty 1 - 3 z 3 Zobrazit stránku: 1 Počet zobrazených objektů na stránce: 20

| Kód | Název | Období | Území | T | G | M |
|----------------|--|-----------|------------------------|---|---|---|
| FIN4010UC_KR * | Příjmy krajů, obcí a dobrovolných svazků obcí | 2007-2010 | Česká republika, Kraje | | | |
| FIN4020UC_KR * | Výdaje krajů, obcí a dobrovolných svazků obcí | 2007-2010 | Česká republika | | | |
| FIN4030PU_KR * | Příjmy a výdaje krajů, obcí a dobrovolných svazků obcí podle krajů | 2007-2010 | Česká republika, Kraje | | | |

VDBVO 1.3.22 © Český statistický úřad Vygenerováno 13.04.2012 10:50

Obrázek 20. Veřejná databáze Českého statistického úřadu

¹²⁰ tamtéž

¹²¹ tamtéž

¹²² tamtéž

Výběr parametrů pro tabulku - Příjmy krajů, obcí a dobrovolných svazků obcí (FIN4010UC_KR)

Provedte výběr hodnot nabízených parametrů, poté potvrďte tlačítkem "Zobrazit tabulku"

Zobrazit tabulku

Aktualizovat

Ukazatel: Příjmy krajů, obcí a dobrovolných svazků obcí

Příjmy krajů, obcí a dobrovolných svazků obcí

Měřicí jednotka: v tis. Kč

Klasifikace územních statistických jednotek:

| Ukazatel | 2007 | 2008 | 2009 |
|---|-------------|-------------|-------------|
| Příjmy celkem | 379 389 226 | 404 608 904 | 401 731 386 |
| Daňové příjmy | 187 430 352 | 204 237 936 | 180 334 759 |
| z toho daně z příjmů fyzických osob | 49 320 423 | 48 358 126 | 41 737 833 |
| z toho: ze závislé činnosti a funkčních požitků | 38 733 002 | 36 417 764 | 34 139 012 |
| ze samostatné výdělečné činnosti | 8 105 636 | 8 856 172 | 4 583 246 |
| z kapitálových výnosů | 2 472 743 | 3 077 886 | 3 013 914 |
| daně z příjmů právnických osob | 51 668 473 | 60 447 069 | 43 805 594 |
| z toho: daň z příjmů právnických osob | 45 218 590 | 53 867 205 | 37 739 801 |
| daň z příjmů právnických osob za obce | 6 182 880 | 6 026 856 | 5 764 358 |
| daň z přidané hodnoty | 68 413 875 | 76 973 414 | 76 009 061 |
| daně a poplatky z vybraných činností a služeb | 13 059 795 | 13 354 043 | 12 454 514 |
| z toho: poplatky a odvody v oblasti životního prostředí | 5 771 567 | 5 819 342 | 6 100 102 |
| místní poplatky z vybraných činností a služeb | 2 609 373 | 2 715 631 | 2 487 660 |
| správní poplatky | 3 762 181 | 3 790 217 | 2 843 143 |
| daň z nemovitostí | 4 962 376 | 5 098 966 | 6 324 951 |
| Nedaňové příjmy | 28 156 275 | 30 164 843 | 31 165 331 |
| z toho: příjmy z vlastní činnosti | 6 064 172 | 6 197 281 | 6 179 125 |
| odvody příspěvkových organizací | 1 720 242 | 1 647 469 | 2 225 728 |
| příjmy z pronájmu majetku | 9 757 673 | 10 134 520 | 10 527 796 |
| příjmy z úroků | 1 993 829 | 3 008 196 | 1 680 735 |

Obrázek 21. Veřejná databáze Českého statistického úřadu - příjmy územních celků

Metainformace k hodnotě

| Položka | Hodnota |
|--------------------|------------------------------------|
| Hodnota: | 428 734 485 |
| Ukazatel: | Příjmy krajů a obcí celkem |
| Druhá specifikace: | Ocenění: Běžná cena |
| Měřicí jednotka: | tisíc korun českých |
| Typ údaje: | Definitivní údaj bez další korekce |
| Území: | Česká republika |
| Období: | Rok 2010 |
| Zveřejnění: | 04.11.2011 - 03.11.2061 |
| Zdroj hodnoty: | MF ČR |
| Důvěrnost údaje: | Veřejný |

Zkrácená verze

Zavít

Obrázek 22. Veřejná databáze Českého statistického úřadu - metadata

Ke dni 13. 4. 2012 publikovala veřejná databáze 28 oblastí v 9 tematických okruzích. Celkem bylo zveřejněno 1800 tabulek. Vedle tabulek jsou informace zpracovány i jako grafy a mapové podklady. Seznam tematických okruhů a oblastí je uveden v příloze.

Na portálu je rovněž zveřejněn výsledek Sčítání lidu, domů a bytů 2011. Údaje jsou k dispozici ve formě html tabulek s metadaty nebo jako sešity MS Excel. Portál navštívilo v roce 2011 600 tisíc návštěvníků¹²³ (bez web robotů).

¹²³ <http://vdb.czso.cz/vdb/webstats.jsp?bot=0>.

3.3.2 RÚIAN – Registr územní identifikace, adres a nemovitostí

Je jediný základní registr, do kterého je umožněn veřejný bezplatný přístup¹²⁴, a to na základě zákona¹²⁵. Přístup se provádí pomocí aplikace VDP¹²⁶ přístupnou na adrese vdp.cuzk.cz/. Aplikaci provozuje ČÚZK¹²⁷.

Aplikace pracuje s těmito daty:

- Registru územní identifikace, adres a nemovitostí
- Informačního systému územní identifikace
- Informačního systému katastru nemovitostí

Aplikace poskytuje nahlédnutí k referenčním údajům v těchto kategoriích:

- Vyhledání územního prvku
- Ověření adresy
- Zobrazení mapy
- Výměnný formát RÚIAN/ISUI (změnové věty)

Aplikace VDP má zcela zásadní význam pro otevřená-propojená data, a to proto, že každému adresnému místu v ČR přiřazuje jedinečný trvalý identifikátor. U vyřazených adres umožní zobrazit historická data.

| Kód kraje (VÚSC) | Název kraje (VÚSC) | Název regionu soudržnosti | Detail |
|------------------|----------------------|---------------------------------|--------|
| 19 | Hlavní město Praha | Praha | |
| 35 | Jihočeský kraj | Jihozápad | |
| 116 | Jihomoravský kraj | Jihovýchod | |
| 51 | Karlovarský kraj | Severozápad | |
| 108 | Kraj Vysočina | Jihovýchod | |
| 86 | Královéhradecký kraj | Severovýchod | |
| 78 | Liberecký kraj | Severovýchod | |
| 132 | Moravskoslezský kraj | Moravskoslezsko | |
| 124 | Olomoucký kraj | Střední Morava | |
| 94 | Pardubický kraj | Severovýchod | |
| 43 | Plzeňský kraj | Jihozápad | |
| 27 | Středočeský kraj | Střední Čechy | |
| 60 | Ústecký kraj | Severozápad | |
| 141 | Zlínský kraj | Střední Morava | |

Obrázek 23. Registr územních identifikátorů - veřejný dálkový přístup

¹²⁴ Přístup bude umožněn po spuštění 1. 7. 2012

¹²⁵ §47 zákona 111/2009 Sb., o základních registrech.

¹²⁶ Veřejný Dálkový Přístup k datům základního registru RÚIAN/ISUI

¹²⁷ Český úřad zeměměřičský a katastrální, <http://www.cuzk.cz>

VDP umožňuje vedle vyhledání územních prvků zobrazit také jejich detail, a popis. Ukázka zobrazení detailu města Ostrava je uveden na následujícím obrázku.

Obrázek 24. Registr územních identifikátorů - veřejný dálkový přístup - Detail objektu

3.3.3 OpenData.CZ - Iniciativa za otevřenou veřejnou správu

V roce 2010 v České republice vznikla „Iniciativa za otevřenou veřejnou správu“. Iniciativu můžeme pojmut jako nezávislé sdružení odborníků s cílem prosadit myšlenku a principy otevřených dat v České republice. Na svém webu vymezují svojí činnost takto:

„Naším cílem je vybudování otevřené datové infrastruktury v ČR. Ta bude poskytovat veřejná data komukoliv a kdykoliv a umožní data volně kombinovat. Na základě snadného přístupu k datům může odborná veřejnost svobodně a nezávisle tvořit aplikace ať pro jiné odborníky tak pro laickou veřejnost. Příkladem může být aplikace Rozklikávací rozpočet.“¹²⁸



Obrázek 25. Logo iniciativy OpenData.cz

3.3.4 Příklady existujících aplikací

Na základě už dnes dostupných dat vzniklo několik aplikací ilustrujících možnosti, které otevřená data přinesou. Stát přitom nestál ani korunu. Po „otevření“ dat veřejnosti lze očekávat jejich prudký nárůst. (1)

¹²⁸ Zdroj: <http://www.opendata.cz>, 10. 3. 2012.

- BudováníStátu.cz – Michal Škop – grafické přehledy o hospodaření státu a utrácení veřejných prostředků státního rozpočtu, rozklikávací rozpočet ČR
- VášMajetek.cz – Jiří Skuhrovec – dražby, aukce a prodej majetku státu, krajů, měst a obcí a exekučně zabaveného majetku na jednom místě
- MapaZakázek.cz – Pavel Nohejl – databáze veřejných zakázek, propojená s obchodním rejstříkem a mapovými podklady. 40 tisíc veřejných zakázek v databázi, 13.500 dodavatelů, 5.000 zadavatelů, mezi nimiž je možné vyhledávat podle rozličných kritérií, zadavatele, dodavatele, ceny zakázek atd.

3.4 Otevřená data ve světě

Ve světě jsou budovány datové katalogy na různých úrovních veřejné správy a jsou pro ně využívány různé nástroje. Tato kapitola popisuje základní východiska, praktiky a přístupy publikace otevřených dat a jejich katalogizaci ve vybraných zemích. Oblast publikace a katalogizace prochází vývojem, a proto je cílem této kapitoly poskytnout základní a ucelený přehled.

Aktivity a iniciativy orientované na otevřená data veřejné správy, a s nimi spojené datové katalogy, můžeme z hlediska zaměření rozdělit na lokální, národní a mezinárodní. Lokální datové katalogy se zaměřují na otevřená data měst či regionů, mezi které lze počítat i samostatné územní celky států s federativním zřízením. Národní datové katalogy se zaměřují na otevřená data z celého státu. Nadnárodní datové katalogy se pak zaměřují na otevřená data pocházející z více států. (14)

3.4.1 Situace ve Spojených Státech Amerických

Spojené Státy Americké patří mezi průkopníky publikace otevřených dat ve světě. Projekt formálně odstartoval 44. prezident USA Barack Obama memorandem¹²⁹ „Transparency and Open Government“, které bylo v roce 2009 adresováno všem vedoucím organizačních složek v zemi a zveřejněno ve Federálním registru US¹³⁰. V memorandu se uvádí tyto politické cíle:

- Government should be transparent
- Government should be participatory
- Government should be collaborative

Administrativa prezidenta Obamy vydala v této souvislosti nařízení „Open Government Directive“¹³¹, které konkrétně popisuje, jaké kroky budou učiněny. Součástí tohoto nařízení je i „Open Government Plan“. V nařízení jsou zmíněny kapitoly:

- Publish Government Information Online
- Improve the Quality of Government Information
- Create and Institutionalize a Culture of Open Government
- Create an Enabling Policy Framework for Open Government

Otevřenost je realizována jednotlivými ministerstvy a úřady. V dokumentu „Open Government Directive: Senior Accountable Officials (SAOs)“¹³² je uveden

¹²⁹ Memorandum for the heads of executive departments and agencies. Dostupné online na http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government/

¹³⁰ Federal Register /Vol. 74, No. 97 /Thursday, May 21, 2009 /Notices. Dostupné online na adrese www.fas.org/sgp/news/2009/05/opengov.pdf nebo www.fas.org/sgp/crs/secretary/R41361.pdf

¹³¹ <http://www.whitehouse.gov/open/documents/open-government-directive>

¹³² Dostupné online na <http://www.whitehouse.gov/open/documents/open-government-directive/officials>

seznam všech přímo zainteresovaných složek. Ministerstvem spravedlnosti US je např. otevřenost dat podpořena materiálem „The Freedom Information Act (FOIA)¹³³.

Prezident Obama bere otevřená data velmi vážně a zmiňuje je i ve svých věřených vystoupeních.

*The way to make government responsible is to hold it accountable. And the way to make government accountable is make it transparent so that the American people can know exactly what decisions are being made, how they're being made, and whether their interests are being well served.*¹³⁴

*My administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government*¹³⁵.

Zároveň je však potřeba zmínit, že i přes formálnost prohlášení a nařízení, tak i aktivní zapojení administrativy na všech úrovních, je iniciativa právně nevymahatelná a stojí na aktivním zapojení a dostatku volných financí. Vyloučení právní vymahatelnosti je například uvedeno přímo v klíčovém memorandu¹³⁶:

This memorandum is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by a party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

Portál otevřených dat je v USA umístěn na adrese data.gov. Vedle federálního portálu otevřených dat jsou v provozu portály jednotlivých států¹³⁷, měst¹³⁸ a vládních institucí¹³⁹.

¹³³ Dostupné online na <http://www.justice.gov/ag/foia-memo-march2009.pdf>.

¹³⁴ Barack Obama, 21. Leden 2009.

¹³⁵ Zdroj: <http://www.whitehouse.gov/open>.

¹³⁶ Memorandum for the heads of executive departments and agencies. Dostupné online na http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government/

¹³⁷ Ke dni 16.4. 2012 bylo v provozu 31 portálů států U.S.

¹³⁸ Ke dni 16.4. 2012 bylo v provozu 15 portálů měst U.S.

¹³⁹ Ke dni 16. 4. 2012 bylo v provozu 172 portálů vládních institucí v USA.

Raw Data

Use the catalog below to access U.S. Federal Executive Branch datasets. Click on the name of a dataset to view additional metadata for that dataset. By accessing the data catalogs, you agree to the Data Policy. The Raw Data catalog provides an instant download of machine readable, platform-independent datasets.

« Go back to the main data catalog.



Search

View Types

- Datasets
- External Datasets
- Files and Documents
- Filtered Views
- Charts
- Maps
- Calendars
- Forms

Filter By

- Raw Data
- Data Extraction Tools
- RSS Feeds
- Widgets and Gadgets

Agency

- Broadcasting Board of Governors
- Commodity Futures Trading Commission
- Corporation for National and Community Service
- Court Services and Offenders

Browse Raw Datasets

Most Relevant

| Name | Popularity | Type |
|--|---------------|------|
| 1. Worldwide M1+ Earthquakes, Past 7 Days Geography and Environment ANSS, geologist, plate, real time, environment, ... Real-time, worldwide earthquake list for the past 7 days | 165,132 views | CSV |
| 2. Federal Data Center Consolidation Initiative (FDCCI) Data Center Closings 2010-2013 Federal Government Finances and Employment fddoi, ... Federal Data Center Consolidation Initiative (FDCCI) Data Center Closings | 30,546 views | XML |
| 3. TSCA Inventory Geography and Environment new chemicals, manufactured chemicals, ... This dataset consists of the non confidential identities of chemical | 24,246 views | CSV |
| 4. US DOE/NNSA Response to 2011 Fukushima Incident: Radiological Air Samples Geography and Environment radiation, Japan, nuclear, Tohoku, ... Field Samples are physical media collected during the response which are | 21,627 views | CSV |
| 5. Data.gov Catalog Other dataset, metadata, catalog, data extraction tool, ... An interactive dataset containing the metadata for the Data.gov raw datasets | 20,521 views | XML |
| 6. US DOE/NNSA Response to 2011 Fukushima Incident: Field Team Radiological Measurements Geography and Environment Japan, nuclear, Tohoku, radiation, ... Field Measurements describe α and β activity and γ | 20,178 views | CSV |
| 7. Federal Executive Branch Internet Domains Federal Government Finances and Employment .gov, domains, agencies, federal, registered | 16,263 views | XML |

Obrázek 26. Portál otevřených dat USA - data.gov

V dokumentu „Data.gov Concept of Operations“ je velmi podrobně rozpracována strategie pro realizaci otevřených dat¹⁴⁰.

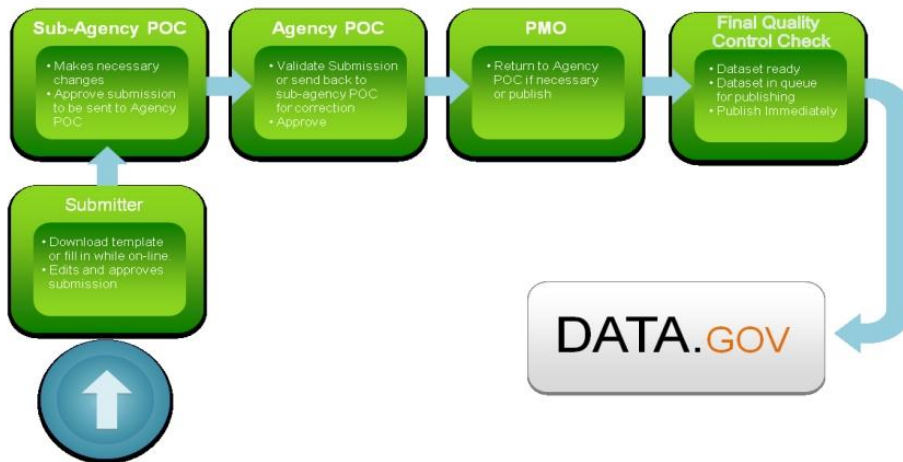
Nejvýznamnější agenturou podporující myšlenku otevřených dat je NASA, její portál nalezneme na adrese <http://open.nasa.gov>. Na portálu nalezneme nejenom publikovaná data, ale i metodiku a plán podpory myšlenky otevřených dat na úrovni agentury. Mezi vědecko-výzkumné instituce můžeme např. zařadit iniciativu Linking Open Government Data, která působí na Rensselaer polytechnic institute v New York USA. Informace o iniciativě nalezneme na adrese logd.tw.rpi.edu.

Technická realizace

Data jsou publikovaná pod otevřenou licenci, většinou se jedná o Creative commons. Odkazy na data jsou uvedeny v katalogu otevřených dat – data.gov. Proces uveřejnění dat je následující:

¹⁴⁰ Dostupné online na http://www.data.gov/documents/data_gov_conops_v1.0.pdf.

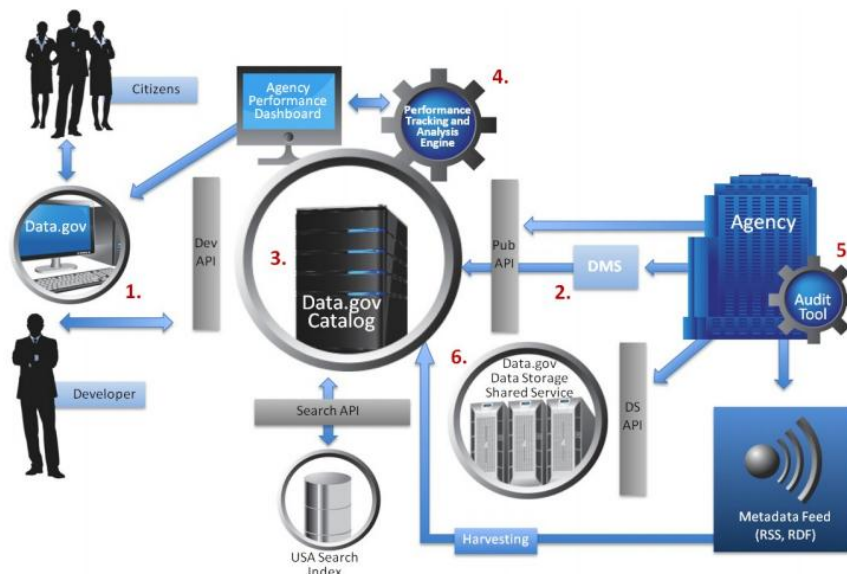
Process for Dataset Management System (DMS)



Obrázek 27. Proces publikace otevřených dat v USA

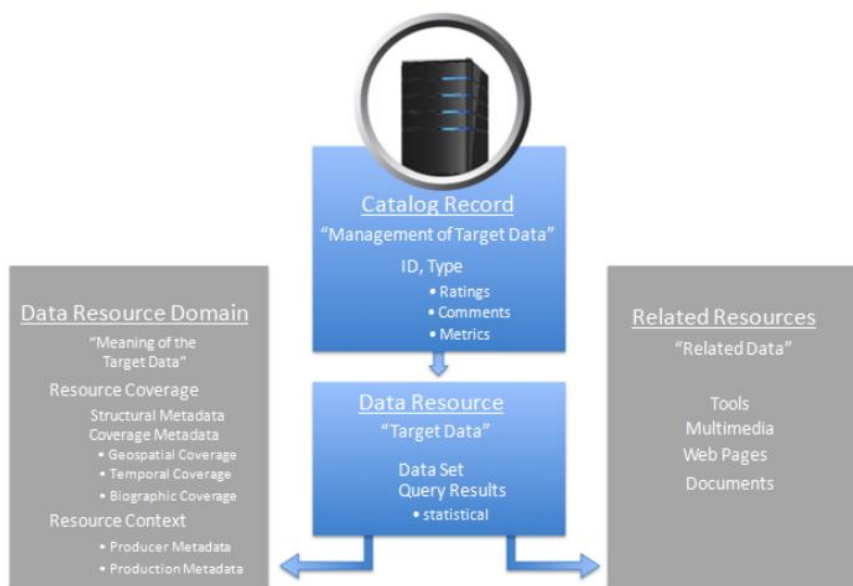
Proces (schvalování) publikace otevřených dat v USA je řešen vertikálně na několika úrovních, nejprve dojde k návrhu na publikaci v místě vzniku dat a následně jsou postupně schvalovány a redigovány nadřazenými subjekty až nakonec schváleny a zveřejněny v katalogu otevřených dat – data.gov. Tento proces je náročný na množství zainteresovaných subjektů a finance. Na druhou stranu zajišťuje, že v místě vzniku dat jsou lidmi, kteří (svým) datům rozumí, správně popsána z hlediska jejich oblasti a následně metodicky redigována lidmi, kteří spravují celé kategorie nebo katalog.

Konceptuální architekturu otevřených dat v USA zobrazuje následující schéma. Ze schématu je patrné, že jednotlivé „bloky“ jsou od sebe odděleny a komunikují mezi sebou skrze popsané rozhraní.



Obrázek 28. Konceptuální architektura data.gov

Struktura datasetu používaných na data.gov je zobrazena na diagramu. Detailní popis je přístupný v publikaci (16).



Obrázek 29. Struktura datasetu katalogu dat data.gov

3.4.2 Situace ve Velké Británii

Velká Británie přijala myšlenku otevřených dat jako druhá země světa, jen pár měsíců po USA. Britský katalog otevřených dat dostupný na adrese <http://data.gov.uk> je v provozu od ledna 2010. K dubnu 2012 se na něm nachází 5400 datasetů. Proces oficiálně zahájil v květnu 2010 premiér David Cameron úředním dopisem „Letter to Government departments on opening up data“¹⁴¹. V dopise se přímo uvádí:

Greater transparency across Government is at the heart of our shared commitment to enable the public to hold politicians and public bodies to account; to reduce the deficit and deliver better value for money in public spending; and to realise significant economic benefits by enabling businesses and non-profit organisations to build innovative applications and websites using public data.

The Government must set new standards for transparency, and our Coalition Programme for Government sets out a number of specific commitments. The Government's initial transparency commitments are set out below, alongside deadlines for publication. Limited exemptions on national security and personal privacy grounds will be permitted.

Následně tento krok Úřad vlády hodnotí v dokumentu „Making Open Data Real: A Public Consultation“ vydaným v srpnu 2011¹⁴². Britský premiér David Cameron propagoval myšlenku otevřených dat na prestižní konferenci TED¹⁴³ v únoru 2010¹⁴⁴

¹⁴¹ <http://www.number10.gov.uk/news/letter-to-government-departments-on-opening-up-data/>.

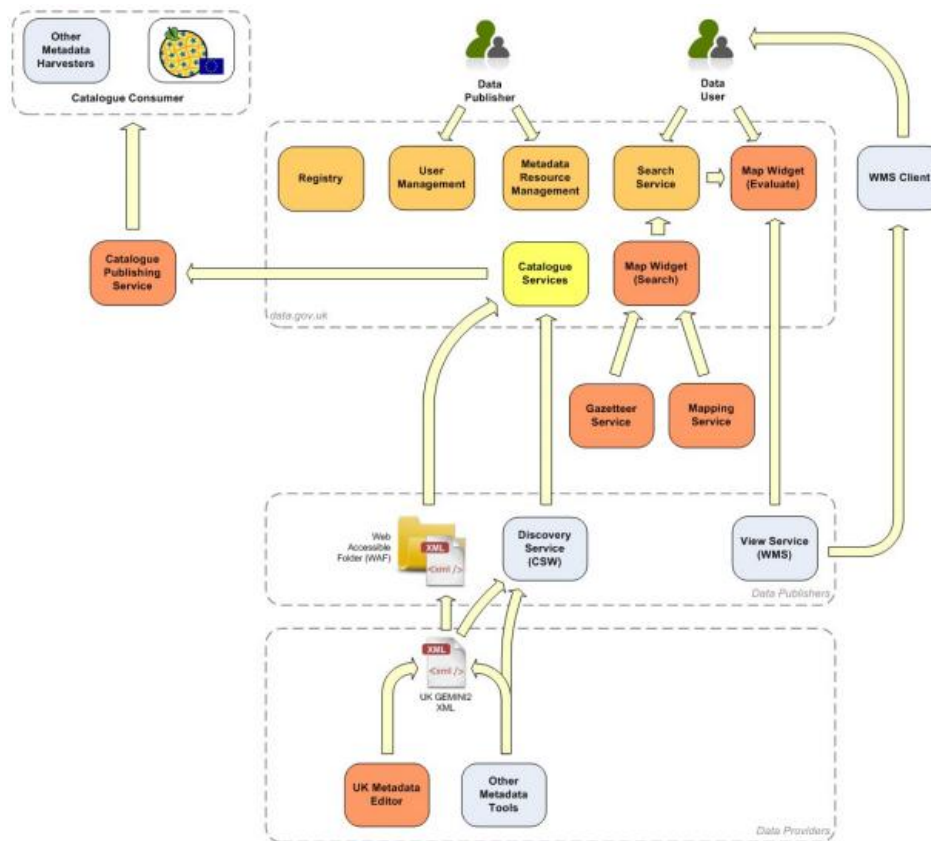
¹⁴² <http://www.cabinetoffice.gov.uk/sites/default/files/resources/open-data-consultation.pdf>

¹⁴³ Mezinárodní konference, <http://www.ted.com>.

¹⁴⁴ Premiérem UK se stal o tři měsíce později.

svým vystoupením „The Next Age of Government“¹⁴⁵. Proces otevřených dat v UK zaštituje jako poradce zakladatel W3C Tim Berners-Lee.

Architektura publikace otevřených dat ve Velké Británii je vyobrazena na následujícím schématu. Data.gov.uk je servisně orientovaný. Detailní popis struktury a služeb data.gov.uk je v publikaci (17).



Obrázek 30. Architektura publikace dat ve Velké Británii

3.4.3 Situace v Evropské unii

Hlavním cílem strategie EU 2020 je dosažení vysokého a udržitelného růstu evropské ekonomiky. Za tímto účelem bude muset Evropa posílit svůj inovační potenciál a využívat své zdroje co nejlepším způsobem. Jedním z těchto zdrojů jsou veřejné informace. Tyto informace mají významný – v současné době nevyužitý – potenciál pro opakované použití pro nové produkty a služby a zvýšení účinnosti správy. Zpřístupnění tohoto zdroje by mohlo hospodářství EU přinést zisk až 40 miliard EUR ročně a zároveň posílí účast občanů na politickém a společenském životě¹⁴⁶.

Místopředsedkyně Evropské komise pro Digitální agendu Neelie Croes na konferenci Lift-Off towards Open Government konané 15. prosince 2010 v Bruselu prohlásila:

¹⁴⁵ http://www.ted.com/talks/david_cameron.html.

¹⁴⁶ Zdroj: Sdělení Komise Evropskému parlamentu, Radě, Evropskému hospodářskému a sociálnímu výboru a Výboru regionů – Veřejně přístupné údaje. V Bruselu 12.12 2011. KOM(2011) 882 v konečném znění.

*Také the example of public sector information – possibly a €30 billion market in Europe. I have said it before, and I say it again: YES TO OPEN DATA!*¹⁴⁷

Hlavním portálem pro otevřená data na úrovni Evropské unie je Europe's public data, dostupný na adrese <http://www.publicdata.eu>.

The screenshot shows the publicdata.eu website interface. At the top, there is a search bar and navigation links for 'Search', 'Groups', 'About', 'Apps', and 'Ideas'. The main heading is 'Airport traffic Czech republic - Výkony letišť ČR'. Below this, there is a 'Resources' section with a list of XLS files for each year from 2002 to 2010. To the right, there is a 'Metadata source' section with details about the dataset, including the country 'Czech Republic' and various tags. At the bottom right, there is an 'Other access' section listing JSON and RDF/XML formats.

| Field | Value |
|----------------|---|
| Source | https://www.zydos.cz/os/ctvlepr.htm |
| Author | Author not given |
| Maintainer | Maintainer not given |
| Version | 1.0 |
| Country | Czech Republic |
| Harvest Source | Dataset page on CKAN Czech Republic |
| Frequency | Yearly |
| Language | Czech |

Obrázek 31. Portál otevřených dat EU - publicdata.eu

Oficiálním projektem EU je ESD-toolkit dostupný na adrese <http://www.esd-toolkit.org.uk>. Projekt je podrobněji popásán v samostatné kapitole.

V následujícím seznamu jsou uvedeny významnější projekty realizované za podpory grantů EU:

- Planet Data, www.planet-data.eu
- Linked open data around the clock (LATC), www.latc-project.eu
- Linking Open Data (LOD2), www.lod2.eu
- TaxonConcept, www.taxonconcept.org

Vzhledem k omezenému rozsahu práce nebudu projekty blíže popisovat.

3.4.4 Effective Service Delivery - ESD toolkit

Effective Service Delivery (ESD) toolkit je komplexní soubor osvědčených nástrojů, modelů, případových studií a doporučení, z nichž část je volně dostupná. ESD je určena poskytovatelům veřejných služeb za účelem zvýšení jejich účinnosti, produktivity a účelnosti. ESD definuje řadu redigovaných slovníků a obchodních

¹⁴⁷ Zdroj: <http://europa.eu/rapid/pressReleasesAction.do?reference=speech/10/752>

modelů popisujících veřejnou správu¹⁴⁸. Nejzajímavější model je „Local Government Bussiness Model¹⁴⁹“ zobrazený na následujícím diagramu.

Diagram souhrnně nazvaný jako Obchodní model místní samosprávy definuje prvky veřejné správy v oblasti poskytování služeb na místní úrovni. Model zahrnuje soubor kontrolovaných seznamů a vztahů mezi nimi. Tento model může být vnímán jako ontologie.

Model má tři části:

- Lidé a místa (People and places)
- Organizační rámec (Organisation scope)
- Organizace (Organisation)

Pro příklad ustanovení zákona (legislation) definuje povinnost (Power/Duty). Povinnost je vykonávána/poskytována službou (Service). Z druhé strany občan se nachází v životní situaci (Life event), která mění jeho životní stav (Circumstance), a ten vyvolává potřebu (Need). Potřeba je řešena službou.

Např. Ztratil jsem práci – Jsem nezaměstnaný – Potřebuji najít práci – úřad práce poskytne službu – ta je poskytována ze zmocnění – zmocnění vychází ze zákona. Konkrétně si situaci v různých počátečních variantách můžeme představit takto:

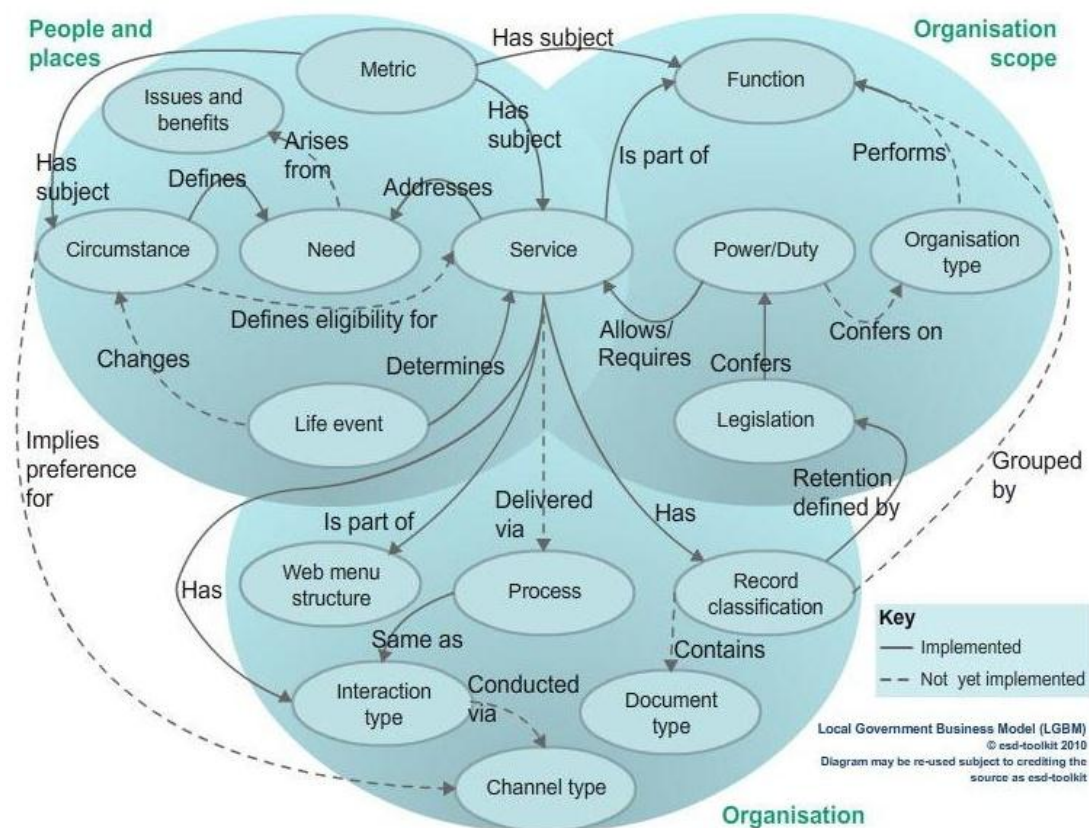
| Stav | Situace |
|-------------------------|---|
| Životní situace | <ul style="list-style-type: none"> - Ztratil jsem práci - Ukončil jsem pracovní poměr - Zaměstnavatel ukončil činnost - Dostudoval jsem - Ukončila jsem mateřskou - Změna bydliště (a ukončení stávající práce) |
| Důsledek (Životní stav) | - Jsem nezaměstnaný |
| Potřeba | - Potřebuji práci |
| Služba | - Poskytnutí pomoci při hledání práce, např. pomocí Úřadu práce |
| Povinnost | - Zprostředkování zaměstnání ¹⁵⁰ |
| Zákon | - 435/2004 Sb., zákon o zaměstnanosti |

Tabulka 9. Ukázka životní situace

¹⁴⁸ Zdroj: standards.esd.org.uk.

¹⁴⁹ Dostupné na standards.esd.org.uk/LGBMDiagram.aspx.

¹⁵⁰ Dle § 14 zákona č. 435/2004 Sb., o zaměstnanosti.



Obrázek 32. Obchodní model vlády podle ESD toolkitu

3.4.5 Organizace Spojených národů, Světová banka a OECD

Portál otevřených dat OSN se nachází na adrese <http://data.un.org>. Ke dni 13. 4. 2012 obsahoval 24 datasetů s více než 60 miliony záznamů. Mezi zařazená data patří i data z České republiky. Otevřená data z České republiky jsou spravovány ČSU¹⁵¹. Podrobnější popis otevřených dat ČSU je uveden v kapitole 3.3.1.

Světová banka má portál svých otevřených dat umístěný na adrese <http://data.worldbank.org>. OECD má portál svých otevřených dat umístěný na adrese <http://stats.oecd.org>.

3.4.6 Aktuální situace v Rakousku

V Rakousku jsou otevřená data implementována na dvou úrovních¹⁵²:

- Města, jako hlavní iniciátoři myšlenky otevřených dat (Vídeň a Linz již provozují své internetové portály s otevřenými daty, v Salzburgu a Grazi na portálech pracují).
- Téma je prioritou pro rakouské federální kancléřství, které ohlásilo prototypovou verzi vládního portálu otevřených dat na duben 2012 a plně funkční verzi na červen 2012. Tento prototyp bude oficiálně představen na konferenci OGD 2012¹⁵³.

¹⁵¹ <http://www.czso.cz>.

¹⁵² Zdroj: Seminář Otevřená data a veřejná správa: rakouské zkušenosti a jejich využití v ČR, 28. 2. 2012.

¹⁵³ <http://www.ogd2012.at>.

V Rakousku také stále roste komunita okolo problematiky otevřených dat. Patří do ní politici, úředníci, občané, ale i průmysl a podnikatelé. Ti se pravidelně setkávají a diskutují své potřeby, požadavky a také další kroky v oblasti otevřených dat. Velmi aktivní politik v oblasti otevřených dat je např. člen Federal Council Marco Schreuder.

Stále je však v Rakousku co zlepšovat. Chybí totiž zatím oficiální strategie v uplatňování otevřených dat z nejvyšších politických míst¹⁵⁴.

Významné zdroje otevřených dat v Rakousku jsou tyto portály:

- Národní portál: <http://gov.opendata.at/>.
- Portál hlavního města Vídně: <http://data.wien.gv.at/>.
- Portál města Lince: <http://data.linz.gv.at/>.

3.4.7 Ostatní státy podporující myšlenku otevřených dat

Vzhledem k omezenému rozsahu práce nebudu blíže popisovat portály ostatních zemí světa, které podporují myšlenku otevřených dat. Portály jednotlivých zemí jsou uvedeny v příloze. Na jednotlivých portálech najdeme nejenom publikována otevřená data dané země, ale zpravidla i metodiku zveřejňování, doporučení, „best practice“ a výhled do budoucna.

V příloze je rovněž uveden seznam dosud vydaných aplikací, které pracují s otevřenými daty.

3.5 Nástroje pro katalogizaci otevřených dat

3.5.1 Comprehensive Knowledge Archive Network

Comprehensive Knowledge Archive Network (CKAN) je software pro vytváření a správu datového katalogu. Software CKAN je navržen tak, aby s ním mohli pracovat jak lidé, tak i další programy a aplikace. Za tímto účelem CKAN disponuje webovým rozhraním pro interakci s člověkem a aplikačním rozhraním (API)¹⁵⁵.

Bližší informace o CKAN je možné nalézt na webových stránkách věnovaných tomuto softwaru (ckan.org) nebo na wiki (wiki.ckan.org).

Open Knowledge Foundation provozuje instanci CKAN také pro Českou republiku. Ta je dostupná na adrese cz.ckan.net.

Záznamy o datech (datových množinách) lze do CKAN vkládat, upravovat a lze v nich také vyhledávat. Datové množiny v datovém katalogu lze kategorizovat pomocí skupin a štítků (tagů). Uživatelé mohou k záznamům o datových množinách (datasetech) přidávat svoje komentáře a mohou také datasety hodnotit.

¹⁵⁴ Zdroj: Seminář Otevřená data a veřejná správa: rakouské zkušenosti a jejich využití v ČR, 28. 2. 2012.

¹⁵⁵ Zdroj: <http://www.opendata.cz/>.

Obrázek 33. Služba CKAN Czech Republic

CKAN DataSet

Základní jednotkou pro popis dat v software CKAN je tzv. dataset¹⁵⁶. Tento dataset obsahuje odkaz na datové zdroje (jejich URI), prostřednictvím kterých lze získat datové množiny, a dále je v datasetu obsažen také popis těchto datových množin.

Každý CKAN dataset má několik atributů, jejichž hodnoty je možné nastavit během zakládání CKAN datasetu, nebo upravit u již existujících CKAN datasetů.

CKAN datasety se vytvářejí pro datové množiny, které mají dobře strojově zpracovatelnou formu. Například se jedná o soubory tabulkového procesoru (MS Excel - xls, xlsx apod.), XML soubory, tabulková data v souborech oddělených čárkou či jiným oddělovačem (csv).

CKAN dataset je možné založit i pro službu (aplikaci), prostřednictvím které lze data získat. Jako zdroj je pak uvedeno API služby. Do CKAN se nekládají samotné datové soubory, ale pouze odkazy na ně. Odkaz na stažitelný soubor nebo na API je v terminologii CKAN označován jako "zdroj".

3.5.2 Socrata

Socrata je platforma pro správu a publikaci dat veřejné správy, která je vyvíjená a poskytována společností Socrata Inc. se sídlem v USA. Jedná se o proprietární řešení. Socrata je nejrozšířenějším nástrojem pro otevřená data v rámci vlády USA.

¹⁵⁶ Pro účely odlišení od obecného pojmenování datové množiny se v metodice CKAN používá označení CKAN dataset.

Nástroj Socrata je orientován nejen na katalogizaci vlastních dat a datových zdrojů, ale také na prezentaci obsahu dat. Nabízí tak rozsáhlé funkce, které umožňují uživatelům procházet a analyzovat obsah katalogizovaných datových zdrojů. Subjekty publikující data také mohou uživatelům nabídnout například vizualizaci dat na mapovém podkladě. Stejně jako CKAN i Socrata disponuje aplikačním rozhraním pro interakci s dalšími aplikacemi. Socrata využívá model Software-as-a-Service (SaaS, software jako služba) a primárně je dostupná jako služba placená (lze si vybrat z několika cenových tarifů, pro údaje o cenách je třeba kontaktovat představitele společnosti Socrata Inc.). K dispozici je i služba zdarma, která je ale omezena z hlediska funkcí, objemu spravovaných dat i tím, že obsah katalogu může spravovat a vytvářet pouze jeden uživatel. (14)

Více informací o nástroji Socrata získáme z adresy <http://www.socrata.com/>.

3.6 Existující ontologie a slovníky

Pro popis entit veřejné správy ČR je nezbytné sestavení systému ontologii, které budou poskytovat meta-model. Je vhodné, aby tento systém ontologii byl v co největší míře provázán s již existujícími ontologiemi, a to za účelem snadnějšího propojení a přenosu informací.

3.6.1 Core vocabularies

V rámci komise EU vnikla iniciativa joinup¹⁵⁷, jejíž hlavním posláním je vytvoření základního názvosloví, které by mělo umožnit opakované použití a sdílení dat veřejné správy napříč rezorty i hranicemi.

Konkrétně se jedná o tři ontologie s označením osoby, povinnosti a polohy. Každý tento slovník obsahuje určité charakteristiky, které pokud by byly zařazeny do příslušného katalogu veřejné správy, umožní sjednocený popis příslušného objektu. Například je tak možné jednotně popsat osobu, včetně data a místa narození, pohlaví ale i dalších skutečností, jako zda se daná osoba zúčastnila voleb, byla někdy nezaměstnaná, nemocná, atp.¹⁵⁸

Slovníky byly ve své základní podobě sestaveny třemi pracovními skupinami, do nichž bylo zapojeno celkem 67 expertů. V každé skupině byli specialisté z Evropské komise, mezinárodních normativních úřadů, akademické obce a experti z 21 států Evropské unie.

Významným prvkem projektu je skutečnost, že je zapojen do projektu EU – ISA (Interoperability Solutions For European Public Administrations¹⁵⁹). To mimo jiné znamená, že do projektu jsou zapojeny účastníci z většiny zemí EU a to včetně veřejnosti, která může celý vývoj připomínkovat.

Projekt se zabývá popisem tří domén:

- Fyzická osoba¹⁶⁰
- Právnícká osoba¹⁶¹
- Poloha¹⁶²

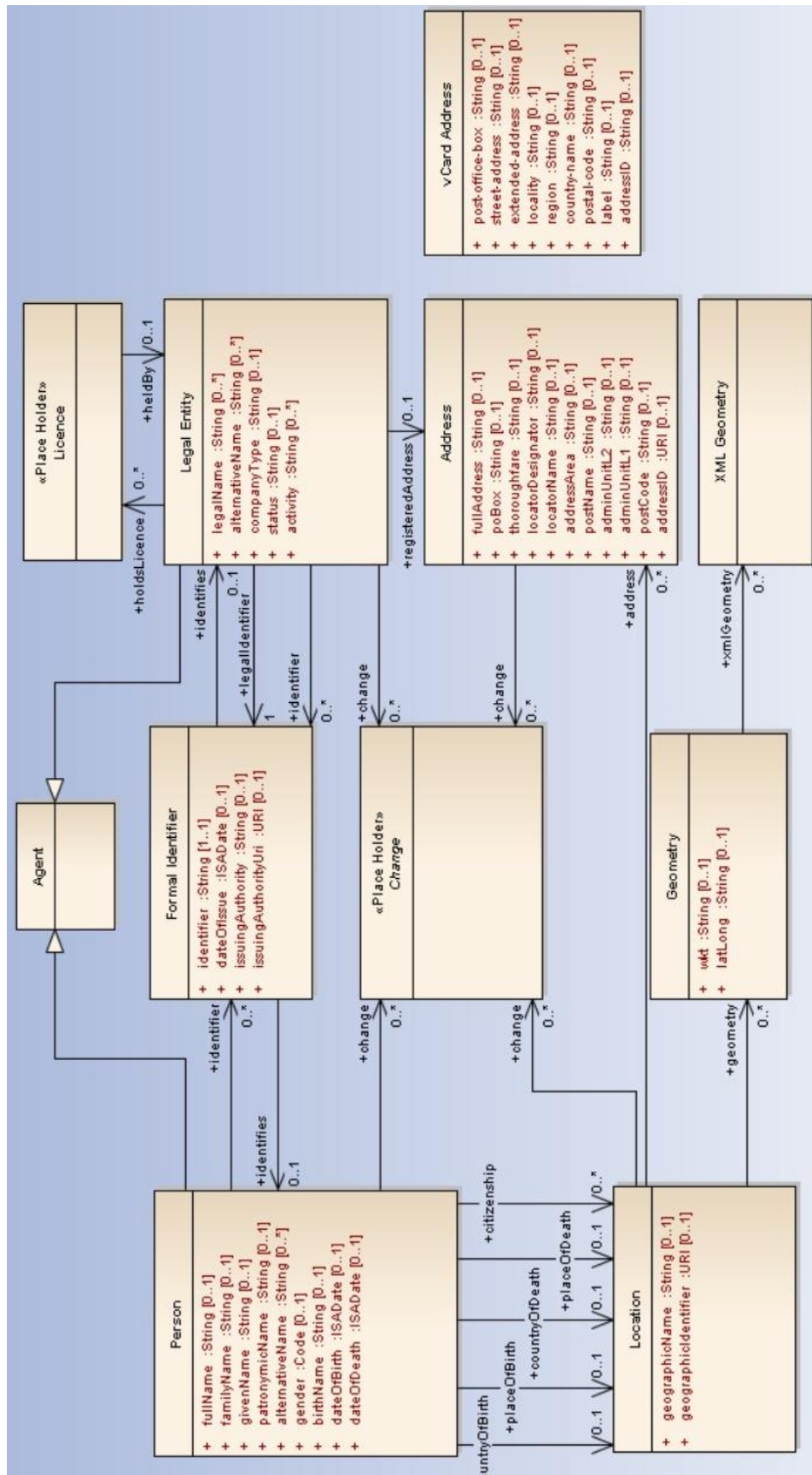
¹⁵⁷ joinup.ec.europa.eu.

¹⁵⁸ Zdroj: Časopis eGovernment – elektronizace veřejné správy č. 1/2012, dostupné na www.egovernment.cz.

¹⁵⁹ Více informací na adrese ec.europa.eu/isa/.

¹⁶⁰ https://joinup.ec.europa.eu/asset/core_person/description.

¹⁶¹ https://joinup.ec.europa.eu/asset/core_business/description.



Obrázek 34. Diagram Core Vocabularies

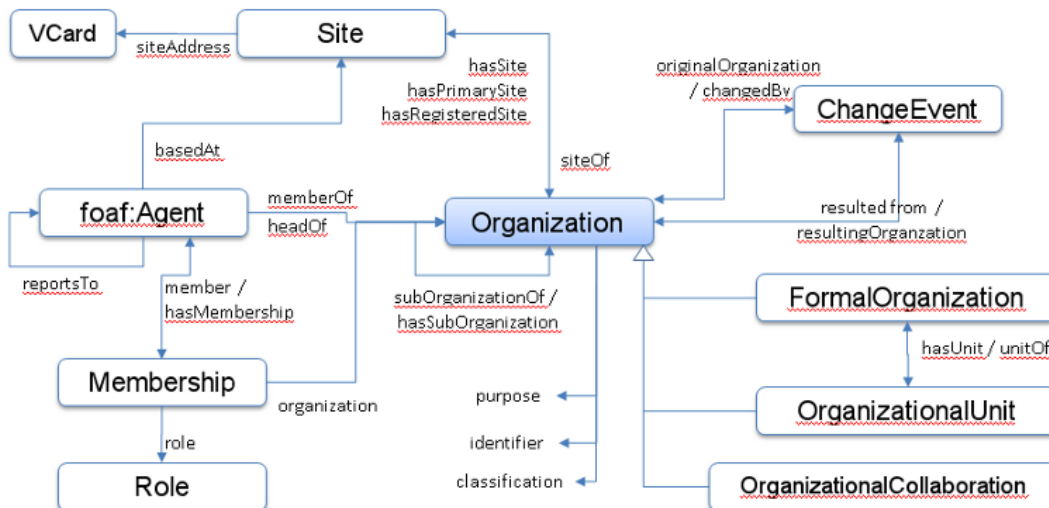
¹⁶² https://joinup.ec.europa.eu/asset/core_location/description.

3.6.2 Organization ontology

Tato ontologie byla původně motivovaná potřebou zveřejnovat informace týkající se veřejné správy v rámci iniciativy data.gov.uk. Jedná se o soubor malých, znovupoužitelných ontologií pro popis organizační struktury veřejné správy ve Velké Británii, použitelný i pro jiné země.

Ontologie popisuje základní pojmy z oblasti veřejné správy, jako jsou:

- Organizační struktura
 - o Základní popis organizační jednotky
 - o Rozklad na dílčí organizace a jednotky
 - o Význam (účel) organizace
- Popis (vnitřní) struktury
 - o Osoby v organizaci
 - o Role a vztahy osob a organizace
- Informace o umístění
 - o Budovy, apod.
- Historie organizace
 - o Přejmenování, sloučení/rozdělení organizací, změna účelu



Obrázek 35. Struktura Organization ontology

3.6.3 Public Contracts Ontology a Business Register Ontology

Public Contracts Ontology (PCO) je ontologie pro publikování veřejných zakázek jako Linked Data. Pokrývá základní vlastnosti veřejných zakázek, jejich části, zadavatele, dodavatele, atd.¹⁶³

Business Register Ontology (BRO) je ontologie pro zveřejnění obchodního rejstříku jako Linked Data. Pokrývá základní vlastnosti organizací, které se obvykle v obchodních rejstřících objevují. Také pokrývá osoby v organizacích a jejich vazby na organizace (majitel, člen správní rady, atd.).¹⁶⁴

¹⁶³ Zdroj: www.opendata.cz.

¹⁶⁴ Zdroj: www.opendata.cz.

3.6.4 Dublin Core

Dublin Core je soubor metadatových prvků, jehož záměrem je usnadnit vyhledávání elektronických zdrojů. Původně byl vytvořen jako popis zdrojů na WWW sestavený přímo autorem, postupně ale zaujal instituce zabývající se formálním zpracováním zdrojů, jako jsou muzea, knihovny, vládní agentury a komerční organizace¹⁶⁵.

Vedení Dublin Core sídlí ve Spojených státech amerických, konkrétně v OCLC¹⁶⁶ ve státě Ohio v Úřadu pro výzkum a speciální dokumenty. Struktura Dublin Core je v současnosti používána ve 20 zemích Severní Ameriky, Evropy, Asie a Austrálie a počet zemí se postupně zvyšuje.

Na základě výzvy pracovní skupiny "Dublin Core in Multiple Languages" pracuje Knihovnicko-informační centrum Masarykovy univerzity v Brně ve spolupráci se specialisty v oblasti knihoven na vytvoření české verze metadatového standardu 'Dublin Core' pro popis a podporu vyhledávání elektronických informačních zdrojů v českém prostředí.

Standard Dublin Core je jeden ze široce přijímaných návrhů systému dokumentace zdrojů na Webu. Používá transparentní jednoduchou sémantiku, schopnou pokrýt většinu potřeb pro metadata schémat ISVS (18).

3.6.5 FOAF – Friend Of A Friend

FOAF je strojově čitelnou ontologií popisující osoby, jejich činnost a vztahy mezi ostatními lidmi a objekty. Jedná se o nejběžnější ontologii pro popis této domény. Ostatní ontologie ji zpravidla mapují přímo, nebo z ní odvozují. Detailní informace o samotném projektu a ontologii nalezneme na adrese www.foaf-project.org.

Ukázka použití FOAF¹⁶⁷:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <foaf:Person>
    <foaf:name>David Banner</foaf:name>
    <foaf:title>Mr</foaf:title>
    <foaf:firstName>David</foaf:firstName>
    <foaf:surname>Banner</foaf:surname>
    <foaf:nick>hulk</foaf:nick>
    <foaf:homepage rdf:resource="http://www.davidbanner.com"/>
    <foaf:weblog rdf:resource="http://www.davidbanner.com/blog"/>
    <foaf:schoolHomepage rdf:resource="http://www.mit.edu"/>
    <foaf:workplaceHomepage rdf:resource="http://www.gamma-rays-r-us.com"/>
    <foaf:workInfoHomepage rdf:resource="http://www.gamma-rays-r-us.com/~banner/crazy-experiments.html"/>
    <!-- etc -->
  </foaf:Person>
</rdf:RDF>
```

Obrázek 36. Ukázka použití ontologie FOAF v RDF

¹⁶⁵ Zdroj: Dublin Core Czech, www.ics.muni.cz/dublin_core/

¹⁶⁶ <http://www.oclc.org/about/default.htm>

¹⁶⁷ Převzato z wiki.foaf-project.org/w/DescribingAPerson

 **The FOAF card** what's this?

Timothy Berners-Lee

Nick: TimBL, timbl
E-mail: timbl@w3.org
Phone: +1-(617)-253-5702
WWW: <http://www.w3.org/People/Berners-Lee/>
Blog: <http://dig.csail.mit.edu/breadcrumbs/blog/4>
Workplace: <http://www.w3.org/>

Office address:
 MIT CSAIL Room 32-G524
 32 Vassar Street
 Cambridge 02139
 USA

Knows:
[Coralie Mercier](#), [Dean Jackson](#), [Edd Dumbill](#), [Henry Story](#), <http://danbri.org/foaf.rdf#danbri>, [James Martin](#), [John Gage](#),
[John Klensin](#), [John Markoff](#), [John Seely Brown](#), [Tim Bray](#), [Joe Lambda](#), [Robert Hoffmann](#), [Ralph R. Swick](#), [Yosi Scharf](#),
[Daniel Krech](#), <http://heddley.com/edd/foaf.rdf#edd>, [Christoph Bussler](#), [Nicholas Gibbins](#), [Wendy Hall](#), [Nigel](#)



Obrázek 37. Ukázka použití ontologie FOAF

3.6.6 GoodRelations

GoodRelations je strojově čitelnou ontologií pro popis věcí v oblasti e-komerce. Je pouze ve formátu OWL DL¹⁶⁸. Je podporován internetovými vyhledávači jako je Google nebo Yahoo. Je dostupná pod licencí Creative Commons Attribution v 3.0.

¹⁶⁸ Ontology Web Language Description Logics

3.6.7 EuroVoc – mnohojazyčný tezaurus Evropské unie

EuroVoc je vícejazyčný polytematický tezaurus s terminologií oblastí, kterými se zabývá Evropská unie. Zaměřuje se především na parlamentní činnost. EuroVoc je k dispozici ve 22 úředních jazycích Evropské unie, v jednom jazyce kandidátské země (chorvatštině) a v jednom jazyce třetí země (srbštině)¹⁶⁹.

EuroVoc spravuje Úřad pro publikace. Správa tezauru nyní volí ontologický přístup a v souladu s doporučeními konsorcia World Wide Web (W3C) a s nejnovějšími trendy v tvorbě tezaurů používá sémantické webové technologie¹⁷⁰.

Tezaurus EuroVoc používají mimo jiné Evropský parlament, Úřad pro publikace, parlamenty členských států EU a jejich regionů, správní orgány členských států a soukromí uživatelé z členských i nečlenských zemí¹⁷¹.

Slovník je možné procházet podle oblastí a filtrovat jazyky, které potřebujeme. Slovník je také možné stáhnout jako celek ve formátu SKOS/XML. Slovník je poskytován pod licenci¹⁷², která umožňuje s obsahem nakládat při uvedení zdroje a zachování obsahu, zároveň je ale možné vytvářet překlady pro jazyky které nejsou součástí slovníku.

3.7 Existující licence pro publikaci dat

Publikace dat (obecně díla) je u nás upravena autorským zákonem¹⁷³, jejich další použití a způsob nakládání může být upraven licencí. Ve světě i u nás jsou obecně nejpoužívanější licence Creative Commons a GNU Free Documentation, blíže popsané dále v textu.

Za účelem svobodného šíření znalostí při zachování základních atributů autorství stanovila nadace Open knowledge definici kterou musí splňovat publikace díla, aby ji bylo možné považováno za otevřeno.

Některé země mají své vlastní licence, které odpovídají podmínkám uvedeným v Open Knowledge a jsou plně v souladu s jejich národní legislativou. Mezi takové licence např. patří britská Open Government Licence for public sector information¹⁷⁴ a částečně také Non-Commercial Government License for public sector information¹⁷⁵.

3.7.1 Open knowledge – Otevřené znalosti

Nadace Open Definitions¹⁷⁶ uvádí následující definici pojmu otevřená znalost.

Terminologie: Termín znalost v sobě zahrnuje:

- obsah jako jsou hudba, filmy, knihy
- data, ať už vědecká, historická, geografická nebo jiná
- informace z veřejné správy a administrativy

¹⁶⁹ Zdroj: <http://eurovoc.europa.eu/drupal/?q=cs>.

¹⁷⁰ tamtéž

¹⁷¹ tamtéž

¹⁷² <http://eurovoc.europa.eu/drupal/?q=cs/legalnotice>

¹⁷³ Zákon 121/2000 o právu autorském, právech souvisejících s právem autorským a o změně některých zákonů.

¹⁷⁴ Dostupné z <http://www.nationalarchives.gov.uk/doc/open-government-licence/>.

¹⁷⁵ Dostupné z <http://www.nationalarchives.gov.uk/doc/non-commercial-government-licence/>.

¹⁷⁶ Zdroj: <http://www.opendefinitions.org>.

Software je vynechán, i přes jeho zjevné důležitosti, protože už byl dostatečně popsán v jiných dokumentech. Termín dílo je určen k označení vložené části znalosti, nebo znalosti jako celku. Termín soubor může značit sbírku děl. Samozřejmě tento soubor může být považován rovněž za dílo sám o sobě. Termín licence se vztahuje na právní licenci, pod kterou je dílo zpřístupněno. V případě, že nebyla žádná licence stanovena, mělo by se předpokládat, že práce je zpřístupněna za standardních právních podmínek (např. copyright).

Definice: Dílo je otevřené, pokud splňuje následující podmínky:

1. Dostupnost
2. Redistribuce
3. Opětovné použití
4. Neexistence technologických omezení
5. Uvedení autora
6. Integrita
7. Zákaz diskriminace osob a skupin
8. Šíření licence
9. Licence nesmí být specifická pro soubor
10. Licence nesmí omezovat distribuci dalších prací.

Úplnou definici pojmu „otevřená znalost“, včetně popisu jednotlivých podmínek naleznete na webovém portálu www.opendefinition.org/okd/czech/.

3.7.2 Open Data Commons

Je nadace, která zahájila svoji činnost v prosinci 2008 a spravuje sadu licencí vhodných pro publikaci otevřených dat a databází. Mezi licence v současné době patří:

- PDDL - Public Domain Dedication and License
- ODC-By – Attribution License
- ODC-ODbL – Open Database License

Podrobné informace včetně textu licencí a způsobu použití naleznete na webu nadace <http://opendatacommons.org>.

3.7.3 Creative Commons

Creative Commons je americká nezisková organizace, která spravuje sadu licencí pro různou úroveň otevřené publikace.

Licence Creative Commons vznikají kombinací několika základních vlastností, které popisují, jaká práva si chce držitel autorských práv podržet, a jakých si přeje se vzdát. Existují čtyři základní vlastnosti¹⁷⁷:





Attribution (zkratka *by*): Uveďte autora — Máte povinnost uvést údaje o autorovi a tomto díle způsobem, který stanovil autor nebo poskytovatel licence (ne však tak, aby vznikl dojem, že podporují vás nebo způsob, jakým dílo užíváte).



Noncommercial (zkratka *nc*): Neužívejte dílo komerčně — Toto dílo nesmíte využít pro komerční účely.

¹⁷⁷ Zdroj: <http://creativecommons.org>.

 **No Derivative Works** (zkratka *nd*): Nezasahujte do díla — Toto dílo nesmíte pozměňovat, doplňovat, využívat celé nebo částečně v jiných dílech.

 **Share Alike** (zkratka *sa*): Zachovejte licenci — Pokud toto dílo jakkoliv upravíte nebo použijete ve svém díle, máte povinnost výsledek své práce šířit pod stejnou nebo slučitelnou licenci.

Kombinací může vzniknout až 16 různých typů licencí, z nichž ovšem pět nedává smysl. Licence bez atributu „by“ se zpravidla nepoužívá a nadace Creative Commons jej implicitně nenabízí.

Licence nadace Creative Commons mají oproti jiným dnes používaným licencím řadu výhod, která je činí oblíbenými. Jde především o tři druhy popisu licence:

- Právní popis
- Popis srozumitelný „běžnému“ člověku
- Strojově čitelný popis

3.7.4 GNU Free Documentation License

Jedná se o starší licenci než Creative Commons, týká se textů a užívá se např. na Wikipedii.

Podmínky použití:

- Uvedení původního autora (autorů)
- Uvedení licence
- CopyLeft – odvozená díla musí zachovat licenci

Její značnou nevýhodou a důvodem proč se od ní upouští je podmínka uvedení textu licence spolu s publikovaným textem, což je v mnoha případech značně komplikované¹⁷⁸.

3.8 Zhodnocení současných praktik publikace dat

Tato kapitola popsala současný stav publikování dat, a to jak z pohledu užívaných technologií, tak z pohledu významných systémů, které svá data zveřejňují. Rovněž popsala stav v ostatních zemích světa, které patří k tahounům iniciativy OpenGovernment partnership.

U katalogů dat, a s nimi spojených katalogizačních aktivit realizovaných u sledovaných zemí, lze vysledovat následujících sedm hlavních praktik (14):

- využívání samostatných katalogizačních nástrojů,
- katalogizace dat na různých úrovních veřejné správy,
- propojování katalogů na různých úrovních,
- potřeba definovaných procesů a legislativně podpořených procesů tvorby a údržby katalogu,
- existence jednoznačné metodiky určující co katalogizovat,
- použití jednotných taxonomií pro klasifikaci dat,
- jasné vymezení práv a povinností při využívání otevřených dat.

¹⁷⁸ Např. v případě vložení obrázku do prezentace je nutné vložit i celý text licence.

3.8.1 Zhodnocení situace v ČR

Veřejná správa je významným producentem dat, a již nyní svá data z části publikuje. Problematickou však zůstává koordinace, způsob a jednotnost publikace. Řada institucí VS si neuvědomuje význam správné publikace dat. Obecně můžeme identifikovat následující problémy:

- Publikovaná data jsou těžko dohledatelná, neexistuje prostředek, který by umožnil data vyhledávat na základě jejich vlastností (producent dat, metadata, kategorie a typ dat, atd.).
- Data jsou zveřejňována ve formátech, která neumožňují (dobrou) strojovou čitelnost (jako např. oskenované tištěné dokumenty, PDF, DOC, nestrukturované HTML, atd.).
- Data nejsou publikována pod vhodnou licencí, což znemožňuje veřejnosti data volně využívat, vyhledávat souvislosti a nalezené souvislosti zaznamenávat a dále zpracovávat.

4 Popis řešení

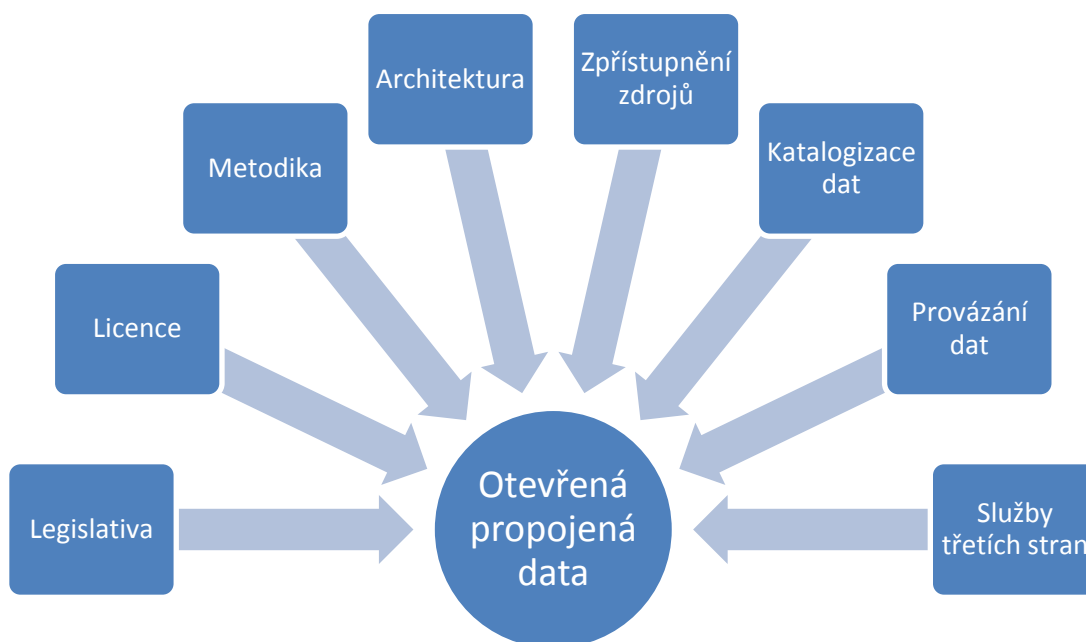
Pravidla umožňující publikaci a konzumaci informací musí být součástí systému řízení, a proto je nutné vybudovat odpovídající informační, organizační a sociální infrastrukturu. Tato infrastruktura musí mít stabilní, ale rozšiřovatelnou architekturu. Je také nutné stanovit závaznou metodiku, která sjednotí postupy napříč jednotlivými producenty dat.

Tato kapitola staví na úvodním popisečném textu, identifikuje základní problémy a hledá na ně odpověď. Zároveň uvádí jednotlivé dílčí kroky, které je nutné naplnit.

Kapitola není detailním rozбором problematiky, spíše se snaží vystihnout základní aspekty a otevřít diskuzi nad dalším postupem.

4.1 Rámec řešení

Otevřená propojená data veřejné správy v sobě zahrnují několik aspektů, které je nutno řešit. Každý aspekt samostatně přispívá k realizaci myšlenky otevřených propojených dat ve veřejné správě. Vynechání některého z aspektů bude mít za následek neúspěšnou realizaci záměru a to proto, že spolu úzce souvisí. Jednotlivé aspekty jsou znázorněny v následujícím diagramu. Postup je znázorněn na samostatném diagramu harmonogramu realizace.



Obrázek 39. Rámec problematiky otevřených propojených dat

4.1.1 Legislativa

Z právního pohledu můžeme na informace pohlížet ze dvou uhlů pohledu a to podle práva na:

- přístup k informacím,
- možnost dalšího použití.

Přístup k informacím je upraven řadou zákonů popsanou níže. Možnost dalšího použití je obecně upraveno autorským zákonem a následně licencí, pod kterou jsou informace publikovány, rozbořem tohoto aspektu se zabývá následující kapitola.

Otevřená data nejsou záležitostí pouze technickou, ale i právní, neboť je třeba zajistit, aby jejich uživatelé měli potřebná oprávnění k užití dat v rozsahu, který odpovídá požadované otevřenosti dat v souladu s právní ochranou a licencí a zároveň je potřeba „vynutit“ povinné subjekty k publikování svých dat, a to ve formě a s náležitostmi umožňujícími efektivní zpracování.

Právo na přístup k informacím, resp. jeho odejmutí lze rozdělit na několik kategorií podle své povahy, a to na:

- Informace osobní povahy
- Informace utajované povahy
- Informace autorské povahy
- Informace pro vnitřní potřebu
- Informace ostatní

Informace osobní povahy: Jsou takové informace, na které se vztahuje zákon o ochraně osobních údajů¹⁷⁹, Směrnice Evropského parlamentu a Rady¹⁸⁰, a dále které jsou vymezeny mezinárodními smlouvami, kterými je Česká republika vázaná¹⁸¹.

Osobním údajem se podle zákona¹⁸² rozumí jakákoliv informace týkající se určeného nebo určitelného subjektu údajů. Subjekt údajů se považuje za určený nebo určitelný, jestliže lze subjekt údajů přímo či nepřímo identifikovat zejména na základě čísla, kódu nebo jednoho či více prvků, specifických pro jeho fyzickou, fyziologickou, psychickou, ekonomickou, kulturní nebo sociální identitu.

Z tohoto zákona vyplývá i povinnost „nesdružovat osobní údaje, které byly získány k rozdílným účelům“¹⁸³.

Informace utajované povahy: Jsou takové informace, na které se vztahuje zákon o ochraně utajovaných skutečností¹⁸⁴.

Utajovanou informací se podle zákona¹⁸⁵ rozumí informace v jakékoliv podobě zaznamenaná na jakémkoliv nosiči označená v souladu s tímto zákonem, jejíž vyřazení nebo zneužití může způsobit újmu zájmu České republiky nebo může být pro tento zájem nevýhodné, a která je uvedena v seznamu utajovaných informací podle § 139 tohoto zákona.

¹⁷⁹ Celým názvem: Zákon č. 101/2000 Sb., o ochraně osobních údajů a o změně některých zákonů.

¹⁸⁰ Směrnice Evropského parlamentu a Rady 95/46/ES ze dne 24. října 1995 o ochraně jednotlivců v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů.

¹⁸¹ Úmluva o ochraně osob se zřetelem na automatizované zpracování osobních dat č. 108, vyhlášená pod č. 115/2001 Sb. m s.

¹⁸² § 4 zákona č. 101/2000 Sb., o ochraně osobních údajů.

¹⁸³ § 5 odst. 1) písm. h) zákona 101/2000 Sb., o ochraně osobních údajů.

¹⁸⁴ Celým názvem: Zákon č. 412/2005 Sb., o ochraně utajovaných skutečností, ve znění pozdějších předpisů.

¹⁸⁵ §2 písm. a) zákona č. 412/2005 Sb., o ochraně utajovaných skutečností.

Informace autorské povahy: Jsou takové informace, na které se vztahuje zákon o právu autorském¹⁸⁶. Tyto informace mohou být zveřejněny, ale na jejich užití se vztahují podmínky licence (v tomto případě nemůžeme hovořit o otevřených datech).

Předmětem práva autorského¹⁸⁷ je dílo literární a jiné dílo umělecké a dílo vědecké, které je jedinečným výsledkem tvůrčí činnosti autora a je vyjádřeno v jakékoli objektivně vnímatelné podobě včetně podoby elektronické, trvale nebo dočasně, bez ohledu na jeho rozsah, účel nebo význam (dále jen "dílo"). Dílem je zejména dílo slovesné vyjádřené řečí nebo písmem, dílo hudební, dílo dramatické a dílo hudebně dramatické, dílo choreografické a dílo pantomimické, dílo fotografické a dílo vyjádřené postupem podobným fotografii, dílo audiovizuální, jako je dílo kinematografické, dílo výtvarné, jako je dílo malířské, grafické a sochařské, dílo architektonické včetně díla urbanistického, dílo užitého umění a dílo kartografické.

Ochrana podle práva autorského se nevztahuje na úřední dílo, jímž je právní předpis, rozhodnutí, opatření obecné povahy, veřejná listina, veřejně přístupný rejstřík a sbírka jeho listin, jakož i úřední návrh úředního díla a jiná přípravná úřední dokumentace, včetně úředního překladu takového díla, sněmovní a senátní publikace, pamětní knihy obecní (obecní kroniky), státní symbol a symbol jednotky územní samosprávy a jiná taková díla, u nichž je veřejný zájem na vyloučení z ochrany¹⁸⁸. Užití takové dílo lze jen způsobem nesnižujícím jeho hodnotu.

Informace pro vnitřní potřebu: Jsou takové informace, které ze své povahy nejsou zveřejněny. Ať už rozhodnutím původce nebo vnitřním nařízením úřadu. Jedná se zpravidla o informace pracovního charakteru, informace, jejíž zveřejnění může způsobit újmu, apod. V tomto případě sice není zákonná překážka informace zveřejnit, ale na druhou stranu neexistuje právní síla, která by zveřejnění nařizovala¹⁸⁹.

Informace ostatní: Jsou takové informace, u kterých neexistuje zákonný důvod bránící jejich zveřejnění. Tyto informace jsou přístupné na základě zákona o svobodném přístupu k informacím¹⁹⁰ nebo Směrnice Evropského parlamentu a Rady¹⁹¹. Pokud to jejich povaha umožňuje, nebo se jedná o často žádané informace, jsou přístupné online prostřednictvím protokolu sítě internet.

4.1.2 Licence

V současné době není v platnosti žádný předpis, který by ukládal orgánům veřejné moci povinnost publikovat informace pod určitou licenci. Dochází k situacím, kdy nelze jednoduše rozpoznat, zda je informace volně dostupná, nebo je chráněna např. autorským zákonem.

To značně komplikuje publikování informací jako otevřená-propojená data, protože nemáme právní jistotu, že smíme informace dále využívat a zpracovávat.

¹⁸⁶ Celým názvem: Zákon č. 121/2000 Sb., zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, ve znění pozdějších předpisů.

¹⁸⁷ § 2 odst. 1) zákona č. 121/2000 Sb. Autorský zákon.

¹⁸⁸ § 3 odst. 1) zákona č. 121/2000 Sb. Autorský zákon.

¹⁸⁹ Tzv. jedná se o informace, které nejsou zpřístupněné na základě zákona č. 106/1999 Sb., o svobodném přístupu k informacím.

¹⁹⁰ Celým názvem Zákon č. 106/1999 Sb., o svobodném přístupu k informacím, ve znění pozdějších předpisů.

¹⁹¹ Směrnice Evropského parlamentu a Rady č. 2003/98/ES ze dne 17. listopadu 2003 o opakovaném použití informací veřejného sektoru.

Za tímto účelem je nezbytné stanovit povinnost orgánům veřejné moci uvádět licenci, pod kterou jsou informace publikována a informace, které jsou volně dostupné publikovat pod otevřenou licenci.

Souhrnný popis vlastností, jenž musí splňovat licence vhodná pro publikaci otevřených-propojených dat je uvedena v kapitole 3.7.1.

Tato práce si neklade za cíl důkladnou právní analýzu, a proto návrh možné licence pro publikaci dat veřejné správy ČR není zpracován a odkazujeme pouze na ostatní zdroje. Z rámcového studia problematiky však vyplývá, že jako nejvhodnější bude využít licenci Creative commons s úpravou pro podmínky České republiky, tzv. v úpravě vedoucí k souladu českém právem.

4.1.3 Metodika

Ne všechna data VS ČR je možné publikovat jako otevřená data. Pro identifikaci takovýchto překážek, ale i proto aby data měla odpovídající kvalitu, byla správně a dostatečně popsána metadaty a správně vložena do katalogu, je potřebné mít oficiální metodiku.

Prosazování metodiky musí být ve dvou úrovních, a to jednak jako doporučení, a jednak jako nařízení. Forma nařízení klade podmínky na zásadní prvky publikace, jako je forma, licence, druh dat apod. Forma doporučení pomáhá producentům dat data publikovat s jejich maximální hodnotou.

Aby mohla publikace a katalogizace dat VS efektivně probíhat, je třeba formulovat příslušné procesy pro publikaci a katalogizaci těchto dat a stanovit odpovědnosti za provádění těchto procesů. Publikaci a katalogizaci dat VS, a s ní spojené procesy a kroky může být nutné promítnout do legislativy nebo do vnitřních procesů subjektů provádějících katalogizaci.

Metodika by měla odpovědět minimálně na tyto otázky:

- Jak identifikovat data vhodná pro publikaci?
- Jak data čistit a publikovat s maximální informační hodnotou?
- Jak data publikovat ve vhodném formátu?
- Pod jakou licenci data publikovat?
- Jaká pravidla musí splňovat publikovaná data?
- Jak využít slovníky a ontologie?

Metodika není přímo součástí této práce, ale lze ji sestavit na základě popisu a shrnutí uvedených v této práci s doplněním „best practises“.

4.1.4 Architektura

Architektura je klíčovým aspektem celé problematiky. Její vhodný návrh umožní plné využití potenciálu otevřených propojených dat v dlouhodobém horizontu a umožní snížit náklady na provoz i při navýšení počtu producentů či konzumentů dat.

Architekturu hodnotíme podle kvalitativních atributů uvedených v následující tabulce. Volba jednotlivých atributů SW architektury byla volena s ohledem na současné praktiky a zkušenosti v této oblasti.

| Atribut | Popis |
|-------------------------------------|---|
| Modifikovatelnost | Schopnost architektury reagovat na změny v jejím účelu, způsobu použití a v rozšíření. Tzn. Odolnost vůči změnám. |
| Přenositelnost a znouvopoužitelnost | Schopnost architektury být implementována v prostředí jiné veřejné správy. |
| Integrovatelnost | Schopnost architektury využít již hotové komponenty a kooperovat se stávajícími systémy. |
| Modulárnost | Schopnost architektury fungovat jako sada nezávislých modulů, které vyžadují samostatný přístup a schopnost poskytovat (část) své funkcionality při vyřazení některého z modulu. |
| Realizovatelnost | Celkové náklady nutné na vybudování nezbytných součástí, poskytující klíčovou (povinnou) funkcionality a s ohledem na nezbytné změny ve stávajícím systému ve vztahu k aktuálním možnostem. |
| Provozovatelnost | Celkové náklady na provoz nezbytných součástí ve vztahu k aktuálním možnostem. |

Tabulka 10. Kvalitativní atributy SW architektury

Data publikovaná v rámci této architektury mohou být hodnocena podle těchto kvalitativních atributů:

| Atribut | Popis |
|---------------------|--|
| Propojitelnost | Data mohou být propojena s daty jiných producentů. |
| Dohledatelnost | Data propojená na jiná data se díky napojení stávají snadněji dohledatelná. |
| Pluralita | Lze publikovat více (i protichůdných) dat o jedné entitě. |
| Neúplnost propojení | Propojení dat může být vynecháno nebo odloženo do doby až bude možné vzhledem k ostatním datům nebo výhodné např. z ekonomického hlediska. |
| Transparentnost | Víme, kdo a kdy publikuje jaká data. |
| Distribuuovatelnost | Data mohou být publikována po částech na různých místech |
| Modifikovatelnost | Datová infrastruktura je odolná vůči změnám. Změna struktury dat neohrozí dat v jiných setech. |
| Granularita | Data mohou být publikována s různou úrovní granularity, např. rozpočet za celou ČR, nebo podle jednotlivých územních celků. |

Tabulka 11. Kvalitativní atributy publikace dat

Konkrétní návrh architektury je uveden v samostatné kapitole 5.

4.1.5 Zpřístupnění zdrojů

V okamžiku, kdy budou odstraněny legislativní problémy a dojde k vydání vhodné licence a metodiky pro publikaci otevřených dat a bude připravena nezbytná infrastruktura, je veřejná správa připravena zpřístupnit hlavní zdroje dat k publikaci. Zdroje dat vhodné pro publikaci byly navrženy v akčním plánu vlády a jsou popsány v kapitole 3.1 a v příloze.

Zpřístupnění zdrojů nespočívá pouze v samotném veřejném publikování, ale, a to především, v odstranění technických překážek v procházení a stahování datasetů a zajištění (dobré) strojové čitelnosti. Vlastnosti (dobré) strojové čitelnosti jsou vysvětleny v kapitole 2.2.3.

Formáty, ve kterých je možné publikovat data, jsou popsány v kapitole 3.2, kde je rovněž jejich zhodnocení a použitelnost pro publikaci jako otevřená-propojená data.

Jako nejvhodnější se jeví publikace dat pomocí technologie RDF, která dokáže nést data i jejich sémantiku a data je možné mezi sebou snadno propojovat a to i napříč různými datasety. Na druhou stranu je taková publikace technicky náročnější a zpravidla vyžaduje i úpravu stávajících systémů.

Dalším vhodným způsobem je publikace ve strukturované podobě např. pomocí technologie XML. V tomto případě sice nemáme tak silné možnosti jak doplnit sémantiku, ale nemusíme zpravidla upravovat stávající systémy.

Dobrá strojová čitelnost je mimo jiné zajištěna použitím ontologií pro popis dat, z vymezených cílů je do budoucna nutné připravit minimálně tyto ontologie:

- Lokace
 - o Adresa (adresní místo)
 - o Uzemní celek
- Organizace a vztahy
 - o Instituce veřejného sektoru (OVM, a některé další)
 - o Soukromé společnosti
 - o Nestátní neziskové organizace
- Služby
 - o Typy služeb
- Veřejnost
 - o Občan
 - o Životní události
 - o Životní situace
 - o Životní potřeby

Některé potřebné ontologie již existují, jedná se např. o ontologii pro popis veřejných zakázek nebo rozpočtu.

4.1.6 Katalogizace dat

Zpřístupněné zdroje (konkrétní datasety) je vhodné zanést do katalogu dat a to za účelem zvýšení jejich dohledatelnosti. Katalog umožní vyhledávání datasetů nejen pomocí jejich názvu, ale i popisu a přidávaných metadat, zároveň je možné kontrolovat přístupnost dat v čase apod. Jednotlivé přístupy ke katalogizaci a rozpracování vhodného přístupu jsou uvedeny v kapitole 5.

Přínosem katalogizace dat veřejné správy je podle (14):

- usnadnění přístupu k datům veřejné správy,
- vytvoření předpokladu pro snazší opětovné použití dat veřejné správy,
- vytvoření předpokladu pro využívání otevřených propojených dat v rámci VS ČR,
- vyšší transparentnost veřejné správy,
- zlepšení sdílení a správy dat uvnitř veřejné správy.

Povinnost katalogizovat data by měla být součástí legislativy a to v podobě že „všechna nově publikovaná otevřená veřejná data jsou do určité doby po své publikaci zanesena do katalogu dat“.

Vyhledatelnost dat z určité oblasti a jejich propojení selepší, pokud pro klasifikaci dat i pro klasifikaci záznamů o těchto datech budou používány jednotné taxonomie-slovníky a ontologie. Je nutné, aby v rámci VS ČR byly slovníky a ontologie harmonizovány. V případě mezinárodní spolupráce je harmonizace vhodná a to minimálně na úrovni propojení s portálem otevřených dat EU. Příkladem takovéto harmonizace taxonomií je použití slovníku EUROVOC.

Slovníky a ontologie, které jsou spravovány správcem katalogu.

4.1.7 Provázání dat

Samotné zveřejnění informací v rozumném formátu, tzv. zajištění (dobré) strojové čitelnosti, je krok směrem k otevřeným datům. Pro vybudování otevřených propojených dat je nutné zveřejněné informace provázat s jinými daty. Díky tomu vznikne síť dat různých producentů a různých oblastí, které se ale díky vazbám vzájemně obohacují a zvyšují svojí hodnotu.

Po vybudování datového katalogu je tedy dalším stupněm vytvoření datové infrastruktury otevřených-propojených dat.

Pro provázání dat je nutné data a případně jejich vlastnosti jednoznačně a trvale identifikovat pomocí URI. Principy tvorby URI jsou naznačeny v samostatné kapitole 6.1.4.

4.1.8 Služby třetích stran

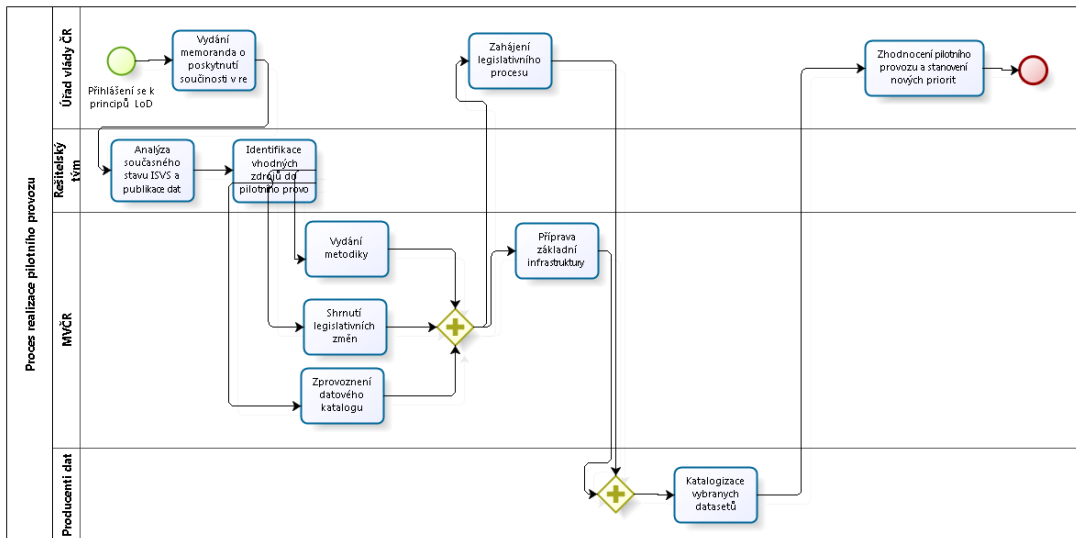
V okamžiku kdy jsou data přístupná má veřejnost sice možnost data procházet, ale vzhledem k obrovskému množství a různých specifických vazeb bude pro především laickou veřejnost problematické z dat získat rozumné informace. K tomuto účelu slouží aplikace a služby postavené nad těmito daty. Tyto služby mohou být budované a následně provozované kýmkoliv jak veřejnou správou, tak veřejností a mohou poskytovat své „obohacené“ výstupy zdarma, nebo za úplatu.

Příklady již existujících aplikací využívající otevřená-propojená dat je uveden v kapitole 3.3.4 a v příloze.

4.2 Harmonogram realizace

Realizace vyžaduje zapojení mnoha subjektů VS na mnoha úrovních řízení. Jednak je zcela nutné se jednoznačně přihlásit k záměru realizovat otevřená propojená data v České republice, následně zahájit legislativní proces u souvisejících zákonů, doporučit vhodnou licenci pro publikaci dat a připravit závaznou metodiku. Teprve poté je možné zpřístupnit vybrané zdroje a zahájit jejich katalogizaci. Opět u vybraných zdrojů je vhodné data publikovat jako propojená.

Samotný proces je stručně znázorněn na následujícím diagramu.



Obrázek 40. Harmonogram realizace otevřených-propojených dat v ČR

5 Katalog otevřených dat veřejné správy

Datový katalog je jednotným místem pro vyhledávání otevřených dat veřejné správy a pro získávání údajů o těchto datech (tj. získávání metadat). Díky tomu datový katalog usnadňuje nejen vyhledávání otevřených dat veřejné správy, ale díky poskytování metadat, jako jsou například autor, datum vytvoření a publikace, používané datové formáty, schémata apod., usnadňuje potenciálním uživatelům i pochopení dat. (14)

Stěžejním posláním katalogu je poskytnout snadnou dohledatelnost dat veřejné správy. Za tímto účelem katalog umožní vložení záznamu o datech, jejich označení a uživatelsky přívětivé vyhledávání nad těmito záznamy. Označením dat chápeme jejich pojmenování, připojení značek (tagů) a nastavení parametrů jako je autor, správce, licence, apod.

Katalog musí / měl by poskytnout možnost indexovat otevřená data ve všech jejich variantách, tedy jako odkaz na statická data, odkaz na dynamicky generovaná data a popisem rozhraní pro přístup ke službám.

Pro katalog jsou podstatné tyto otázky:

- Bude existovat více specializovaných katalogů, pokud ano bude se replikovat do centrálního (nadřazeného) katalogu?
- Kde budou umístěny datasety? V katalogu, u zdroje, kombinace obou?
- Kdo bude moci vkládat záznamy, bude probíhat redakce?
- Kdo katalog vybuduje?
- Kdo bude katalog provozovat?
- Kdo bude zodpovídat za správnost katalogu?

Na jednotlivé otázky odpovídají následující kapitoly.

5.1.1 Vlastnosti katalogu

Katalog umožňuje propojení dat na úrovni datasetu. Nejedná se o propojení ve smyslu LinkedData ale o hrubší, méně detailní vazbě na úrovni celého datasetu. Můžeme hovořit o vazbách mezi množinami dat A a B typu:

- A a B jsou totožné
- A je v rozporu s B
- A je podmnožinou B
- A je agregací dat z B
- A mnohé další specializovanější vazby

5.2 Varianty provedení datového katalogu

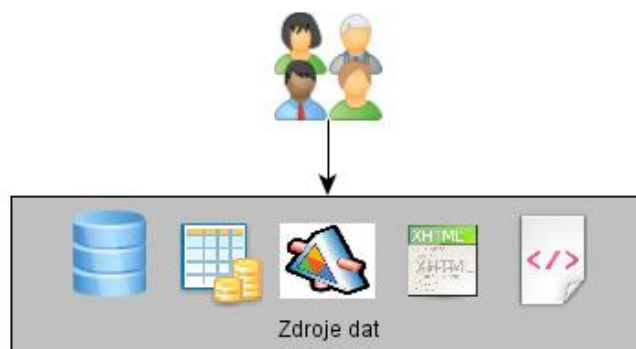
Realizace datového katalogu je možná ve třech hlavních provedeních, jedná se o:

- Provoz bez datového katalogu
- Centralizovaný provoz
- Distribuovaný provoz

Architektura jednotlivých řešení je popsána a znázorněna v následujících kapitolách. Hlavním rozdílem jednotlivých variant je ve formě přístupu konzumenta dat a producenta dat, tzv. ve vazbě mezi katalogem a publikovanými daty.

5.2.1 Provoz bez datového katalogu

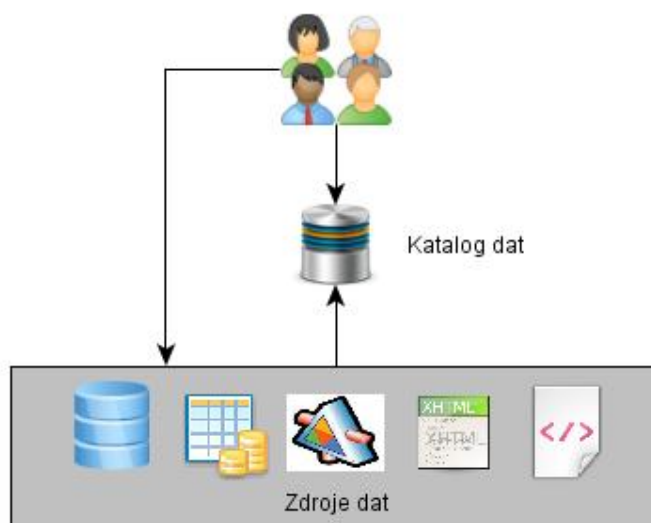
V případě provozu bez datového katalogu nemá konzument dat možnost vyhledávat daty jinak než procházením jednotlivých zdrojů dat. Vzhledem k neexistenci jednotného popisu dat není možné snadno dohledat podobné daty. Tato varianta sice nemá žádné nároky na vybudování nebo provoz, avšak nemá žádný přínos, vzhledem k tomu že budování datové infrastruktury otevřených propojených dat bez katalogu dat je velmi nešťastné nebudeme tuto variantu dále rozebírat.



Obrázek 41. Katalogizace - varianta bez katalogu dat

5.2.2 Centralizovaný přístup

Varianta centralizovaného katalogu předpokládá jeden hlavní katalog pro všechny producenty dat. Nespornou výhodou tohoto řešení je snadná integrovatelnost (neexistují jiné katalogy, které je nutné integrovat) a také správa. Diskutabilní výhodou je fakt, že v katalogu budou indexovány daty z různých domén zájmu a katalog tak nebude moci snadno poskytovat pro tyto data nadstandardní služby, jako je například vizualizace apod. přímo v katalogu.

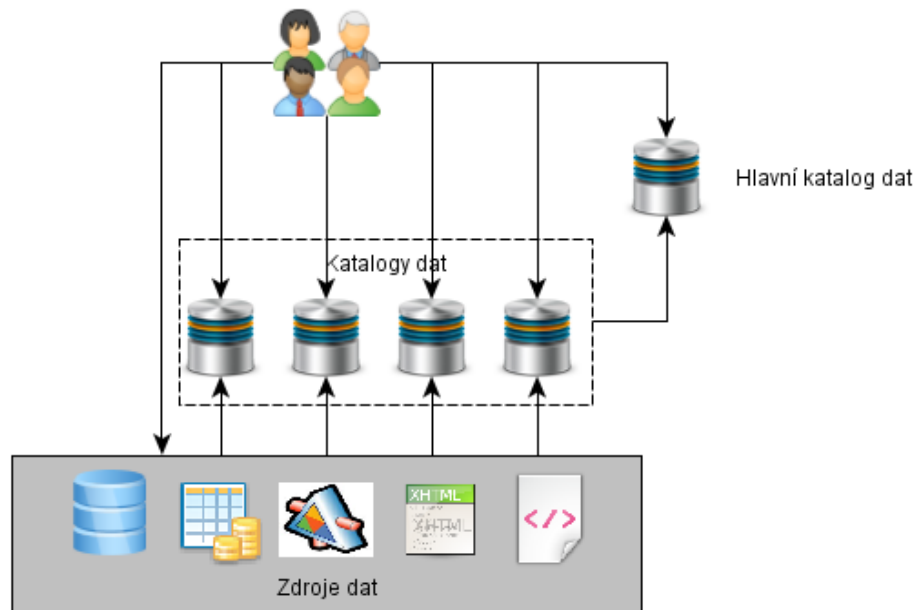


Obrázek 42. Katalogizace - varianta s hlavním katalogem dat

5.2.3 Distribuovaný přístup

Varianta distribuovaného přístupu předpokládá existenci několika katalogů zaměřených na určitou oblast zájmu. Volitelně je možné provozovat hlavní katalog, který bude jednotlivé katalogy integrovat bez poskytnutí dodatečných nadstandardních služeb.

Tato varianta přináší řadu výhod, a to především dobrou škálovatelnost, možnost poskytnutí specializovaných služeb, apod.. Drobnou nevýhodou je větší počet (specificky zaměřených) katalogů a tím i náročnější integrace jejich metadat.



Obrázek 43. Katalogizace- varianta s distribuovanými katalogy

V případě distribuovaného přístupu nás rovněž zajímá, zda jednotlivé specializované katalogy jsou provozovány jako homogenní nebo hybridní řešení, tzn., zda jsou katalogy provozovány jedním nástrojem s doplněním nadstandardních služeb, anebo zda jsou katalogy provozovány na různých nástrojích a platformách.

5.3 Varianty replikace dat

Replikací dat je myšleno umístění samotných dat vůči katalogu. Uvažujeme tyto případy:

- Data jsou umístěna u producenta, v katalogu je popis a odkaz na data
- Data jsou replikována do katalogu, v katalogu je tedy jak popis, tak samotná data
- Katalog umožní provoz v obou z výše uvedených variant

V případě že, jsou data umístěna u producenta snižujeme kapacitní nároky a také režii na katalog. Na druhou stranu v případě výpadku, nebo ukončení činnosti producenta, mohou být data ztracena.

Varianta, že jsou data replikována, v zásadě obrací výhody a nevýhody první varianty, zároveň ale hrozí i neaktuálnost dat v případě, že data byla aktualizována, aniž by se jednalo o nový dataset. Tato varianta je navíc přípustná pouze pro statické

datasets. V případě dynamických dat, nebo přístupu prostřednictvím služeb, musí být řešeno formou první varianty.

Kombinace obou přístupů spojuje jejich výhody a zároveň minimalizuje nevýhody. Na druhou stranu musí katalog disponovat rozsáhlejší funkcionalitou, a mít větší rezervu provozu. Jeho rozsáhlost se projeví i ve větších nárocích na správu katalogu.

5.4 Varianty správy datového katalogu

Při správě datového katalogu rozlišujeme úroveň přístupu jednotlivých uživatelů k uveřejnění dat, resp. proces uveřejnění dat podle jednotlivých uživatelů.

Hlavními uživateli pro publikaci dat jsou:

- Veřejnost
- Ověřený uživatel
- Zástupce OVM
- Redaktor (ověřený uživatel katalogu)

5.4.1 Veřejnost

Do této skupiny řadíme všechny neověřené uživatele. Výhodou povolení této skupiny je rychlé a levné naplnění katalogu informacemi, nicméně zde hrozí vkládání nesprávných, neúplných nebo zavádějících informací, případně nedodržování předepsané metodiky.

Jistým zvýšením důvěryhodnosti této skupiny může být zavedení role redaktora, který jednotlivé vkládané údaje ověří. To ale vytváří náklady, které snižují výhodnost zřízení této skupiny. Případně je možné umožnit veřejnosti pouze omezenou funkcionalitu, která neohroží bezproblémový chod katalogu. Tato funkcionalita ale zvyšuje komplikovanost katalogu a jeho pořizovací cenu.

5.4.2 Ověřený uživatel

Ověřený uživatel je uživatel z řad veřejnosti, který byl ověřen a schválen správcem datového katalogu. Takový uživatel vystupuje pod svým jménem a je zavázán dodržovat předepsanou metodiku.

Dá se předpokládat, že tato skupina tvořena především zástupci akademické sféry a technicky zdatnými jedinci veřejné správy, bude dodržovat metodiku práce s katalogem, zároveň svou činností pokryjí oblasti, ve kterých zatím nedochází k publikaci (např. proto že daný OVM zatím nezahájil spolupráci). Výhody této skupiny jsou obdobné jako u veřejnosti s rozdílem, že není potřeba zřizovat roli redaktora, a tím navyšovat provozní náklady.

5.4.3 Zástupce OVM

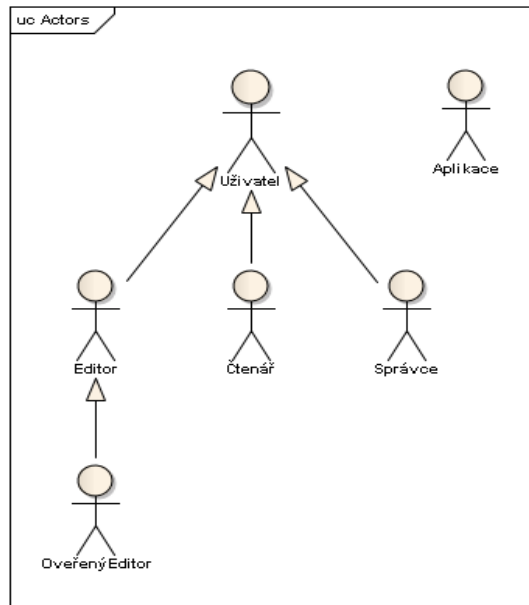
Jedná se o zaměstnance instituce, která je původcem dat. U této skupiny se dá očekávat nejvyšší informovanost o datech konkrétní instituce, zároveň ale není možné předpokládat shodnou úroveň technické znalosti.

5.4.4 Redaktor

Je zaměstnancem provozovatele katalogu (případě osobou již byla svěřena důvěra). Má oprávnění schvalovat příspěvky vložené uživateli z kategorie „veřejnost“.

5.4.5 Struktura aktérů pro správu katalogu

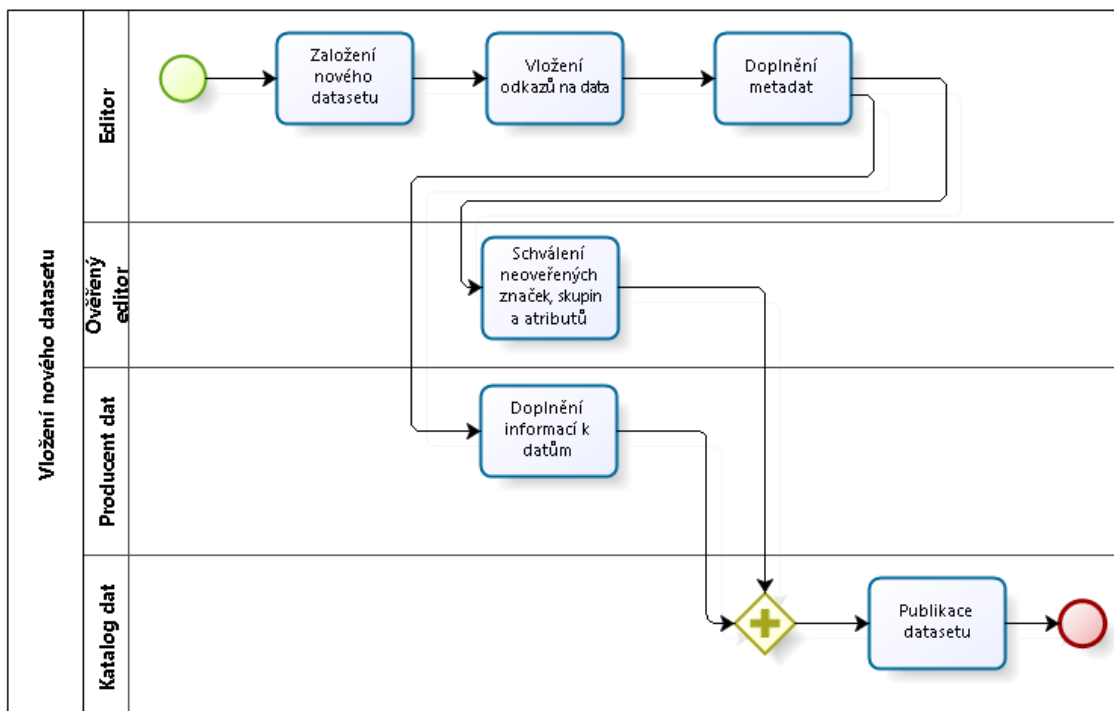
Ve správě katalogu budou vystupovat aktéři (role uživatelů) uvedení na diagramu.



Obrázek 44. Role uživatelů pro správu katalogu otevřených dat

Na tyto aktéry budou mapováni jednotliví uživatelé ze skupin uvedených v předchozích kapitolách. Každá skupina bude mít přiřazena určitá oprávnění pro správu katalogu.

Proces naplňování katalogu daty můžeme navrhnout takto:



Obrázek 45. Proces plnění katalogu daty

5.5 Varianty plnění datového katalogu

Plněním katalogu je myšleno jeho naplňování daty. Katalog je možné plnit třemi způsoby, a to:

- Ruční
- Automatizované
- Polo automatizované

5.5.1 Ruční plnění katalogu

Při této variantě jsou jednotlivé příspěvky vkládány do katalogu ručně. Tato varianta je vhodná pro většinu stávajících systémů, které nepodporují notifikaci změny obsahu, případně nebudou přímo napojeny na katalog. Jedná se o časově náročnější variantu, která ale nevyžaduje změnu ve stávajících systémech. Varianta zvyšuje náklady na provoz při vkládání jednotlivých záznamů, ale ušetří se náklady na změny stávajících systémů.

5.5.2 Automatizované plnění katalogu

Tato varianta předpokládá, že systémy, které publikují data, jsou přímo napojeny na katalog a vkládají nové příspěvky, vč. popisu a označení automatizovaně. Tato varianta vyžaduje rozsáhlé změny ve stávajících systémech a zároveň nutnou podporu ze strany katalogu. Zvýšené počáteční náklady se odrazí ve snížení nákladů na správu. Tento přístup plnění je vhodný zejména pro nově vyvíjené systémy.

5.5.3 Polo automatizované plnění katalogu

Je kompromisním řešením automatizované a ruční varianty. Varianta předpokládá, že systém publikující data oznámí nově publikovaná data katalogu formou oznámení – skutečnosti že v daném systému došlo ke změně. Následně redaktor katalogu vloží do katalogu nový příspěvek s patřičným popisem, odkazem na data a označením. Tato varianta předpokládá pouze dílčí změny ve stávajících systémech, zároveň umožňuje udržet katalog aktuální i v oblastech kde není vhodné zaškolovat pracovníka OVM pro plnění katalogu.

5.6 Varianty konzumace obsahu prostřednictvím katalogu

Existence katalogu umožní vedle přímého vyhledání dat tyto přívětivější způsoby:

- Vyhledání dat a následná práce přímo se zdrojem
- Vyhledání dat a práce s daty prostřednictvím katalogu

V obou případech dojde k vyhledání odpovídajících dat. V první variantě převezmeme odkaz na vyhledaná data a dále v aplikaci již pracujeme přímo se zdrojem dat, tzv. dále již nevyužíváme služeb katalogu. Tato varianta má výhodu v menším nároku na vytížení katalogu a jednodušší aplikaci. Nevýhodou je změna umístění datasetu.

Druhá varianta předpokládá, že konzument dat se nejprve dotáže katalogu na aktuální umístění dat a následně s nimi bude po určitou dobu pracovat.

5.7 Varianty pořízení datového katalogu

Za účelem ochrany investic a maximálního využití již provozované infrastruktury je vhodné prozkoumat možnost provozu datového katalogu na současných systémech. V případě že provozovatelem datového katalogu bude stát, bude tento krok nutné provést v případové studii, která jednotlivé navržené kroky zdůvodní.

V současnosti jsou v provozu dva významné systémy, které mohou být využity jako datový katalog. Jedná se o portál Veřejné správy a Informační systém o informačních systémech veřejné správy.

5.7.1 Portál veřejné správy

Portál je popsán v kapitole 3.1.1. Portál slouží k „univerzální“ komunikaci s širokou veřejností a firmami, klade si za cíl být středobodem komunikace mezi veřejnou správou a okolím. Jedná se však primárně o informační zdroj směrem k veřejnosti.

Portál svým současným zaměřením nepřináší pro proces budování nebo provozu datového katalogu žádné přínosy. Provozování katalogu na portálu by narušilo jeho integritu.

5.7.2 Informační systém o informačních systémech veřejné správy

Systém je popsán v úvodních kapitolách. IS ISVS slouží jako registr provozovaných IS. Z důkladného prostudování systému vyplývá, že se jedná o uzavřený systém navržený pro jednorázový účel. Rozšíření IS ISVS o funkcionalitu katalogu není vhodné, a to z důvodu efektivnosti vynaložených prostředků, s ohledem na budoucí rozšíření katalogu a zvyšující se provoz a kapacitu dat.

Vzhledem ke svému zaměření je IS ISVS vhodným prostředkem pro identifikaci datových zdrojů vhodných k publikaci. Je zde rovněž možné posoudit, zda konkrétní IS má vhodné rozhraní k přímému napojení na katalog apod.

5.8 Využití portálu

Mezi předpokládané uživatele-konzumenty dat a jejich využití katalogu můžeme zařadit následující skupiny:

| Uživatel | Způsob použití |
|-------------------|---|
| Veřejnost | Vyhledávání a stahování dat. Prostřednictvím aplikací třetích stran, vizualizace a strukturované zobrazení souvisejících dat. |
| Komerční sféra | Provoz služeb využívajících data veřejné správy. Vyhledání nových obchodních příležitostí |
| Novináři | Vyhledávání informací pro investigativní žurnalistiku, žurnalistika řízená daty. |
| Odborná veřejnost | Analýzy a statistika. Vyhledávání souvislostí |
| Neziskový sektor | Podobně jako novináři, především pak vyhledávání a analýza využívání prostředků veřejné správy, odhalování |

| | |
|----------|--|
| | korupce nebo posuzování efektivity vynaložených prostředků. |
| Vývojáři | Vývoj aplikací a poskytování služeb využívajících nezpracovaná veřejná data. |
| Vláda | Hodnocení kvality VS. Posílení poskytovaných služeb. Adresování služeb, řízení výkonosti. Efektivita využívání veřejných prostředků. Zvýšení kvality života. |

Tabulka 12. Využití katalogu otevřených dat

5.9 Zhodnocení a doporučení optimální varianty

Zhodnocení optimální varianty pro katalogizaci dat veřejné správy vychází z posouzení těchto aspektů:

- Přístup a celková spokojenost (úspěšnost) katalogizace u hlavních zemí, která svá data katalogizují.
- Zhodnocení specifik veřejné správy ČR
 - o nelze předpokládat aktivní zapojení všech subjektů VS
 - o Infrastrukturu je nutné budovat postupně
- Zhodnocení na základě odhadu celkových nákladů na vybudování a následný provoz, rychlosti nasazení, potenciálu pro budoucí růst, integrity dat, propojitelnosti s jinými evidenčními systémy, bezpečnosti, technické proveditelnosti a množství personálu pro obsluhu.

Detailní rozbor včetně diskuze jednotlivých aspektů je uveden v materiálu (14), na kterém autor spolupracoval. Na základě tohoto hodnocení a vlastního přístupu je doporučena tato varianta katalogizace:

- Distribuované řešení (více specializovaných katalogů),
- Jeden hlavní integrující katalog,
- Katalogy využívají společné služby,
- Katalogy využívají společná metadata.
- Katalog může editovat kdokoliv, editace veřejnosti podléhá schválení redaktorem.

Architektonické rozpracování katalogu je začleněno a rozpracováno v kapitole 6.

Věcným správcem katalogu bude podle kompetenčního zákona¹⁹² pověřeno Ministerstvo vnitra a to protože je ústředním orgánem státní správy pro oblast informačních systémů veřejné správy¹⁹³ a také plní koordinační úlohu pro informační a komunikační technologie¹⁹⁴.

¹⁹² Zákon č. 2/1969, o zřízení ministerstev a jiných ústředních orgánů státní správy České republiky, ve znění pozdějších předpisů.

¹⁹³ §12, odst. 1, písm. o) zákona č. 2/1969, o zřízení ministerstev a jiných ústředních orgánů státní správy České republiky, ve znění pozdějších předpisů.

¹⁹⁴ §12, odst. 6 zákona č. 2/1969, o zřízení ministerstev a jiných ústředních orgánů státní správy České republiky, ve znění pozdějších předpisů.

6 Architektura otevřené veřejné správy

Globální architektura byla navržena s ohledem na cíle vytýčené v úvodu a současné záměry České republiky v oblasti informatiky a e-governmentu. Architektura zohledňuje ekonomická opatření vytýčená vládou. Je modulární – servisně orientovaná a identifikuje priority, podle kterých je možné ji implementovat.

Jak bylo v úvodních kapitolách této práce zmíněno, můžeme v rámci veřejné správy identifikovat tyto vlastnosti:

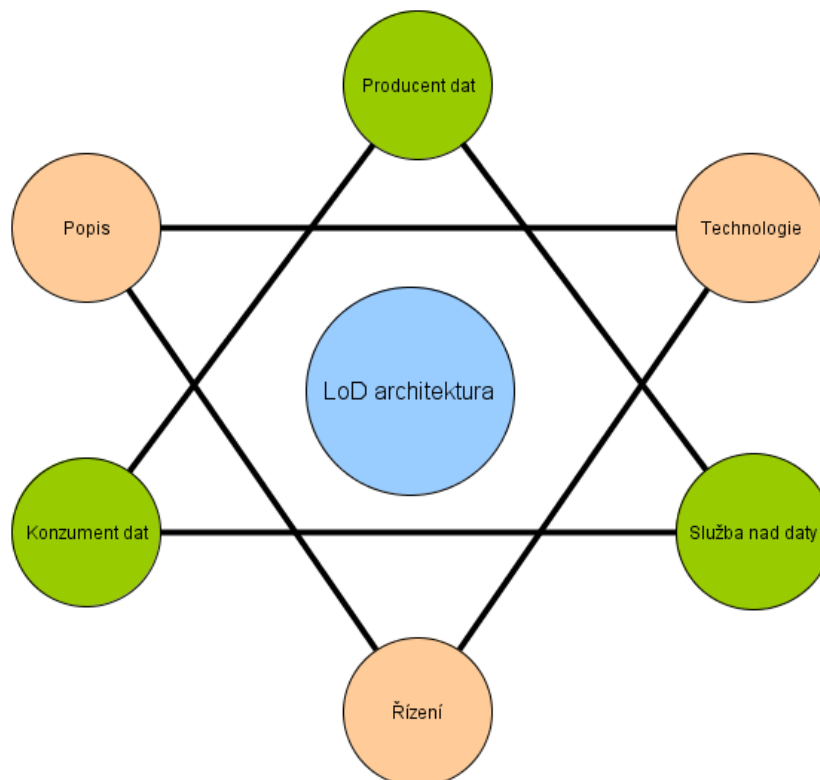
- VS je velmi rozsáhlá a to jak horizontálně tak vertikálně.
- Úroveň jednotlivých systémů VS je značně rozdílná.
- Neexistuje jednotná koncepce oblasti IT pro VS.
- Řízení IT ve VS je decentralizované s různou úrovní přístupu.
- VS je nehomogenním prostředím, používá se zde mnoho platform, technologií i architektur.
- Financování rozvoje IT je omezeno a soustředí se na klíčové systémy.

Ze shora uvedených vlastností můžeme stanovit tato pravidla pro návrh vhodné architektury:

- Architektura musí být nezávislá na platformě a technologii.
- Architektura musí být modulární, kde moduly jsou na sobě co nevíce nezávislé.
- Infrastruktura dokáže fungovat s omezeným počtem funkčních modulů.
- Náklady na realizaci nezbytné infrastruktury musí být dostatečně nízké.
- Náklady na provoz musí být dostatečně nízké a musí umožnit zapojení subjektů mimo VS.

Výše uvedená omezení vedou k rozhodnutí zvolit koncept servisně orientované architektury. Servisně orientovaná architektura se vyznačuje modulárností, znouvopoužitelností volně vázaných na sobě nezávislých a bezstavových služeb.

Konceptuálně můžeme na architekturu VS nahlížet ze šesti provázaných hledisek. Jednotlivé pohledy jsou naznačeny na diagramu.



Obrázek 46. Konceptuální architektura otevřených-propojených dat veřejné správy

Producent dat je libovolný subjekt poskytující data. Jedná se zde především o:

- Ústřední orgány státní správy
- Státní instituce
- Vládu
- Obce, města, kraje a další územní samosprávné celky

A v neposlední řadě také o:

- Subjekty neziskového sektoru
- Výzkumné a akademické organizace
- Veřejnost

Konzument dat je libovolný subjekt konzumující data, a to ve formě prostých dat, nebo služeb postavenými nad daty. Jedná se především o tyto skupiny:

- Novináři
- Neziskový sektor
- Veřejnou správu
- Komerční sféru
- Vývojáře
- Obecně o veřejnost

Využití otevřených resp. otevřených propojených dat a v neposlední řadě také katalogu dat je uvedeno dále v textu.

Služba nad daty je míněna služba v obecném slova smyslu, která poskytuje jeden nebo více služeb tohoto typu:

- Přímo poskytuje data,
- Umožňuje data vyhledávat na základě metadat (tzv. katalog),
- Umožňuje data procházet na základě souvislostí a vazeb,
- Vizualizuje data do map nebo jiných vhodných zobrazení,
- Obecně služba, která data skládá, filtruje, hodnotí, vizualizuje, a to za účelem zvýšení jejich hodnoty nebo snazšího pochopení.

Popis je prvkem architektury podporující sjednocení popisu dat s cílem jejich lepšího pochopení, možnosti propojení s jinými daty a v neposlední řadě umožnění strojového automatizovaného zpracování. Jedná se především o tyto prvky:

- Slovníky
- Ontologie
- Taxonomie
- Popis rozhraní pro komunikaci
- Metriky
- Dohoda o úrovni poskytovaných služeb

Řízení je základním prvkem pro udržení shody návrh s reálným stavem, vynucením povinných subjektů zveřejňovat svá data v souladu s metodikou a v neposlední řadě také dohled nad touto činností. V oblasti řízení otevřených dat veřejné správy identifikujeme tyto prvky:

- Metodiky
- Standardy
- Dohled
- Směrnice
- Řízení metadat a identifikátorů
- Ukázky a doporučení správného použití
- Legislativa

Technologie je nositelem implementační části celé architektury, obsahuje v sobě použité nástroje, dokumentaci k nástrojům apod.

6.1 Enterprise architektura

Dle popsaných požadavků na infrastrukturu a katalogizaci dat je architektura navržena takto:

- Data jsou fyzicky umístěna u producenta dat
- Katalog je navržen jako distribuovaný s jedním hlavním katalogem
- Katalogy využívají společnou základnu služeb a infrastrukturu metadat a identifikátorů

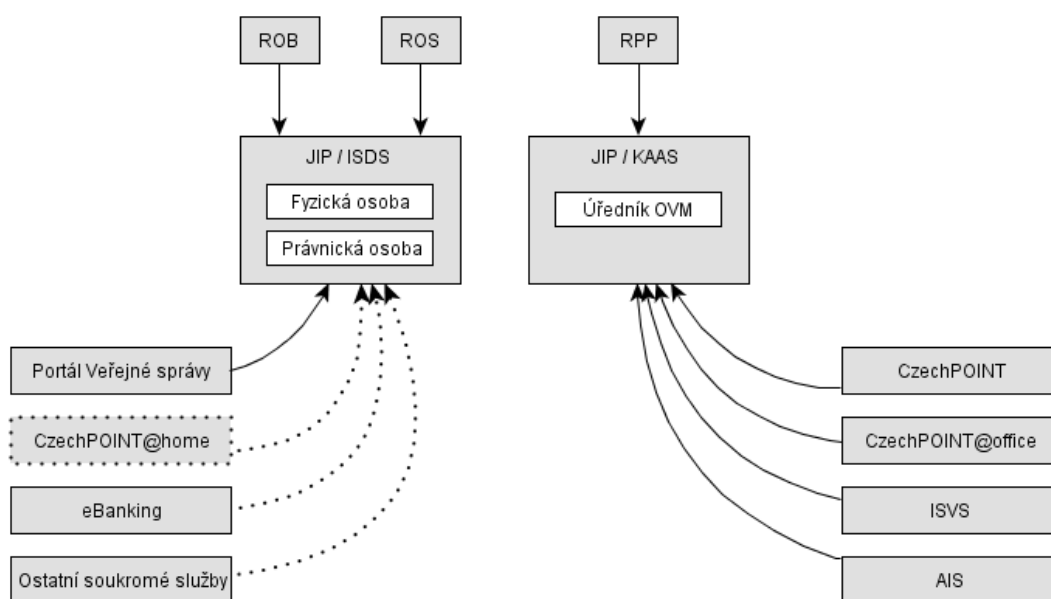
6.1.1 Sdílené služby katalogu dat

Katalogy mají řadu společných vlastností a vyžadují podobnou funkcionalitu, proto bude znovupoužitelná funkcionalita poskytována globálně. Jednotlivé společné oblasti jsou vystiženy v následující tabulce.

| Služba | Popis |
|----------------------------|---|
| Jednotný identitní prostor | Zabezpečená adresářová služba obsahující údaje pro autentizaci a autorizaci uživatelů. Je použita pro identifikaci jednotlivých uživatelů pracujících v katalogu. |
| Správa metadat | Metadata datových katalogů, která jsou znovupoužitelná budou částečně společná a budou podléhat centralizované správě. Účelem této centralizace je snazší integrovatelnost zdrojů v podřazených katalogích do katalogů nadřazených. |
| Ověření existence datasetu | Služba ověřující existenci a některé vlastnosti datasetu v průběhu jeho katalogizace v datovém katalogu. |

Tabulka 13. Sdílené služby katalogu otevřených dat

Veřejná správa ČR provozuje dva odlišné Jednotné identitní prostory a to JIP/ISDS¹⁹⁵ a JIP/KAAS¹⁹⁶. Kde první je postaven pro identifikaci fyzických a právnických subjektů mimo VS a je provázán s existencí DS v ISDS a druhý pro identifikaci úředníku VS postavený nad RPP. Konceptuální schéma je zobrazeno na následujícím diagramu.



Obrázek 47. Konceptuální schéma Jednotného identitního prostoru VS

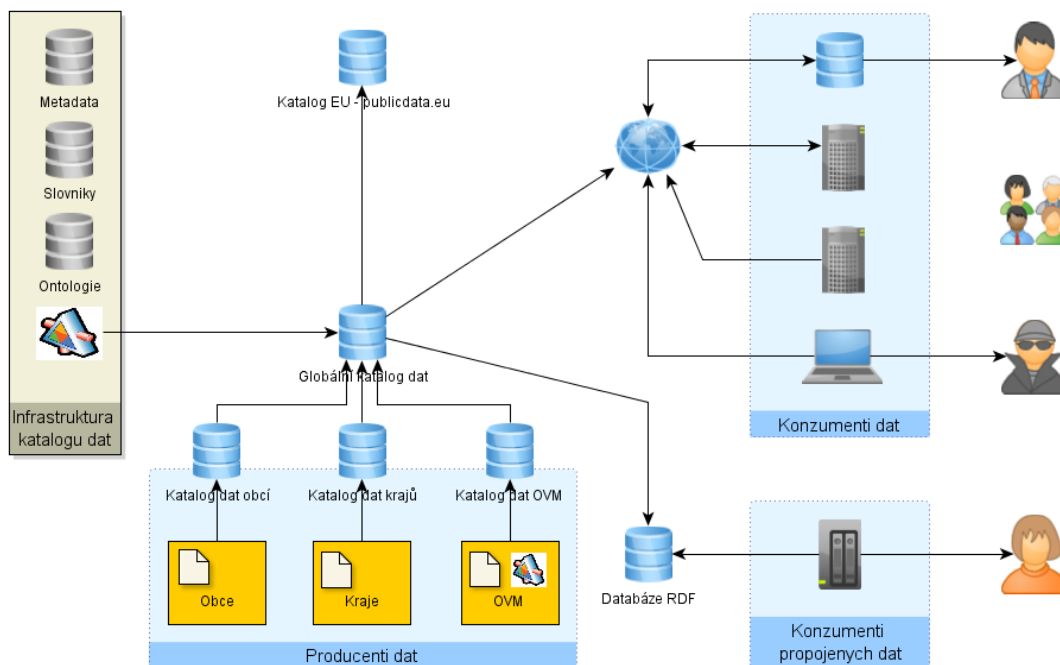
Čárkované vazby a CzechPOINT@home nejsou v době psaní textu realizovány, ale počítá se s nimi v horizontu tří let. Datový katalog bude využívat služeb obou identitních prostorů a to tak že JIP/KAAS bude sloužit pro identifikaci redaktorů a JIP/ISDS pro identifikaci veřejnosti.

¹⁹⁵ Jednotný identitní prostor / informačního systému datových schránek

¹⁹⁶ Jednotný identitní prostor / Katalog autentizačních a autorizačních služeb

6.1.2 Návrh katalogu

Podle analýzy uvedené v kapitole 5 a zhodnocení uvedené v kapitole 5.9 je zvolena distribuovaná varianta datového katalogu. Konceptuální schéma a vazba mezi hlavním datovým katalogem a katalogem dat EU a specializovanými katalogy jakož i napojení na infrastrukturu katalogu dat a způsob konzumace dat je uvedeno v následujícím diagramu.



Obrázek 48. Architektura katalogu dat

Z diagramu je patrné že všechny katalogy mají společnou základnu metadat. Specializované katalogy publikují své záznamy do katalogu globálního, který je napojen na katalog EU. Konzumenti data mohou pomocí katalogů data vyhledávat, případně je využít k extrakci sémantických informací ve formě otevřených-propojených dat.

6.1.3 Návrh ontologie

Návrh ontologie je velmi složitý a komplexní problém. Proces návrhu můžeme shrnout takto:

1. Vymezení rozsahu a účelu ontologie
2. Identifikace entit specifických v dané doméně
3. Uspořádání entit do hierarchie
4. Nalezení průniku naší domény s doménou popsanou v již existující ontologii
5. Definice entit a jejich vlastností
6. Identifikace vztahů

6.1.4 Principy tvorby URI

Na základě analýzy existujících přístupů v jiných zemích (zejména Velké Británie) a současného stavu a potřeb v ČR je stanovena tato metodika pro tvorbu URI.

Principy vychází především z doporučení Chief Technology Officer Council sepsanými v dokumentu (19).

Obecná doporučená struktura pro URI je následující:

http:// {oblast} . {domena} / id / {typ objektu} / {identifikátor objektu}

Data, která jsou pod přímou kontrolou VS a mají globální povahu, případně nejsou v konfliktu budou využívat jako doménu portál otevřených dat a to: data. gov.cz. Oblast bude blíže určovat doménu, o které data hovoří, jako např. adresa, doprava apod..

V případě že se jedná o oblast, která není pod přímou kontrolou veřejné správy, dochází ke konfliktu více producentů nebo se jedná o rozsáhlá a členitá data je vhodné je publikovat pod jinou specializovanou doménou. Zde se jedná např. o oblast veřejných zakázek, rozpočtů a dalších.

Zásady navrhování URI

Následující zásady byly odvozeny ze stávajících osvědčených postupů a revidovány pro potřeby užití ve veřejné správě ČR. Zakládá se především na principech tvorby URI uvedeném v dokumentu (19).

| Zásada | Vyžadován |
|--|-----------|
| Použití protokolu HTTP takovým způsobem aby URI mohlo být dosaženo, tzv. aby pojmenovaná entita mohla být vyhledána. | Vyžadován |
| Použití konzistentní struktury, tak aby typ URI byl jednoznačně identifikován. | Doporučen |
| Struktura URI je vhodná pro znovupoužití jinými orgány veřejné správy nebo veřejnosti. | Doporučen |
| Autor struktury URI uvede jaká je životnost a možnosti znovupoužitelnosti URI. | Vyžadován |
| Ty URI, které jsou navrženy pro znovupoužití, jsou navrženy tak aby jejich životnost byla minimálně 10 let. | Doporučen |
| V případě více typů obsahu pro jedno URI, je vhodný obsah vrácen pomocí metody content negotiation ¹⁹⁷ . | Doporučen |
| Struktura URI neobsahuje technické specifika a implementaci. | Doporučen |
| Struktura URI bude trvalá a neměnná, nebude obsahovat žádné dočasné části jako například sessionID, apod.. | Vyžadován |
| Je-li to vhodné je struktura URI „lidsky“ srozumitelná. | Doporučen |

Tabulka 14. Zásady navrhování URI

¹⁹⁷ Metoda sjednání obsahu, je standardní složka komunikace mezi (webovým) serverem a klientem k určení vhodného typu dokumentu, jako je například jazyková verze nebo formát dokumentu.

6.1.5 Struktura URI pro konkrétní a abstraktní typy zdroje

Obecné „věci“ je vhodné identifikovat genericky a přímo k doméně data.gov.cz dle následující tabulky. Některé entity nemají jednoznačně (centrálně) určitelný identifikátor a ten jim proto bude muset být přiřazen až v průběhu. Tyto entity se budou zpravidla identifikovat vůči svému originálnímu zdroji. Pro konkrétní entity je doporučena struktura URI uvedená v následující tabulce.

| Typ zdroje | Formát URI |
|--|--|
| Adresa | Adresa.data.gov.cz/{KodAdresnihoMista} Kód adresního místa je jedinečný stálý identifikátor z RÚIAN dostupný prostřednictvím VDP. |
| Územní jednotka (obec ¹⁹⁸ , městská část, obvod, kraj, ulice ¹⁹⁹) | Lokace.data.gov.cz/{KodUzemniJednotky} Kód územní jednotky je jedinečný stálý identifikátor z RÚIAN dostupný prostřednictvím VDP. |
| Účelový územní prvek (volební okrsek, ...) | Lokace.data.gov.cz/{KodUcelovehoUzemnihoPrvku} Kód účelového územního prvku je jedinečný stálý identifikátor z RÚIAN dostupný prostřednictvím VDP. |
| Silnice, dálnice | Doprava.data.gov.cz/komunikace/{ID} Kde ID představuje identifikátor přidělený pozemní komunikaci, například D1, R68. |
| Stavební objekt | Stavba.data.gov.cz/{ID_objektu} ID_objektu je jedinečný stálý identifikátor z RÚIAN dostupný prostřednictvím VDP. |
| Orgán veřejné moci | Data.gov.cz/ovm/{ICO} IČO je identifikační číslo přidělené |
| Útvar orgánu veřejné moci | Data.gov.cz/ovm/{ICO}/{ZkratkaUvaru} Zkratka útvaru je označení jednotlivého útvaru dle organizačního řádu daného úřadu (dle spisového a skartačního řádu). |

Tabulka 15. Struktura URI pro konkrétní typy zdroje

¹⁹⁸ Podle zákona 128/2000 Sb., o obcích je obec základním územním samosprávným společenstvím občanů; tvořící územní celek, který je vymezen hranicemi území obce. Pod pojem spadá město, městys, obec, vojenský újezd a další územní celky.

¹⁹⁹ Ulice je členěna podle územních celků, kam spadá, čili jedné ulici může připadat více identifikátorů

V následující tabulce je uvedena doporučená struktura URI pro abstraktní typy zdroje.

| Typ zdroje | Formát URI |
|-------------------------------------|---|
| Abstraktní věci uvedené ve slovníku | Data.gov.cz/{typ_slovníku}/{ID} Kde typ_slovníku je jedinečný identifikátor konkrétního slovníku a jeho případné verze a ID je jedinečný identifikátor záznamu ve slovníku katalogu dat. |
| Životní situace | portal.gov.cz/portal/obcan/situace/{ID kategorie}/{ID podkategorie}/ID situace Kde ID kategorie, podkategorie a konkrétní životní situace jsou identifikátory přidělené v rámci portálu veřejné správy dle interní metodiky. Např. pro životní situaci „vydání občanského průkazu bude URI http://portal.gov.cz/portal/obcan/situace/152/160/4279 |
| Datum a čas | http://placetime.com/instant/gregorian/{datum_a_čas} Kde datum_a_čas je ve formátu yyyy-mm-ddThh:nn:ssTZ ²⁰⁰ . |
| Časový interval | http://placetime.com/interval/gregorian/{datum_a_čas}/{trvání} Formát položky datum_a_čas je vysvětlen v předchozím záznamu. Položka trvání je ve formátu PyYmMdDThHmMsS kde P označuje začátek celé sekvence, y / m / d / h / m / s je počet let / měsíců / dnů / hodin / minut / sekund a Y / M / D / H / M / S oddělují jednotlivé položky. |
| Světové strany | Pro sever: http://dbpedia.org/resource/North Pro jih: http://dbpedia.org/resource/South Pro východ: http://dbpedia.org/resource/East Pro západ: http://dbpedia.org/resource/West |

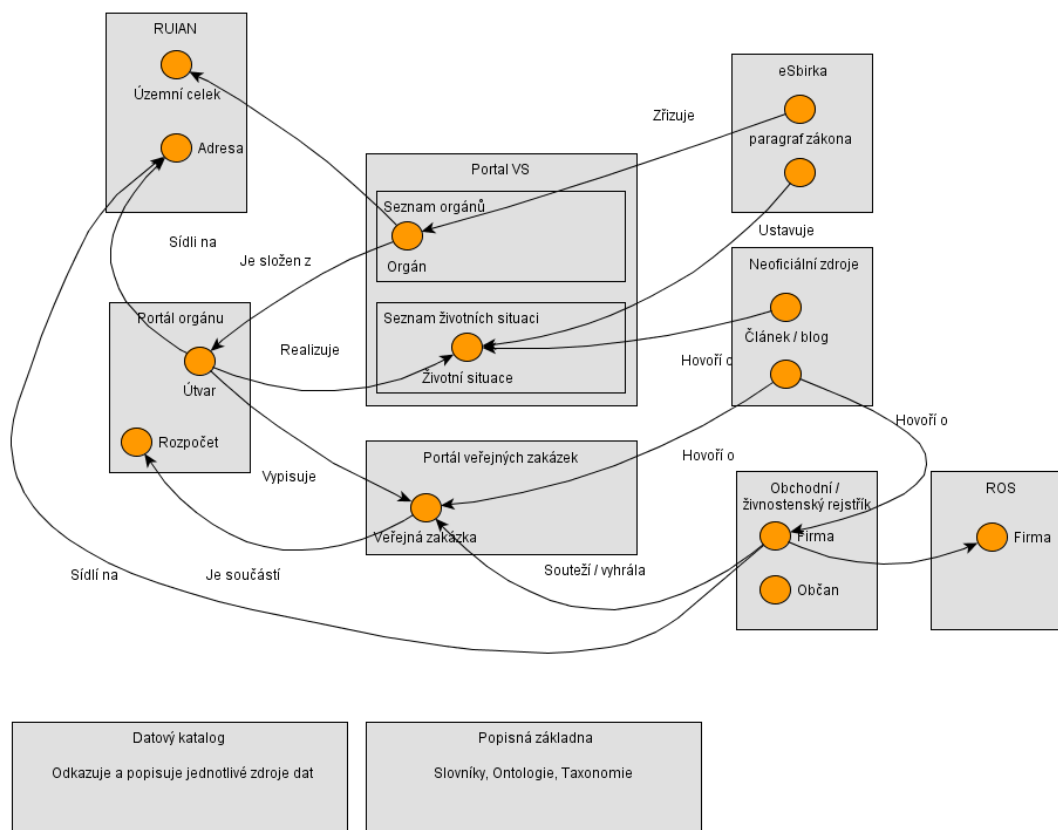
Tabulka 16. Struktura URI pro abstraktní typy zdroje

6.2 Datový model otevřených-propojených dat

Po označení jednotlivých entit jedinečnými identifikátory (URI) je dalším krokem propojení entit mezi sebou tak, aby data byla čerpána z primárního zdroje a zároveň se data uváděla do širšího kontextu. Díky tomu vznikne rozsáhlá globální síť propojených dat. Výřez možné datové struktury je uveden na následujícím diagramu. Diagram zachycuje producenty dat a jejich datové zdroje a možná propojení těchto datových zdrojů mezi sebou. V modelu je zachycena oblast veřejných zakázek a životních situací. Ústředním zdrojem informací je portál veřejné správy, který poskytuje informace o jednotlivých OVM, a také seznam životních situací. Detailní informace

²⁰⁰ Detailní popis formátu a případných dalších podob je uveden na <http://placetime.com/instant/gregorian/>.

OVN jsou pak dostupné na jejich samostatných portálech. Veřejné zakázky se zveřejňují na portálu veřejných zakázek. Subjekty, které jsou ve vztahu k veřejné zakázce (uchazeči, realizátoři, apod.) jsou dostupní z obchodního nebo živnostenského rejstříku. Každá entita může být doplněna (např. komentována) na externích zdrojích (např. v internetových novinách apod.). Nezastupitelnou roli zde hrají základní registry, především RUIAN, který referencuje veškeré adresy, lokace apod. Vazba na ROS je neveřejná, ale poslouží oprávněným subjektům VS.



Obrázek 49. Zjednodušený datový model otevřených-propojených dat

V rámci celé datové infrastruktury vystupuje „Popisná základna“, která obsahuje ontologie, slovníky a taxonomie jako základ pro metadata. Díky tomuto popisu je možné propojení a integrace dat. Pro vyšší dohledatelnost je zde datový katalog, který odkazuje na zajímavé datové zdroje.

7 Praktické řešení

7.1 Mapa Czech POINT

V rámci praktické části diplomové práce – odstranění duplicit dat ve veřejné správě ČR, jsem připravil projekt „Mapa Czech POINT“ dále jen mapa CzP. Projekt původně vznikl jako záměr MVČR zveřejnit informace o umístění jednotlivých poboček CzP. Vzhledem k úsporným opatřením bylo rozhodnuto, že mapa nebude vybudována z prostředků MV, ale data budou poskytnuta soukromým provozovatelům mapových aplikací. Tato skutečnost paradoxně zvýšila hodnotu tohoto projektu a přiblížila jej k myšlence otevřených-propojených dat. Během svého působení v odboru Hlavního architekta eGovernmentu jsem projekt upravil na plnohodnotnou službu postavenou na otevřených-propojených datech. Ke dni²⁰¹ psaní tohoto textu je záměr schválen, předpokládá se, že bude realizován v posledním čtvrtletí roku 2012.

7.1.1 Zadání projektu

Poskytnout veřejnosti / subjektům veřejné správy informace o pobočkách kontaktního místa veřejné správy - Czech POINT. Především pak o jejich adrese, umístění na mapě, otevírací době a poskytovaných službách.

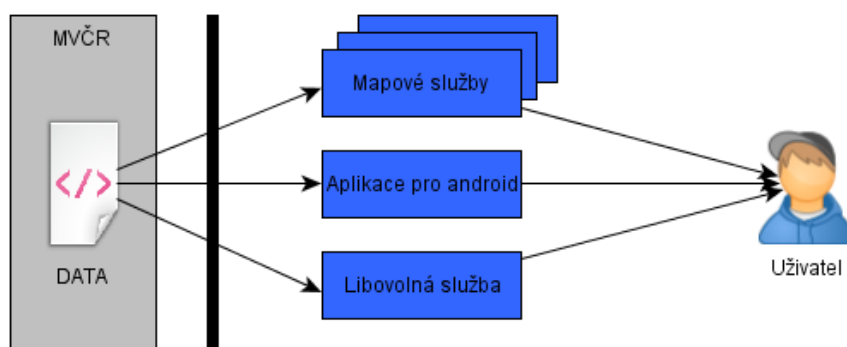
7.1.2 Východiska

Prostřednictvím hlavních poskytovatelů mapových aplikací²⁰² není informace o pobočkách CzP aktuálně přístupná. Informace o pobočkách CzP je pouze na portálu CzP²⁰³ ve formě tabulky s adresou.

7.1.3 Provedení

Informace o jednotlivých CzP budou fyzicky umístěny na datových serverech Ministerstva vnitra. Správa dat - aktualizace bude podléhat provozovateli CzP. Aplikace a služby nad daty budou v plné zodpovědnosti třetích stran.

Záměr si můžeme představit takto:



Obrázek 50. Architektura publikace mapy veřejné správy

²⁰¹ Projekt byl projednán a bylo doporučeno v něm pokračovat na jednání dne 10. 4. 2012

²⁰² Mapy.cz, maps.google.cz.

²⁰³ <http://www.czechpoint.cz/web/?q=node/62>.

Data budou poskytována jako dataset ve formátu XML. Pro jednotlivé pobočky CzP bude existovat záznam s těmito údaji:

- GPS souřadnice
- Kód adresního místa (identifikátor z RUIAN)
- Identifikátor provozujícího subjektu
- Název provozujícího subjektu
- Název pobočky Czech POINT
- Otevírací doba
- Poskytované služby

7.1.4 Změny v souvisejících systémech

Část informací je možné získat přímo z portálu VS - bude muset dojít k následujícím úpravám funkčnosti:

- Dekódování URI: přístup na informace konkrétního OVM dle jeho ID
- Poskytnutí informací o subjektu ve formátu XML nebo HTML/RDFa.

7.1.5 Návaznost projektu na jiné záměry v gesci vlády nebo MV

Projekt svým provedením přímo zapadá do koncepce projektu OpenGov²⁰⁴, který vychází z aktuálního akčního plánu Vlády ČR. Využívá data ze základních registrů. Volitelně pak využívá portál veřejné správy.

V současnosti jsou informace připraveny ve formátu GPX ze základních údajů takto:

```
<?xml version="1.0" encoding="UTF-8" ?>
<gpx version="1.1" xmlns="http://www.topografix.com/GPX/1/1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.topografix.com/GPX/1/1
  http://www.topografix.com/GPX/1/1/gpx.xsd">
  <wpt lat="49.7865" lon="18.52425">
    <name>Obecní úřad Albrechtice</name>
    <desc>Albrechtice Obecní 186, 73543 Albrechtice</desc>
  </wpt>
  <wpt lat="49.70014" lon="18.09477">
    <name>Obecní úřad Albrechtický</name>
    <desc>Albrechtický 131, 74255 Albrechtický</desc>
  </wpt>
  <wpt lat="50.05997" lon="17.389">
    <name>Městský úřad Andělská Hora</name>
    <desc>Andělská Hora 197, 79332 Andělská Hora</desc>
  </wpt>
```

Obrázek 51. Zamýšlený stav publikace poboček CzechPOINT (bez rozšíření) ve formátu GPX

7.2 Případová studie – duplicita dat

Jedním se závažných nedostatků veřejné správy a důvodem pro její nízkou efektivitu je špatný přístup (správa) informací. Není výjimkou, že konkrétní informace má svého původce-správce a je využívána dalšími subjekty, ať už veřejnou správou nebo veřejností. Zpravidla však dochází k tomu, že publikované informace jsou

²⁰⁴ Pracovní název projektu Otevřených dat VS, v době psaní textu nebyl oficiálně schválen.

kopírovány a zveřejnovány i na jiných portálech (např. krajských, obecních, a dalších úřadů). Tím dochází nejenom k duplicitě ale i k nekonzistentnosti. Informace je nutné aktualizovat na více místech, a to zpravidla jinými osobami dle práv přístupu. Zajištění aktuálnosti informací je značně komplikované.

Typickým a ukázkovým příkladem jsou informace o orgánech veřejné moci. Informace jako jsou kontaktní údaje, otevírací doba apod., je povinně zveřejnována na portálu veřejné správy²⁰⁵. Portál veřejné správy publikuje ověřené a zaručené informace. V mnoha oblastech působnosti se jedná o referenční zdroj informací. Stejně informace (a mnohdy pouze jako čistá kopie) jsou publikovány na krajských a obecních portálech, případně na specializovaných stránkách úřadu.

The screenshot shows a web page from the public administration portal. The breadcrumb trail at the top reads 'KATASTRÁLNÍ ÚŘAD PRO MORAVSKOSLEZSKÝ KRAJ > ZÁKLADNÍ ÚDAJE'. A search bar contains 'Karta subjektu: Základní údaje' with a 'Přejít' button. The main heading is 'Orgány státní správy' followed by 'Katastrální úřad pro Moravskoslezský kraj'. The page is divided into two columns of data:

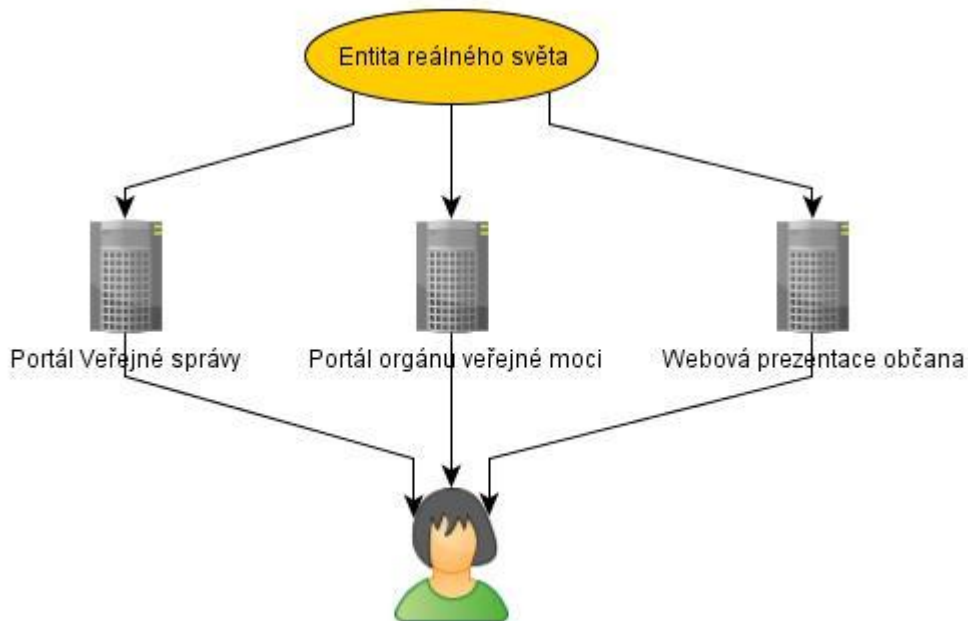
| | | | |
|-------------------|---|--------------------------------|--|
| Název: | Katastrální úřad pro Moravskoslezský kraj | Telefon: | +420 553 698 111 +420 553 698 216 (fax) |
| Adresa sídla: | Praskova 11/194 Opava 74655 | Úřední hodiny: | PO: 8:00 - 17:00 ÚT: 8:00 - 15:00 ST: 8:00 - 17:00 ČT: 8:00 - 15:00 PÁ: 8:00 - 12:00 |
| IČ: | 00849871 | Identifikátor datové schránky: | bxwadpe |
| DIČ: | CZ00849871 | | |
| Bankovní spojení: | 19-10027821 / 0710 | | |
| Kód organizace: | KUMkslszkK | | |
| Typ instituce: | Orgán státní správy | | |
| WWW: | http://www.cuzk.cz/Dokument.aspx?...UVOD | | |
| E-mail: | ku.promoravkraj@cuzk.cz (podatelna) | | |

Obrázek 52. Portál veřejné správy, informace o subjektech VS

Zjednodušeně si můžeme modelovou situaci představit následovně.

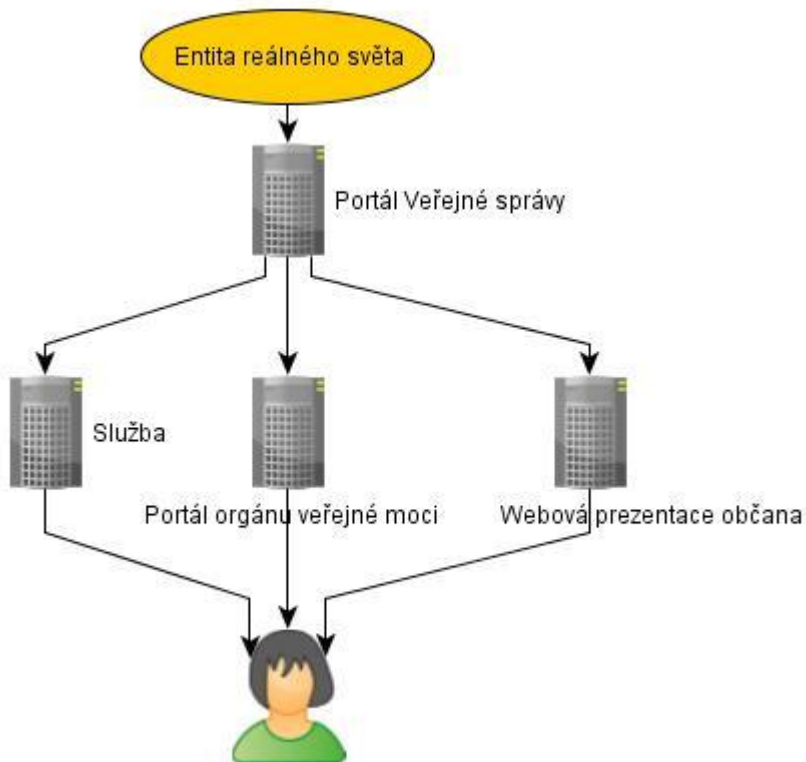
Současný stav: Jedna entita reálného světa popsána několika stejnými datovými entitami s (možností) rozdílné prezentace.

²⁰⁵ Povinnost vychází z toho, že každý OVM má zákonnou povinnost mít zřízenou datovou schránku. A portál VS publikuje seznam všech datových schránek. Součástí publikace je vedle zákonných údajů i doplňující informace jako úřední hodiny, apod.



Obrázek 53. Současný stav publikace dat

Očekávaný stav: Jedna entita reálného světa popsána právě jednou datovou entitou publikovanou portálem VS. Ostatní portály mají možnost si entitu načíst a dle svého uvážení prezentovat.



Obrázek 54. Cílový stav publikace dat

Tohoto stavu je možné docílit pouze za předpokladu, že informace o kterých hovoříme, budou publikovány v souladu s principy otevřených-propojených dat popsanými v této práci.

Ideální způsob je

- Každá potřebná (sada) informace je jednoznačně identifikována
- Informace je dobře strojově čitelná
 - o Služba poskytující informace nabízí vhodné rozhraní pro čtení dat
 - o Informace je popsána slovníkem, ontologií nebo jinými vhodnými metadaty

Portál, který není zdrojem originálních informací a potřebuje je u sebe publikovat si data stahuje z originálního zdroje. Díky tomu informace fyzicky existuje pouze na jednom místě nedochází k duplicitám a je vyloučena nekonzistence. Na druhou stranu to klade větší nároky na zdroje originálních informací a nutnou úpravu stávajících systémů.

V modelovém případě by to znamenalo dvě konkrétní úpravy, a to:

- Portál VS poskytne rozhraní pro přístup k držitelům datových schránek
- Webové prezentace úřadů potřebné údaje získávat z tohoto rozhraní

Vše ostatní, jako je aktualizace dat na portálu VS apod., je v tuto chvíli v běžném provozu.

7.3 Případová studie – Životní situace

Pro běžného člověka je na veřejné správě nejvíce zatěžující to že se velice špatně shánějí informace o tom, co a jak má v jaké situaci udělat²⁰⁶.

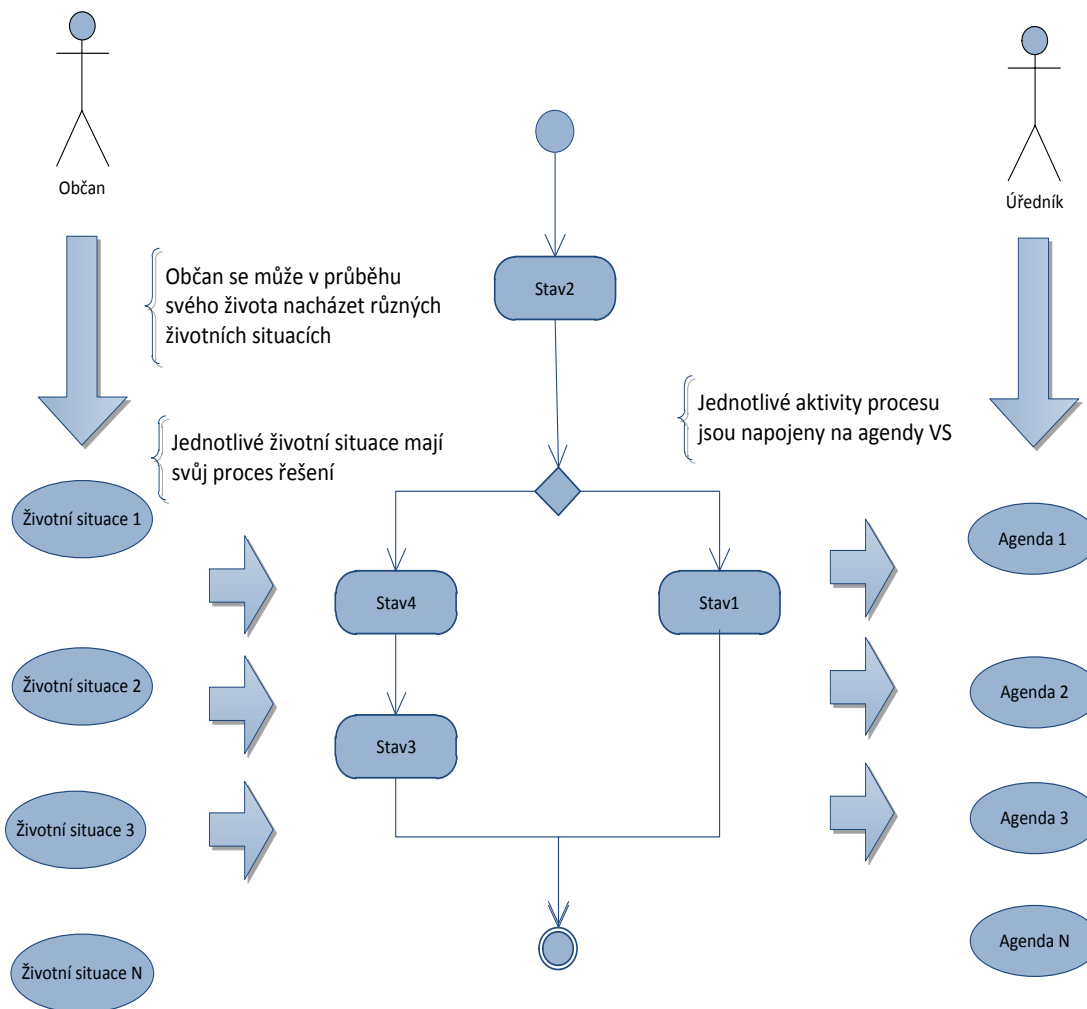
Přínos a využití otevřených propojených dat ukážeme na případu životních situací občana. Portálu veřejné správy a jeho sekce životní situace²⁰⁷.

Občan se v průběhu svého života nachází v řadě životních situací, které musí řešit. Nalezení informací vedoucí k vyřešení dané informace je však zpravidla málo a jsou nedostatečné. Základní informace jsou zveřejněny na portálu VS.

Zjednodušeně si celou problematiku můžeme představit takto: Ke každé životní situaci nabízí portál informace poskytující základní informace vedoucí k vyřešení dané situace. K vyřešení situace musí občan projít procesem, do kterého v jeho průběhu dodává artefakty (dokumenty, žádosti, osvědčení, ...) a kde jednotlivé aktivity procesu spadají (jsou realizovány) do agendy VS, kterou vykonává úřad. Proces je naznačen na následujícím diagramu:

²⁰⁶ Michal Rada, vedoucí iniciativy Informatika pro občany, Informační bulletin ÚOOÚ1/2012.

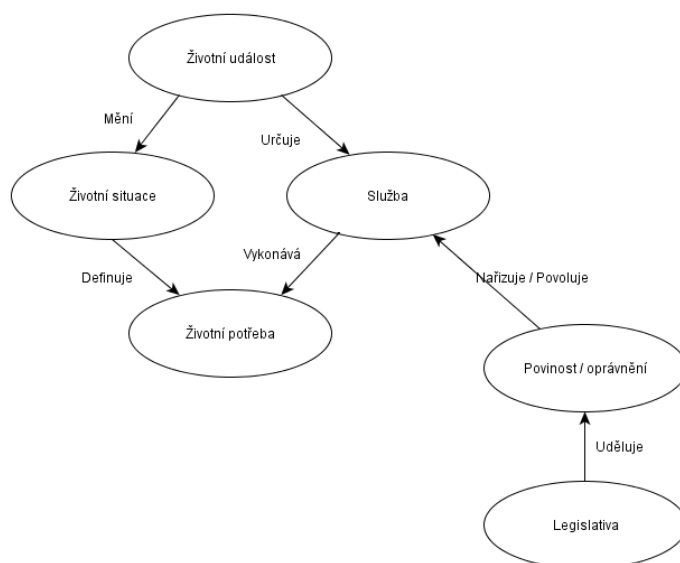
²⁰⁷ Dostupné na adrese portal.gov.cz/portal/obcan/situace/.



Obrázek 55. Proces realizace životní potřeby vůči veřejné správě

Kamenem úrazu celé této problematiky je fakt že (oficiálních) informací vedoucí k vyřešení dané situace je málo, případně nepopisují náš konkrétní případ. Na druhou stranu podobný případ jako my už někdo mohl řešit a mohl jej popsat ve svém soukromém webovém prostoru (webu, blogu, novinovém článku, apod.). Tuto skutečnost my však snadno nezjistíme, protože neexistuje jednoznačná vazba mezi „životní situací“ a neoficiálním zdrojem. V případě že chce subjekt mimo Portál VS hovořit o objektech Portálu VS má značně komplikovanou situaci protože jednotlivé objekty nemají přiřazeno jedinečné URI.

Ve velké Británii tuto problematiku na velmi profesionální úrovni řeší ESD toolkit popsany v kapitole 3.4.4. Ten identifikuje jednotlivé fáze a okolí procesu a snaží se jim přiřadit jednoznačný identifikátor. Vazba jednotlivých fází a okolí je naznačena na následujícím diagramu.



Obrázek 56. Vazba mezi entitami životní situace

Jednotlivé entity je možné (ale pracné) jednoznačně identifikovat a popsat vazby mezi nimi. Díky tomu nám vznikne síť, která říká „V mém životě nastala tato životní událost – budu ji řešit touto službou“

Životní situace jsou rozděleny do tří základních skupin podle typu použití na informace pro:

- občany České republiky
- Podnikatele a živnostníky
- Cizince žijící v České republice

Portál mimo jiné zveřejňuje sadu životních situací ve které se může občan nacházet. Popis sekce životní situace nalezneme na portálu VS²⁰⁸. Ke dni 1. 4. 2012 bylo na portálu uveřejněno v sekci občané 12 kategorií a 106 podkategorií s více než 500 životními situacemi. Kompletní přehled životních situací je uveden v příloze. Životní situace je strukturovaná do 29 položek, které popisují většinu potřebných informací k řešení dané životní situace. Seznam položek s popisem je uveden v příloze.

²⁰⁸ Na adrese <http://portal.gov.cz/portal/obcan/situace/>.

8 Hodnocení a závěr

Technologie a principy OpenData a LinkedData představují klíčový aspekt zvýšení efektivity a přístupnosti veřejné správy. Jejich postupné zavádění umožní nejenom zvýšení stávající hodnoty informací, ale především díky uvedení informací do širšího kontextu a možnosti znovupoužitelnosti, také plné využití hodnoty. Principy OpenData a LinkedData nemají vliv jenom na ekonomickou a provozní stránku veřejné správy, ale v neposlední řadě také dávají příležitost zapojení veřejnosti do správy věcí veřejných a možnosti sledovat, kontrolovat i připomínkovat veřejnou činnost. To vše v konečném důsledku vede k posílení svobody jednotlivce a v souhrnu celé demokratické společnosti.

Cílem této práce bylo provést analýzu přístupu k publikaci dat veřejnou správou s důrazem na principy OpenData a LinkedData, vypracovat analýzu současného stavu jak podle používaných technologií, tak podle konkrétních přístupů ve vybraných zemích a dále provést zhodnocení současného stavu a navrhnou vhodný postup pro zavedení těchto principů ve veřejné správě ČR.

V úvodu je čtenář uveden do problematiky OpenData a LinkedData. Jsou zde zmíněny nejdůležitější důvody proč principy zavádět a jejich východiska. V další části textu jsou vymezeny nejvýznamnější pojmy používané v této práci a související s principy OpenData a LinkedData. Dále pokračuje analytická část, která rozpracovává popis současného stavu řešeného problému jak z hlediska používaných technologií, tak i z pohledu přístupu k OpenData a LinkedData ve vybraných zemích. Úvod je zakončen zhodnocením současných praktik publikace dat ve veřejné správě ČR.

Další část se zabývá řešením samotného problému a je tvořena třemi kapitolami. První kapitola obsahuje popis řešení jako celku, kde jsou identifikovány hlavní problémy a kde je navrženo jejich řešení. Poté následuje kapitola Katalog otevřených dat, která hodnotí jednotlivé přístupy pro katalogizaci a doporučuje nejvhodnější variantu. Trojici uzavírá kapitola Architektura, která navrhuje globální architekturu publikace dat a základní datový model této publikace.

Poslední část této práce se věnuje praktickému řešení a ukazuje zavedení principů OpenData a LinkedData na konkrétních případech.

Cíle vytýčené v úvodu práce se podařilo splnit, nicméně je zde stále prostor pro upřesnění a rozšíření, který by ale již přesahoval rozsah této práce. Návrh architektury popsany v práci je pouze globálním pohledem. Vzhledem k rozsáhlosti a složitosti veřejné správy je prakticky nemožné popsat na tak omezeném prostoru ucelené řešení. Část architektura je natolik zajímavá, že si zaslouží pozornost a budoucí rozpracování v samostatné publikaci.

Citovaná literatura

1. **Berg, Michal, a další, a další.** *Otevřená data ve státní správě: Nová éra rozhodování.* Praha : Nadace Open Society Fund Praha, 2012. ISBN: 978-80-87110-24-9.
2. **Vláda ČR.** Akční plán České republiky - "Partnerství pro otevřené vládnutí". *Vláda ČR.* [Online] 4. Duben 2012. [Citace: 6. Duben 2012.] <http://www.vlada.cz/assets/clenove-vlady/pri-uradu-vlady/karolina-peake/tiskove-zpravy/Akcni-plan-OGP.pdf>.
3. **Evropské společenství.** *Řeč pro Evropu - Jazyky v Evropské unii.* 2008. Dostupné online: http://ec.europa.eu/education/languages/pdf/doc3275_cs.pdf. 978-92-79-09155-1.
4. **Vláda ČR.** Úvodní teze k akčnímu plánu. *Vláda ČR.* [Online] 2012. <http://www.vlada.cz/assets/ppov/boj-s-korupci/otevrene-vladnuti/OGP---Uvodni-teze-k-Akcni-planu.pdf>.
5. **Kalousková, Eva a Poláková, Jiřina.** Data, informace, znalosti - rozdíly, podrobnosti. *Knowledge Managment.* [Online]
6. **Cejpek, Jiří.** *Informace, komunikace a myšlení - Úvod do informační vědy.* Praha : Univerzita Karlova, vydavatelství Karolinum, 1998. ISBN: 80-7184767-4.
7. **Turban, Efraim.** *Managing expert systems.* 1992. 1-878289-11-X.
8. **Horzinková, Eva a Novotný, Vladimír.** *Základy organizace veřejné správy v ČR.* Plzeň : Vydavatelství a nakladatelství Aleš Čeněk, 2010. 978-80-7380-263-9 (brož.).
9. **Průcha, Petr.** *Správní právo. Obecná část.* Brno : autor neznámý, 2003.
10. **Ministerstvo vnitra ČR, Vyšší policejní škola MV v Praze.** *Studijní texty ke vstupnímu vzdělání úvodnímu nových zaměstnanců MV.* Praha : MVČR, 2012.
11. **Lacy, Lee.** *Representing Information Using the Web Ontology Language.* Trafford : autor neznámý, 2005. ISBN: 141203448-5.
12. **Erl, Thomas.** *SOA Servisně orientovaná architektura Kompletní průvodce.* Brno : Computer Press, a.s., 2009. ISBN 978-80-251-1886-3.
13. **Ministerstvo vnitra.** *Analýza aktuálního stavu veřejné správy.* Praha : autor neznámý, 2011.
14. **Chlápek, Dušan; Kroupa, Tomáš; Kučera, Jan; Nečaský, Martin; Weis, Vladimír.** *Koncepce katalogizace otevřených dat VS ČR.* Praha 2012.
15. **Ministerstvo vnitra.** *Dlouhodobé řízení Informačních systémů veřejné správy.* Dostupné online: <http://www.mvcr.cz/soubor/procesni-model-rizeni-isvs-211984.aspx>.
16. **Office of E-Government and IT; Office of Management and Budget.** *Data.gov Concept of Operations .* Washington : autor neznámý, 2010.

17. **Design Coordination Group.** *UK Location Programme - Technical Architecture Overview.*
18. **Ministerstvo vnitra ČR.** *Metodika tvorby XML schémat v oblasti informačních systémů veřejné správy.*
19. **Chief Technology Officer Council.** *Designing URI Sets for the UK Public Sector.* London : autor neznámý, 2010.
20. **Vlada ČR.** Strategie realizace Smart Administration v období 2007-2015. [Online] 2006. http://www.smartadministration.cz/file/9_1_1/download/.
21. **Delquie, Philippe.** *Valuing Information and Options: an Experimental Study.* 2004. Dostupné online: <http://www.insead.com/facultyresearch/resarch/doc.cfm?did=1355>.

Seznam použitých tabulek

| | |
|--|----|
| TABULKA 1. ÚŘEDNÍ JAZYKY VYBRANÝCH MEZINÁRODNÍCH ORGANIZACÍ..... | 4 |
| TABULKA 2. HARMONOGRAM REALIZACE AKČNÍHO PLÁNU VLÁDY | 7 |
| TABULKA 3. PŘIPOJENÍ ORGANIZAČNÍCH SLOŽEK STÁTU A OBCÍ K SÍTI INTERNET | 8 |
| TABULKA 4. ATRIBUTY KVALITY DAT | 12 |
| TABULKA 5. POUŽITÉ POJMY | 29 |
| TABULKA 6. POUŽITÉ ZKRATKY..... | 31 |
| TABULKA 7. VLASTNOSTI FORMÁTŮ DAT..... | 37 |
| TABULKA 8. POPIS A ZHODNOCENÍ FORMÁTŮ DAT | 38 |
| TABULKA 9. UKÁZKA ŽIVOTNÍ SITUACE | 52 |
| TABULKA 10. KVALITATIVNÍ ATRIBUTY SW ARCHITEKTURY | 70 |
| TABULKA 11. KVALITATIVNÍ ATRIBUTY PUBLIKACE DAT..... | 70 |
| TABULKA 12. VYUŽITÍ KATALOGU OTEVŘENÝCH DAT | 81 |
| TABULKA 13. SDÍLENÉ SLUŽBY KATALOGU OTEVŘENÝCH DAT..... | 85 |
| TABULKA 14. ZÁSADY NAVRHOVÁNÍ URI | 87 |
| TABULKA 15. STRUKTURA URI PRO KONKRÉTNÍ TYPY ZDROJE | 88 |
| TABULKA 16. STRUKTURA URI PRO ABSTRAKTNÍ TYPY ZDROJE | 89 |

Seznam použitých diagramů

| | |
|--|----|
| OBRÁZEK 1. VYUŽÍVÁNÍ SLUŽEB NA BÁZI E-BANKING VE SVĚTĚ | 9 |
| OBRÁZEK 2. JEDINCI, KTERÍ VYUŽÍVAJÍ INTERNET KE KOMUNIKACI S VEŘEJNOU SPRÁVOU V ROCE 2010 | 9 |
| OBRÁZEK 3. JEDINCI, KTERÍ POUŽÍVAJÍ INTERNETU K PODÁNÍ VYPLNĚNÝCH FORMULÁŘŮ VEŘEJNÉ SPRÁVY V ROCE 2010 | 10 |
| OBRÁZEK 4. VYUŽÍVÁNÍ INFORMACÍ Z WEBOVÝCH PREZENTACÍ ÚŘADU | 10 |
| OBRÁZEK 5. DIAGRAM VZTAHU MEZI TYPY DAT | 14 |
| OBRÁZEK 6. ARCHITEKTURA PUBLIKACE OTEVŘENÝCH-PROPOJENÝCH DAT | 17 |
| OBRÁZEK 7. ORGANIZAČNÍ VYMEZENÍ VEŘEJNÉ SPRÁVY | 18 |
| OBRÁZEK 8. ORGANIZAČNÍ STRUKTURA VEŘEJNÉ SPRÁVY NA MÍSTNÍ ÚROVNI (OBEC) | 19 |
| OBRÁZEK 9. ZNÁZORNĚNÍ ONTOLOGIE | 19 |
| OBRÁZEK 10. VZTAHY MEZI INSTANCÍ, TŘÍDOU A VLASTNOSTÍ V ONTOLOGII | 20 |
| OBRÁZEK 11. UKÁZKA POUŽITÍ JAZYKA RDFS | 21 |
| OBRÁZEK 12. VZTAH ONTOLOGIE A SÉMANTICKÉHO WEBU | 21 |
| OBRÁZEK 13. ARCHITEKTURA SÉMANTICKÉHO WEBU | 22 |
| OBRÁZEK 14. ARCHITEKTURA SÉMANTICKÉHO WEBU | 22 |
| OBRÁZEK 15. SLUŽBA WOLFRAM ALPHA | 23 |
| OBRÁZEK 16. SOUČASNÝ STAV PUBLIKACE DAT V HTML | 24 |
| OBRÁZEK 17. OČEKÁVANÝ STAV PUBLIKACE DAT POMOCÍ HTML / RDF A | 24 |
| OBRÁZEK 18. PROCES ŽURNALISTIKY ŘÍZENÉ DATY | 26 |
| OBRÁZEK 19. ARCHITEKTURA SYSTÉMU ZÁKLADNÍCH REGISTRŮ | 35 |
| OBRÁZEK 20. VEŘEJNÁ DATABÁZE ČESKÉHO STATISTICKÉHO ÚŘADU | 41 |
| OBRÁZEK 21. VEŘEJNÁ DATABÁZE ČESKÉHO STATISTICKÉHO ÚŘADU - PŘÍJMY ÚZEMNÍCH CELKŮ | 42 |
| OBRÁZEK 22. VEŘEJNÁ DATABÁZE ČESKÉHO STATISTICKÉHO ÚŘADU - METADATA | 42 |
| OBRÁZEK 23. REGISTR ÚZEMNÍCH IDENTIFIKÁTORŮ - VEŘEJNÝ DÁLKOVÝ PŘÍSTUP | 43 |
| OBRÁZEK 24. REGISTR ÚZEMNÍCH IDENTIFIKÁTORŮ - VEŘEJNÝ DÁLKOVÝ PŘÍSTUP - DETAIL OBJEKTU | 44 |
| OBRÁZEK 25. LOGO INICIATIVY OPENDATA.CZ | 44 |
| OBRÁZEK 26. PORTÁL OTEVŘENÝCH DAT USA - DATA.GOV | 47 |
| OBRÁZEK 27. PROCES PUBLIKACE OTEVŘENÝCH DAT V USA | 48 |
| OBRÁZEK 28. KONCEPTUÁLNÍ ARCHITEKTURA DATA.GOV | 48 |
| OBRÁZEK 29. STRUKTURA DATASETU KATALOGU DAT DATA.GOV | 49 |
| OBRÁZEK 30. ARCHITEKTURA PUBLIKACE DAT VE VELKÉ BRITANII | 50 |
| OBRÁZEK 31. PORTÁL OTEVŘENÝCH DAT EU - PUBLICDATA.EU | 51 |
| OBRÁZEK 32. OBCHODNÍ MODEL VLÁDY PODLE ESD TOOLKITU | 53 |
| OBRÁZEK 33. SLUŽBA CKAN CZECH REPUBLIC | 55 |
| OBRÁZEK 34. DIAGRAM CORE VOCABUARIES | 57 |
| OBRÁZEK 35. STRUKTURA ORGANIZATION ONTOLOGY | 58 |
| OBRÁZEK 36. UKÁZKA POUŽITÍ ONTOLOGIE FOAF V RDF | 59 |
| OBRÁZEK 37. UKÁZKA POUŽITÍ ONTOLOGIE FOAF | 60 |
| OBRÁZEK 38. KONCEPTUÁLNÍ SCHÉMA ONTOLOGIE GOODRELATIONS | 61 |
| OBRÁZEK 39. RÁMEC PROBLEMATIKY OTEVŘENÝCH-PROPOJENÝCH DAT | 66 |
| OBRÁZEK 40. HARMONOGRAM REALIZACE OTEVŘENÝCH-PROPOJENÝCH DAT V ČR | 73 |
| OBRÁZEK 41. KATALOGIZACE - VARIANTA BEZ KATALOGU DAT | 75 |
| OBRÁZEK 42. KATALOGIZACE - VARIANTA S HLAVNÍM KATALOGEM DAT | 75 |
| OBRÁZEK 43. KATALOGIZACE- VARIANTA S DISTRIBUOVANÝMI KATALOGY | 76 |
| OBRÁZEK 44. ROLE UŽIVATELŮ PRO SPRÁVU KATALOGU OTEVŘENÝCH DAT | 78 |
| OBRÁZEK 45. PROCES PLNĚNÍ KATALOGU DATY | 78 |
| OBRÁZEK 46. KONCEPTUÁLNÍ ARCHITEKTURA OTEVŘENÝCH-PROPOJENÝCH DAT VEŘEJNÉ SPRÁVY | 83 |
| OBRÁZEK 47. KONCEPTUÁLNÍ SCHÉMA JEDNOTNÉHO IDENTITNÍHO PROSTORU VS | 85 |
| OBRÁZEK 48. ARCHITEKTURA KATALOGU DAT | 86 |
| OBRÁZEK 49. ZJEDNODUŠENÝ DATOVÝ MODEL OTEVŘENÝCH-PROPOJENÝCH DAT | 90 |
| OBRÁZEK 50. ARCHITEKTURA PUBLIKACE MAPY VEŘEJNÉ SPRÁVY | 91 |
| OBRÁZEK 51. ZAMÝŠLENÝ STAV PUBLIKACE POBOČEK CZECHPOINT (BEZ ROZŠÍŘENÍ) VE FORMÁTU GPX | 92 |
| OBRÁZEK 52. PORTÁL VEŘEJNÉ SPRÁVY, INFORMACE O SUBJEKTECH VS | 93 |

| | |
|--|-----|
| OBRÁZEK 53. SOUČASNÝ STAV PUBLIKACE DAT | 94 |
| OBRÁZEK 54. CÍLOVÝ STAV PUBLIKACE DAT | 94 |
| OBRÁZEK 55. PROCES REALIZACE ŽIVOTNÍ POTŘEBY VŮČI VEŘEJNÉ SPRÁVY | 96 |
| OBRÁZEK 56. VAZBA MEZI ENTITAMI ŽIVOTNÍ SITUACE | 97 |
| OBRÁZEK 57. ORGANIZAČNÍ STRUKTURA MĚSTA | 108 |
| OBRÁZEK 58. ORGANIZAČNÍ STRUKTURA MINISTERSTVA | 108 |

Přílohy

Právní úprava informačních systémů veřejné správy ČR

Zákon č. 365/2000 Sb., o informačních systémech veřejné správy a o změně některých dalších zákonů, ve znění pozdějších předpisů.

Zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů, ve znění pozdějších předpisů.

Zákon č. 227/2000 Sb., o elektronickém podpisu.

Zákon č. 111/2009 Sb., o základních registrech.

Zákon č. 300/2008 Sb., o elektronických úkonech a autorizované konverzi dokumentů.

Vyhláška č. 528/2006 Sb., o formě a technických náležitostech předávání údajů do informačního systému, který obsahuje základní informace o dostupnosti a obsahu zpřístupněných informačních systémů veřejné správy (vyhláška o informačním systému o informačních systémech veřejné správy).

Vyhláška č. 529/2006 Sb., o požadavcích na strukturu a obsah informační koncepce a provozní dokumentace a o požadavcích na řízení bezpečnosti a kvality informačních systémů veřejné správy (vyhláška o dlouhodobém řízení informačních systémů veřejné správy).

Vyhláška č. 530/2006 Sb., o postupech atestačních středisek při posuzování dlouhodobého řízení informačních systémů veřejné správy.

Vyhláška č. 53/2007 Sb., o technických a funkčních náležitostech uskutečňování vazeb mezi informačními systémy veřejné správy prostřednictvím referenčního rozhraní (vyhláška o referenčním rozhraní).

Vyhláška č. 52/2007 Sb., o postupech atestačních středisek při posuzování způsobilosti k realizaci vazeb informačních systémů veřejné správy prostřednictvím referenčního rozhraní.

Doporučující normy

ČSN ISO/IEC 11179-5: 1997 - Informační technologie – Specifikace a normalizace datových prvků – Část 5: Identifikační principy a principy tvorby názvů datových prvků.

URL:<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=1677>

Seznam navržených zdrojů k otevření

| Databáze/dataset | Správce dat |
|----------------------|----------------------------|
| Obchodní rejstřík | Ministerstvo spravedlnosti |
| Insolvenční rejstřík | Ministerstvo spravedlnosti |

| | |
|--|--------------------------------|
| Informační systém o veřejných zakázkách | Ministerstvo pro místní rozvoj |
| Výsledky voleb | Český statistický úřad |
| Registr aktivních legislativních prací – RALP (resortní nástroje podporující transparentnost výkonu státní správy a zapojení veřejnosti) | Ministerstvo dopravy |
| Finanční statistika - státní dluh | Ministerstvo financí |
| Finanční statistika - vládní finanční statistika | Ministerstvo financí |
| ÚFIS - účetní záznamy a finanční údaje z CSÚIS | Ministerstvo financí |
| Online přístup k údajům o financování politických stran | Ministerstvo vnitra |
| Centrální registr dotací | Ministerstvo financí |

Aplikace využívající principy otevřených dat

| Název a odkaz | Popis |
|--|--|
| Recovery track the money http://www.recovery.gov | Online portál americké vlády sloužící k poskytování jednoduchého přístupu k datům o poskytování vládních dotací s podporou nahlášení jejich možného zneužití |
| WeThePeople Your Voice In Our Government https://www.whitehouse.gov/petitions | Aplikace americké vlády sloužící k poskytnutí možnosti jednoduchého zakládání peticí a oznamování problémů. V případě dosažení dostatečné podpory pro jednotlivé podmínky jsou následně projednány v expertních skupinách. |
| USASpending http://www.usaspending.gov | Portál sloužící k přehlednému zveřejňování federálního rozpočtu USA. |
| ClearSpending http://sunlightfoundation.com/clearspending/summary/ | Stránka monitoruje a kontroluje využívání amerických veřejných financí. |
| OpenSpending http://OpenSpending.org/ | Portál pracující s různými volně přístupnými dataseťmi monitorující veřejné finance po celém světě |
| PopVox https://www.popvox.com | Americký portál umožňující občanům komentovat a připomínkovat návrhy zákonů |
| UKCrimeStats | Statistika kriminality ve Velké Británii až na |

| | |
|---|--|
| http://www.ukcrimestats.com | úroveň ulic. |
| uk.roadworks http://www.livemap.co.uk/uk/roadworks.html | Statistika opravy silnic ve Velké Británii. |
| CareHomeMap http://www.carehome.co.uk | průvodce po domácí zdravotní péči ve Velké Británii. |

ČSU - Veřejná databáze - tematické okruhy

| Tématiky okruh | Skupiny |
|-----------------------------------|--|
| Souborné informace | Ročenky, Analýzy, Konjunkturální průzkumy, Informace o regionech, městech a obcích, Souborné informace, Případové („ad hoc“) studie, Buletin |
| Životní prostředí, zemědělství | Životní prostředí, Zemědělství, Lesnictví |
| Práce, sociální statistiky | Životní úroveň, Práce a mzdy, Sociální zabezpečení, Školství, kultura, zdravotnictví, Soudnictví, kriminalita, nehody |
| Obyvatelstvo, volby | Demografie, sčítání lidu, domů a bytů, volební statistika, cizinci |
| Makroekonomika | Národní účty, Finanční hospodaření, Investice |
| Zahraniční obchod | Zahraniční obchod se zbožním, Zahraniční obchod se službami |
| Ceny | Ceny výrobců, Ceny spotřebitelského zboží, ceny zahraničního obchodu |
| Průmysl, stavebnictví, energetika | Průmysl, Energetika, Stavebnictví |
| Služby | Tržní služby, Obchod, pohostinství a ubytování, Cestovní ruch, Doprava a spoje, Peněžnictví a pojišťovnictví, Netržní služby, Věda a výzkum, Informační společnost |

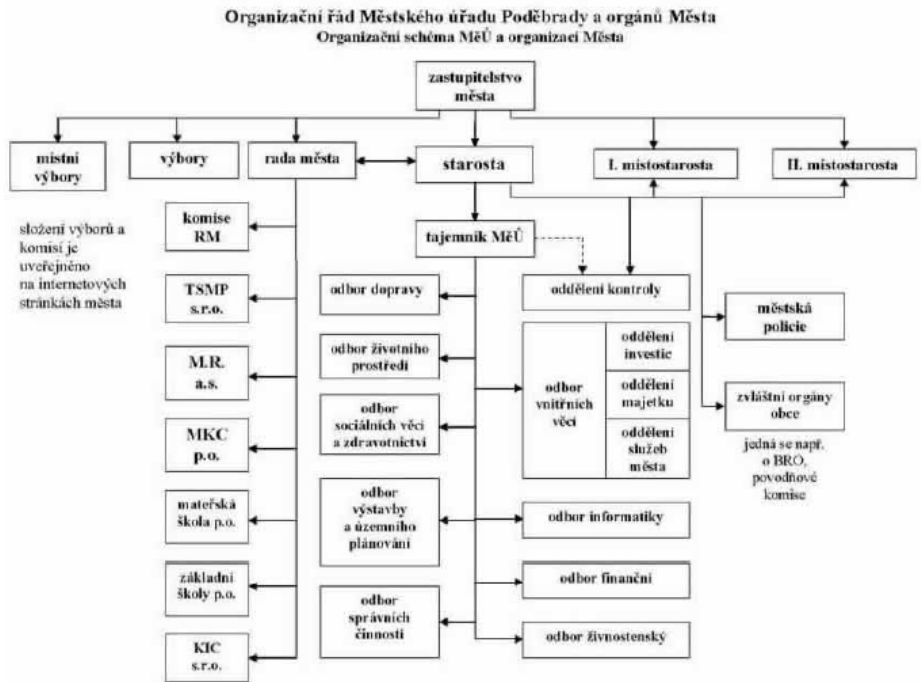
Portály otevřených dat ve světě

| Stát | Katalog | Nástroj |
|-------------------------|---|--------------|
| Belgie | http://data.gov.be | Proprietární |
| Dánsko | http://data.digitaliser.dk | Proprietární |
| Estonsko ²⁰⁹ | http://opendata.riik.ee | Proprietární |
| Finsko | http://data.suomi.fi | Proprietární |
| Francie | http://www.data.gouv.fr | Proprietární |
| Nizozemsko | http://data.overheid.nl | CKAN |
| Portugalsko | http://www.dados.gov.pt | Proprietární |
| Slovensko | http://data.gov.sk | CKAN |
| Španělsko | http://datos.gob.es | Proprietární |
| USA | http://www.data.gov | Socrata |
| Velká Británie | http://catalogue.data.gov.uk | CKAN |

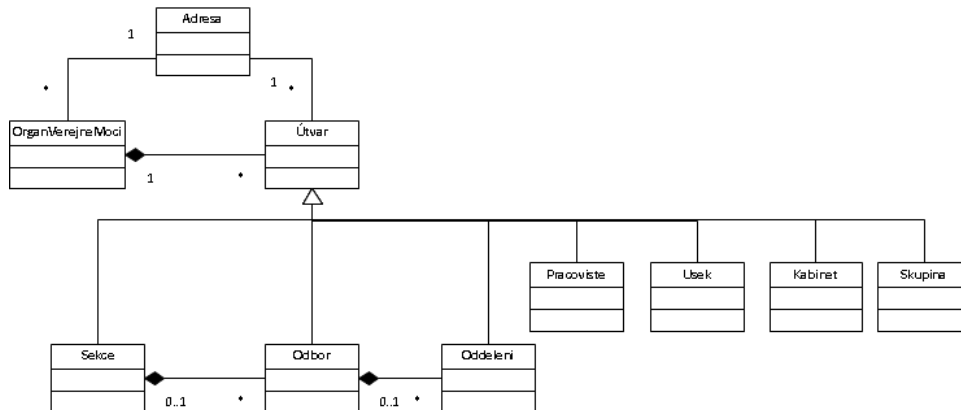
| Stát | Region | Katalog | Nástroj |
|----------------|-------------|---|--------------|
| Finsko | Helsinki | http://www.hri.fi/en/ | CKAN |
| Francie | Paříž | http://opendata.paris.fr/ | Proprietární |
| Německo | Berlín | http://daten.berlin.de/ | CKAN |
| Španělsko | Zaragoza | http://datos.zaragoza.es | Proprietární |
| USA | Stát Oregon | http://data.oregon.gov | Socrata |
| USA | Chicago | http://data.cityofchicago.org | Socrata |
| Velká Británie | Kent | http://picandmix.org.uk/categories/ | Proprietární |

²⁰⁹ Datový katalog ve výstavbě

Organizační struktura města a útvaru ministerstva



Obrázek 57. Organizační struktura města



Obrázek 58. Organizační struktura ministerstva