

Seminář „Zpřístupnění dat a informací“
22.11.2012

Principy a praktické využití otevřených dat

¹Dušan Chlapek, ¹Jan Kučera,
²Martin Nečaský

¹Fakulta informatiky a statistiky,
Vysoká škola ekonomická v Praze

²Matematicko-fyzikální fakulta,
Univerzita Karlova v Praze

Agenda

- Principy a praktické využití otevřených dat
 - Co jsou otevřená data
 - Přínosy otevřených dat
 - Využití otevřených dat
 - Jak publikovat otevřená data
 - (Od otevřených k otevřeným propojitelným datům)
- Metodika publikace otevřených dat bude také k dispozici na [internetových stránkách Úřadu vlády ČR](#)

Otevřená data

- **Otevřená data**
 - Jsou zveřejněna na internetu
 - Způsob jejich uveřejnění a podmínky užití neomezují uživatele ve způsobu použití
 - Podmínky užití opravňují uživatele k jejich dalšímu šíření
 - Je uveden autor dat (i při dalším šíření)
 - Při dalším šíření musí i ostatní uživatelé mít stejná oprávnění s daty nakládat
 - Během šíření dat nesmí dojít např. k omezení jejich využití pouze pro nekomerční účely

Vlastnosti otevřených dat veřejné správy (1)

- **Otevřená data veřejné správy musí být:**
 - **úplná** - data jsou zveřejněna v maximálním možném rozsahu,
 - **snadno dostupná** - data jsou dostupná a dohledatelná běžnými ICT nástroji a prostředky,
 - **strojově čitelná** - data jsou ve formátu, který je strukturovaný takovým způsobem, že pomocí programové aplikace lze z dat získat žádané (vybrané) údaje,
 - **používající standardy s volně dostupnou specifikací** (otevřené standardy),
 - **zpřístupněna za jasně definovaných podmínek užití dat** (licence) s minimem omezení,
 - **dostupná uživatelům při vynaložení minima možných nákladů** na jejich získání.

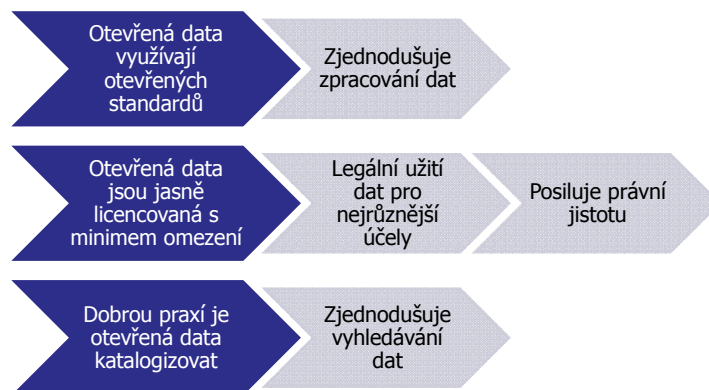
Vlastnosti otevřených dat veřejné správy (2)

- Otevřená data veřejné správy **by dále měla být:**
 - **primární (původní)** - data, která jsou zveřejněna původcem dat v podobě, v jaké byla původcem jako primární (původní) vytvořena,
 - zveřejněná **bez zbytečného odkladu**,
 - **neomezující přístup** - data dostupná způsobem, který nediskriminuje jednotlivce nebo skupinu osob,
 - **stále dostupná** - data jsou dostupná on-line po dobu uvedenou jejich poskytovatelem.

Současné problémy publikování dat VS ČR

- VS ČR publikuje velké množství dat
- Využití těchto dat je náročné, protože
 - je obtížné zjistit, jaká data VS ČR publikuje a kde
 - data jsou zveřejňována v řadě různých formátů a to často takových, které neumožňují opakované strojové zpracování (např. PDF, DOC, HTML, atd.)
 - data nejsou publikována pod vhodnou licenci
- Důsledkem je
 - Náročné vyhledávání dat
 - Pracné, časově a technicky náročné zpracování dat
 - Právní nejistota při využívání dat VS ČR

Proč otevřená data VS ČR?



Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

7

Některá data VS ČR jsou otevřená již dnes

Podmínka otevřených dat	Data z RÚJAN	Data o volbách
1. Úplnost	Splněno	Splněno
2. Primárnost	Splněno - data ze základních registrů jsou považována za primární	Splněno
3. Včasnost zveřejnění	Splněno	Splněno
4. Snadná dostupnost	Splněno	Splněno
5. Strojová čitelnost	Splněno - formát VFR	Splněno - formáty DBF, XML, MS Excel
6. Přístup bez omezení	Splněno	Splněno (v průběhu volebních dnů jsou některá data zpřístupněna po registraci)
7. Otevřené standardy	Splněno - specifikace používaného formátu je bezplatně dostupná	Splněno - formáty DBF a MS Excel jsou rozšířené a zpracovatelné v řadě i bezplatných nástrojů
8. Jasně podmínky užití	Splněno - podmínky využití dat jsou definovány v zák. č. 111/2009 Sb. a také na webových stránkách ČÚZK	Splněno
9. Stálá dostupnost	Splněno	Splněno
10. Dostupnost s minimem nákladů	Splněno - data lze využívat bezplatně	Splněno - data lze využívat bezplatně

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

8

Přínosy otevřených dat (1)

- **Usnadnění opětovného použití dat veřejné správy**
 - Je odhadováno, že na území EU27 může usnadnění využití těchto dat zvýšit přínosy z jejich využití o 40 miliard EUR ročně (Vickery, 2011)
- **Potenciál pro vznik nových aplikací a služeb využívajících data veřejné správy**
 - Odborná veřejnost může tvořit aplikace pro laiky a zpřístupňovat data na platformy jako iOS, Android, atd.
 - Veřejná správa se může soustředit na plnění zákonných povinností
 - Soukromý sektor může obohacovat svoje komerční služby o otevřená data

Vickery, G.: Review of recent studies on PSI re-use and related market developments (2011)

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

9

Přínosy otevřených dat (2)

- **Posilování transparentnosti veřejné správy**
 - Zpřístupňování dat o fungování veřejné správy
 - Kontrola fungování VS ze strany veřejnosti
- **Usnadnění agregací a analýz dat veřejné správy**
 - Lépe zpracovatelná data pro rozhodování podniků
 - Podpora vědecké a výzkumné činnosti

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

10

Příklady aplikací ze světa

- [CareHomeMap](#) – sociální služby (GB)
- [FixMyTransport](#) – hlášení problémů v dopravě (GB)
- [Ozon Sonar](#) – stav ozónu v Berlíně (DE)
- [Evolution of European Union legislation](#) – vývoj evropské legislativy
- [ZNasichDani.sk](#) – veřejné výdaje (SK)

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

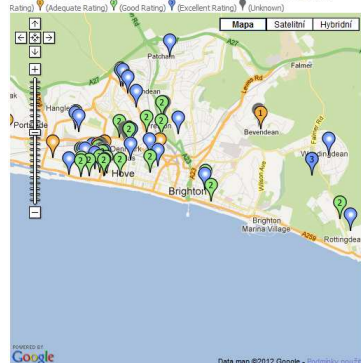
11

Aplikace CareHomeMap

Map Shows lowest **Overall outcome** and **Overall Quality** average rating by Care Home filtered by **Older People**

73 Care Homes found within 5 miles. Click on a Care Home on the map to show details

Key: (Major Concerns) (Moderate Concerns) (Minor Concerns) (No Concerns) (Poor Rating) (Adequate Rating) (Good Rating) (Excellent Rating) (Unknown)



Show Care Home Display Options...

Hide Care Home Details

Latest Outcomes Latest Ratings Details Licences Ratings Feedback: Feedback Graph Glossary

Latest Care Quality Commission (CQC) Outcomes

Outcomes replaced Ratings from August 2010 onwards

Key: (Major Concerns), (Moderate Concerns), (Minor Concerns), (No Concerns)

Service Name	All Older People Care Homes within 5 miles	Click on a Care Home on the map to show details here
Inspection Date		
Overall	Ave: ★★ (2.68)	
Involvement	Ave: ★★★ (2.90)	
Personalised Care	Ave: ★★★ (2.87)	
Safeguarding	Ave: ★★★ (2.68)	
Staffing Suitability	Ave: ★★★ (2.84)	
Management Quality	Ave: ★★★ (2.86)	

Rating, Licensing and Outcome information on this page is compiled from reports which are Copyright of the Care Quality Commission. No warranty is given that information on the site is correct and we encourage users of CareHomeMap to read the full reports and visit chosen care homes to make further assessment.

Copyright© 2009-2012 CareHomeMap

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

12

Příklady tuzemských aplikací

- [Mapa veřejných zakázek](#)
- [Hospodaření obcí](#)
- [Rejstříky škol](#)
- [Vášmajetek.cz](#) – státní majetek na jedno místě
- [Rozklikávací rozpočet obce](#)
- [Rozklikávací rozpočet státu](#)

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

13

Aplikace Hospodaření obcí

HOSPODAŘENÍ OBCÍ.CZ

Zadejte název obce

Hledat

Náklady a výnosy našich obcí pod drobnohledem

DOMŮ

VÝKAZY

STATISTIKY

DUELY

JAK ČÍSLA CHÁPAT

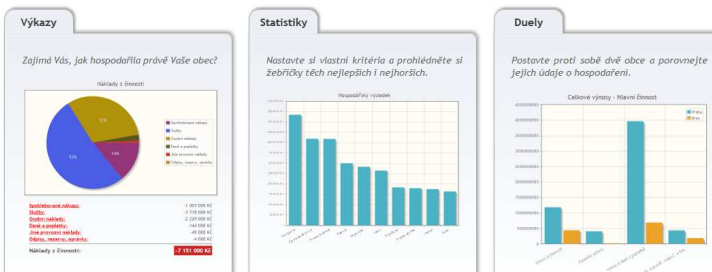
O PROJEKTU

Hospodaření obcí

- Zajímá Vás, jak hospodařily naše obce v roce 2010?
- Chcete se dozvědět, které obce dosáhly nejvyššího/nejnižšího hospodářského výsledku?
- Rádi byste viděli porovnání hospodaření obcí na úrovni celé republiky, krajů i okresů?
- Chtěli byste si prohlédnout strukturu výnosů a nákladů našich obcí v přehledné formě?



Pak jste na správném místě! Na těchto stránkách máte možnost nejen se proklikat výkazem zisku a ztráty každé naší obce, ale také porovnávat na jednom místě jejich údaje podle nejrůznějších kritérií. Podívejte se tedy, jak jsi vedly naše obce v hospodaření.



Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

14

JAK PUBLIKOVAT OTEVŘENÁ DATA?

Seminář „Zpřístupnění
dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

15

Jak publikovat otevřená propojitelná data

1. Analýza dat a možností jejich uveřejnění
2. Výběr vhodného formátu dat
3. Návrh způsobu přístupu k datům
4. Export dat do navrženého formátu
5. Publikace dat
6. Katalogizace dat

Seminář „Zpřístupnění
dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

16

Analýza dat a možností jejich uveřejnění

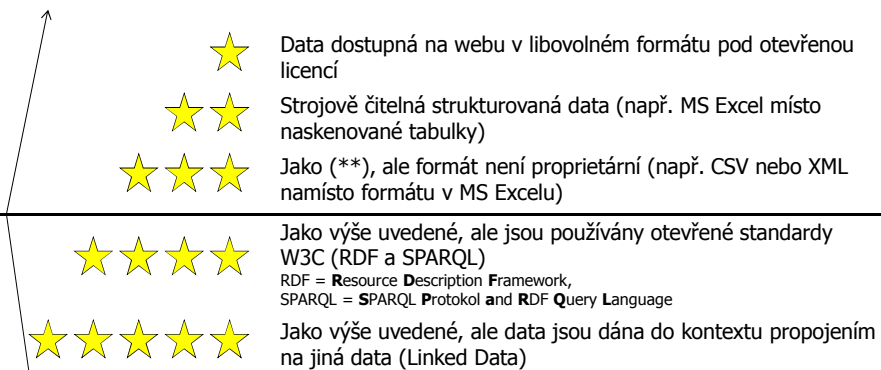
- Zmapování existujících dat
- Analýza struktury a sémantiky dat
 - Návrh konceptuálního modelu
- Popis dat a jejich dokumentace
 - Lze využít navrženého konceptuálního modelu
- Analýza možností uveřejnění
 - Ne všechna data je možné zveřejnit
 - Např. z důvodu ochrany osobních údajů
 - Zvážit uveřejnění upravených dat
 - Např. anonymizovaná dat

Výběr vhodného formátu dat

1. Využití existujících standardních formátů
 - Vhodné je využívat formáty založené na XML nebo RDF
 - Vyhlášené datové prvky v ISDP
2. Kombinace či rozšíření existujících formátů
 - Alespoň část problematiky pokrytá standardním formátem
 - Využití více standardních formátů společně
 - Doplnění standardu o specifické potřeby
3. Vytvoření vlastního formátu
 - Vlastní formát vytvořený na základě konceptuálního modelu
 - Je vhodné využít XML nebo RDF

Úrovně zveřejnění umožněné různými formáty

Otevřená data



Zdroj: Tim Berners-Lee, <http://www.w3.org/DesignIssues/LinkedData.html>

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy

19

Porovnání vlastností formátů dat

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

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

20

Návrh způsobu přístupu k datům

- **Možné způsoby zpřístupnění**
 - Data zpřístupněna jako jeden celek v jednom či skupině souborů a v definovaných časových intervalech přepisovány
 - Tj. vždy bude možné získat všechna data platná k poslednímu datu exportu
 - Vhodné, pokud se mohou měnit data již zveřejněná v minulosti
 - Zpřístupněny přírůstky v definovaných časových intervalech
 - Např. každý měsíc budou zveřejněna nová data a změny
 - Vhodné, pokud se jednou zveřejněná data nemění a pouze přibývají přírůstky
 - Online přístup do databáze pomocí vhodného dotazovacího rozhraní
 - Tj. dotazem je specifikována požadovaná podmnožina dat, která je vyexportována do navrženého formátu a zaslána uživateli
 - Vhodné v případě, kdy se data velmi dynamicky mění a je nutné k nim zajistit aktuální online přístup

Export dat do navrženého formátu

- Tvorba skriptů pro převod dat do zvoleného formátu

Publikace dat

- Volba URL pro publikování dat
- URL by mělo být v rámci domény orgánu VS, který data publikuje
- Z webových stránek orgánu VS musí vést na zveřejněná data dobře viditelný odkaz (nejlépe z hlavní nabídky)
- Nejlépe vytvořit samostatnou stránku na webové prezentaci/portálu určenou pro zveřejňování otevřených dat. Na tuto stránku umístěte odkazy na všechna zveřejňovaná otevřená data.

Katalogizace dat

- Aby byla data co nejlépe dohledatelná, je vhodné je katalogizovat
- Podrobnosti v další části přednášky
- O datech můžete také dát vědět pomocí:
 - kanálu pro odběr novinek (RSS, Atom)
 - sociálních sítích
 - vydání tiskové zprávy
 - mailing listu okfn-cz@lists.okfn.org
 - sdružuje komunitu se zájmem o otevřená data

PUBLINK

- Konzultace a pomoc při publikování otevřených
- Konzultace poskytovány členy LOD2 konsorcia zdarma vybraným zájemcům
- Zaměřeno na pomoc orgánům veřejné správy
- Zájemci se musí přihlásit na základě výzvy
- Podrobnosti o PUBLINK na webu lod2.eu



Poděkování a kontakty

Děkujeme za pozornost

Dušan Chlapek (chlapek@vse.cz)
Jan Kučera (jan.kucera@vse.cz)

Fakulta informatiky a statistiky,
Vysoká škola ekonomická v Praze

Martin Nečaský (necasky@xrg.cz)

Matematicko-fyzikální fakulta,
Univerzita Karlova v Praze

BACKUP

Seminář „Zpřístupnění
dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

27

OD OTEVŘENÝCH DAT K OTEVŘENÝM PROPOJITELNÝM DATŮM

Seminář „Zpřístupnění
dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

28

Otevřená propojitelná data

- Otevřená data, která **navíc** splňující sadu zásad pro publikování a propojování souvisejících dat na webu otevřeným a na standardech založeným způsobem
 - Pojmenování objektů na webu pomocí URI
 - Použití HTTP URI, které umožňují je vyhledat v prostředí dnešního webu
 - Při vyhledání URI jsou uživatelům poskytnuta data o objektu reprezentovaná v modelu RDF
 - Objekty jsou provázány pomocí odkazů mezi HTTP URI, takže je možné objevovat související objekty

Výhody propojitelných dat

- Umožňují propojování dat z různých datových zdrojů
 - Propojení umožňuje dávat data do souvislostí
 - Roste informační hodnota dat
- Standardizovaný přístup k datům
 - Odpadá nutnost budovat individualizované rozhraní ke každému datovému zdroji

Publikace informací o veřejné instituci

○ základní údaje
□ zaměstnanci
◎ oddělení
⬡ veřejné zakázky
△ rozpočet
◇ výdaje

WWW stránky instituce







Obchodní rejstřík

ÚFIS

Profil zadavatele



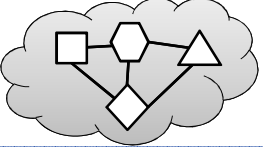
ISVZUS

gov.cz

Uživatel: Dodavatelé do veřejných zakázek MF z Libereckého kraje na Google mapách v iPhone aplikaci. Pro každou zakázku agregace nebo výpis plateb, vazbu na rozpočet a zodpovědnou osobu.

- **Kde** získám data o zakázkách, odpovědných osobách, výdajích a rozpočtu MF?
- Jak mám data **sloučit a provázat**?
- Jak **zobrazit** data v iPhone na mapě?

Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

31

Současný stav

○ základní údaje
□ zaměstnanci
◎ oddělení
⬡ veřejné zakázky
△ rozpočet
◇ výdaje

WWW stránky instituce







Obchodní rejstřík

ÚFIS

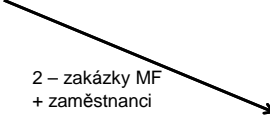
Profil zadavatele

ISVZUS

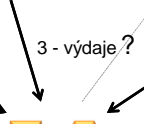
gov.cz


2 – zakázky MF + zaměstnanci

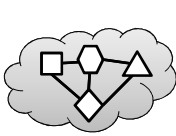



3 - výdaje ?




1 – zakázky MF







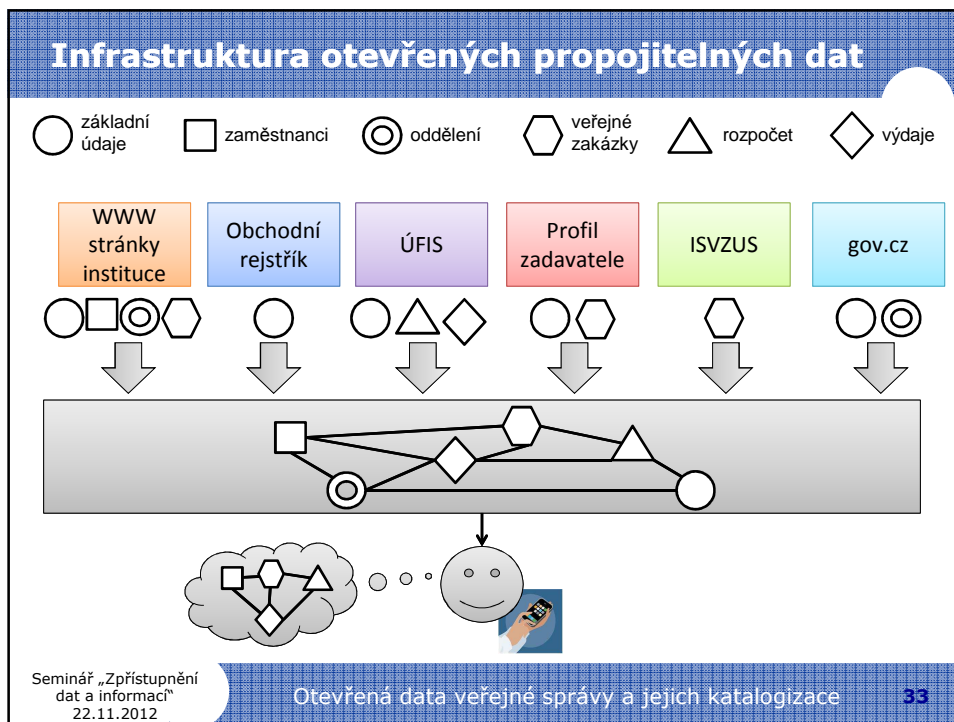
uživatel neobjevil ?



Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

32



Formát RDF

- RDF – Resource Description Framework
- Pracuje s HTTP URI zveřejňovaných objektů a sjednocuje způsob zveřejňování údajů o objektech a souvislostí (tj. propojení) mezi nimi
- Datovým modelem RDF je graf
- Uzly grafu = objekty (identifikované svými HTTP URI) a jejich údaje (texty, čísla, datумы, apod)
- Hrany přiřazují údaje k objektům a také reprezentují propojení mezi nimi.
- Graf je potom popsán v podobě trojic
 - subjekt predikát objekt
- Subjekt je zdrojový uzel hrany, objekt je cílový uzel hrany a predikát specifikuje typ hrany

Seminář „Zpřístupnění dat a informací“ 22.11.2012 Otevřená data veřejné správy a jejich katalogizace 34

Formát RDF – příklad

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

- První trojice spojuje osobu s jejím jménem
 - Osoba identifikovaná URI <http://www.opendata.cz/necasky#me> je spojena s datovou hodnotou „Martin Nečaský“
 - Typ hrany *foaf:name* říká, že subjekt a objekt jsou ve vztahu „subjekt má jméno“
- Druhá trojice spojuje dvě osoby
 - Martin Nečaský zná Dušana Chlapka
 - Osoby jsou identifikovány svými URI
 - Typ hrany *foaf:knows* říká, že subjekt a objekt jsou ve vztahu „subjekt zná objekt“

Ontologie

- Typ hrany RDF grafu = predikát
- Typem hrany může být cokoliv
- Možné typy hran (a další koncepty jako např. třídy objektů) jsou definovány v tzv. slovníku nebo také ontologii
- Ontologie popisuje nejenom strukturu dat (tj. predikáty a jejich zapouzdření do tříd), ale také sémantiku definovaných predikátů a tříd

Projekt LOD2

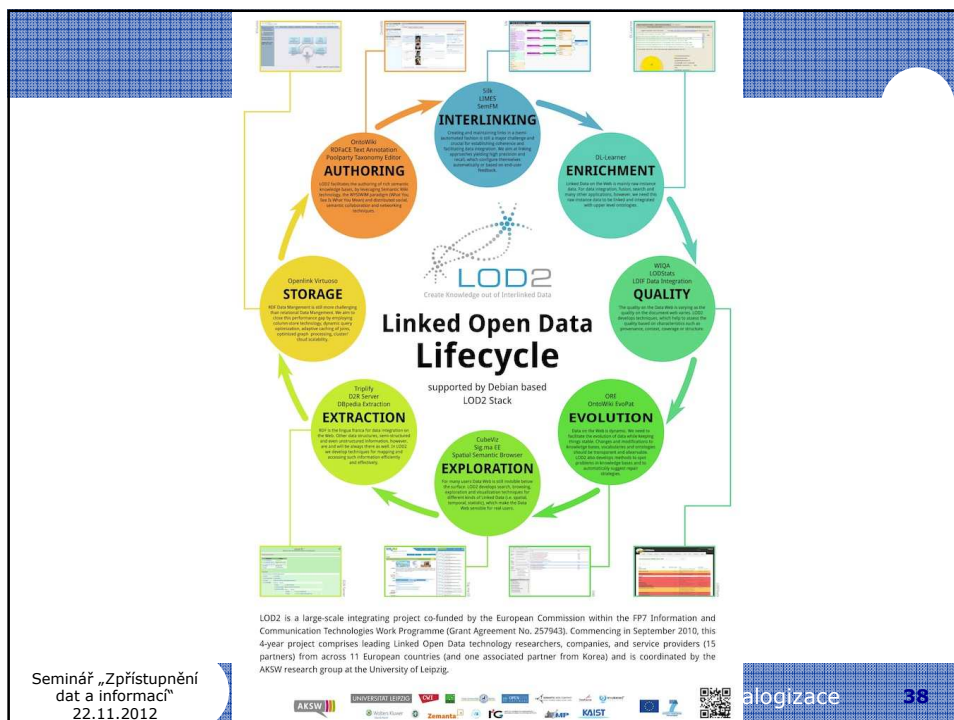
- Projekt 7. rámcového programu EU
- Zaměření
 - Tvorba nástrojů a metodik pro zpracování a řízení velkých objemů strukturovaných dat na webu
 - Tvorba multi-oborových, vícejazyčných ontologií
 - Tvorba nástrojů pro vytváření, vyhledávání a procházení propojitelných dat
 - Vývoj algoritmů pro automatické propojování dat na webu
 - Tvorba standardů a metod pro zajištění kvality, důvěryhodnosti a informací o původu dat
- VŠE Praha je členem LOD2 konsorcia
- <http://lod2.eu>



Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

37



Seminář „Zpřístupnění dat a informací“
22.11.2012

Otevřená data veřejné správy a jejich katalogizace

38

PUBLINK

- Konzultace a pomoc při publikování otevřených propojitelných dat
- Konzultace poskytovány členy LOD2 konsorcia zdarma vybraným zájemcům
- Zaměřeno na pomoc orgánům veřejné správy
- Zájemci se musí přihlásit na základě výzvy
- Podrobnosti o PUBLINK na webu lod2.eu



PUBLINK 2013

- Harmonogram
 - Uzávěrka přihlášek: 31.12.2012
 - Oznámení výsledků uchazečům: 30.1.2013
 - PUBLINK pathfinder: únor až březen 2013
 - PUBLINK kickstart: duben až červen 2013
- PUBLINK pathfinder
 - Jednodenní workshop
 - Tvorba hrubé specifikace a tech. architektury pro publikaci dat
- PUBLINK kickstart
 - (Mini-)Projekty v rozsahu 5 až 10 dní
 - Publikování dat na internetu
 - Integrace nástrojů s nástroji v LOD2 stack

Jak se zapojit do PUBLINK

- Zašlete email na publink@lod2.eu (max. 2 strany), kde uvedete:
 - Stručný profil Vaší organizace
 - Popis dat (kvalita, objem, používané formáty) nebo nástroje (účel, technologie, rozhraní), které chcete zveřejnit
 - Potenciální uživatele Vašich dat nebo nástroje
 - Vaše předpokládané zapojení – např. uveďte, zda poskytnete pracovníky pro podporu PUBLINK projektu a pro údržbu dat po jeho skončení
 - Předpokládaný model licencování (PUBLINK se obecně zaměřuje na publikování pod otevřenými licencemi, viz <http://opendefinition.org>)
- Mail zašlete v angličtině
- Bližší informace na <http://lod2.eu/Article/Publink.html>