

**FAKULTA INFORMATIKY**  
**MASARYKOVA UNIVERZITA**



Library Linked Data

(<http://www.w3.org/2005/Incubator/lld/>)

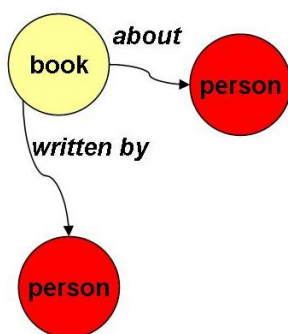
Project of W3C

*David Vávra*

Brno, 4.12.2012

## Úvod:

Library Linked Data, celým názvem Library Linked Data Incubator Group (dále jen LLD) je projekt *webového konsorcia W3C*. Projekt se pohybuje v oblasti digitálních knihoven, převážně v odvětví interoperability. Samotná Linked Data jsou jakousi metodikou či souhrnem pravidel pro to, jak v internetu vytvářet sémantické weby (tj. Weby čitelné lidmi i strojově – roboty či dalšími automatickými nástroji) a propojovat jednotlivé objekty/data určitými predikáty vztahů mezi danými objekty. Pro digitální knihovny jsou Linked Data obzvláště užitečná. Lze to vidět na příkladu knihy a dvou osob, viz následující obrázek.



Pointa tohoto systému je v tom, že jednotlivá data, v předchozím obrázku 2 osoby a kniha, mohou být uložena v různých knihovnách či portálech. Údaje o lokacích mohou být použity z <http://www.geonames.org/>, údaje o autorovi mohou být použity z <http://viaf.org>. Výhodou tohoto systému je to, že nevznikají duplicitní záznamy → není potřeba uchovávat zbytečně velké množství dat a uchovávání dat je finančně velice náročné. Také zde dochází k ucelenému propojení informací v informačním světě.

## Aktuální stav:

Projekt byl ukončen 31. srpna 2011 a byla vydána závěrečná zpráva. Závěrečná zpráva obsahuje doporučení W3C konsorcia, jak tvořit sémantické weby a interoperabilní digitální knihovny.

## Cíle:

Cílem LLD je zlepšit interoperabilitu a vzájemné propojení knihoven a informací. Projekt byl iniciován zaměstnanci knihoven, výzkumných ústavů, univerzit a dalšími lidmi, zájímavými se o tuto problematiku. S tím také souvisí návrh standardů pro tvorbu sémantických webů.

## **Projekt:**

Projekt byl zahájen v květnu roku 2010. Hlavními zakladateli projektu byli Tom Baker, Emmanuelle Bermes a Antoine Isaac. Původní ukončení projektu bylo naplánováno na konec května 2011. Nakonec bylo ukončení projektu odloženo na 31. 8. 2011. Tento termín byl dodržen a tak již nebylo potřeba jej dále posouvat.

Mezi první členy patřily organizace jako Library of Congress, Helsinky University of Technology a Los Alamos National Laboratory. Komunikace jednotlivých členů byla založena na veřejném emailu (public-xg-llid@w3.org a public-llid@w3.org), přes který probíhala emailová komunikace. Dále byl ustanoven týdenní interval telefonických konferencí a roční interval pro fyzickou schůzi všech členů. Další komunikace probíhala přes komunikační protokol IRC.

Důvodem projektu byla potřeba přeorganizovat systém digitálních knihoven kvůli čím dál tím většímu počtu různých zdrojů a množství nových standardů a způsobů tvorby digitálních knihoven. Také problém interoperability začal být aktuální pro mnohem více, převážně kulturních, zařízení a institucí. Například pro různá muzea, archivy, vydavatele, atd...

Základní body postupu projektu:

- Zajistit spolupráci mezi subjekty (knihovny, muzea, vydavatelé, archivy) zainteresovanými do používání Linked Data
- Shromáždit projekty, ve kterých byly úspěšně použity prvky sémantického webu
- Zjistit názvosloví a způsoby budování interoperability pro relevantní datové modely a zjistit, jak tyto modely vylepšit po stránce interoperability
- Vytvořit a vypracovat potřebné standardy a normy
- Ze shromážděných projektů zjistit a určit užitečné technologie sémantického webu
- Navrhnout organizaci která bude v určeném rozsahu pokračovat v udržování a rozvíjení nově vybudovaných standardů a norem

**Projekt ovšem neměl za úkol vytvořit nové názvosloví a nový systém, pouze sjednotit standardy pomocí již existujících a zjistit, jaké jsou optimální cesty k tvorbě sémantických webů a Linked Data systémů v rámci digitálních knihoven.**

Podíl na projektu měly svou spoluprací i další skupiny z webového konsorcia, ale i externí skupiny. Z W3C skupin to byly například tyto: *Semantic Web Activity* (skupina pro tvorbu standardů a názvosloví sémantického webu), *eGovernment Interest Group* a další. Mezi externí skupiny a spolupracující projekty patří například *IFLA* nebo *Dublin Core Metadata Initiative*.

**Jednotlivé výhody Linked Data pro jednotlivé organizace a skupiny lidí:**

- Pro vědce a studenty: Weby, používající Linked Data pomohou propojit celý informační

internetový svět, jak digitální knihovny, tak i jiné informační portály (třeba i zpravodajství), přitom Linked Data netvoří jiný web, pouze vhodným způsobem rozšiřují ten stávající a zvyšují jeho sémantičnost

- **Pro organizace:** Linked Data mohou pomoci organizacím zlepšit vzájemnou propojenost dat, u kterých to je žádáno. Pro knihovny to pak má ještě větší výhody a to například snížení finanční náročnosti jejich informačního systému, nemusí každá knihovna tvořit vše, ale mohou využívat to, co vytvoří jiné knihovny
- **Pro knihovníky a archiváře:** Zde jsou téměř stejné výhody jako ty popsané výše, mohou se využívat záznamy již vytvořené někým jiným
- **Pro vývojáře a distributory:** Nejsou vázáni na specifický datový formát, mohou používat prosté HTTP, které Linked Data podporují. Také nebudou muset používat metadatové formáty, jakým je např. MARC, metadata budou srozumitelná přímo ve webové formě a to pomocí RDF

### Úskalí tradičních knihovních dat:

Data jsou v současných digitálních knihovnách ukládána tak, že je lze procházet webovým rozhraním, ovšem nejsou na sebe vzájemně propojena. V prostředí, které využívá Linked Data se jednotlivé datové informace neopakují, pokud jsou použity vícekrát, například autoři. Současné knihovní standardy jsou určeny především pro samotné knihovníky, například MARC, lze je použít pouze v této oblasti, Linked Data jdou použít všude a mohou vše propojit.

Knihovny své metadatové a datové záznamy spravují lokálně, nepoužívají URI, které by mohlo pomoci ostatním, bylo by pak snadné se odkazovat na již existující záznamy.

\*Dalším problémem je to, že komunita sémantického webu a lidé spravující digitální knihovny používají jinou terminologii, proto je nutné tuto terminologii nejprve sjednotit. Také to závisí na dodavatelích knihovních systémů, kteří musí Linked Data zahrnout do jejich software, aby tento princip mohl fungovat, knihovny by si to měly pohlídat a informovat o tom své softwarové dodavatele.

Úspěch Linked Data bude záviset na schopnosti lidí vyhledat a použít již existující informace a obsahy, využívající Linked Data. Ovšem bylo těžké zjistit, kde jsou již Linked Data používána. LLD proto vytvořila svoji databázi dostupných zdrojů, ovšem nemůže zaručit to, že zde budou všechny zdroje. Proto byla založena skupina, jejíž úkolem je sbírání nových zdrojů a aktualizace těch, které už zde jsou, po tom, co skončil LLD projekt. Databáze je umístěna na službě Data Hub.

Byly specifikovány tyto tři základní druhy zdrojů:

- **Datasets**
- **Value vocabularies**
- **Metadata element sets**

*Datasets* - Datasetsy jsou ekvivalentem seznamu knihovních záznamů, záznam z datasetu může obsahovat například jako předmět metadata převzatá z Dublin Core a hodnotu převzatou z LCSH (*Library of congress Subject Headings*). Mezi příklady takových datasetů v LLD databázi patří *BNB (British National Bibliography)*, *Europeana Linked Open Data*, *Medline* a další.

*Value Vocabularies* – Obsahuje zdroje (např. témata, umělecké styly nebo autory), které jsou použity jako hodnoty v metadatových záznamech. LCSH zde definuje např. zaměstnání, *AAT (Art and Architecture Thesaurus)* definuje jména umělců, *GeoNames* definuje lokace (názvy měst, států, hor, atd...) a další. **Tyto služby jsou užitečné i jinde, než jen v Library Linked Data.**

Metadata element sets - Sem patří například Dublin Core, MARC21 nebo RDA (*Resource Description and Acces*), nejsou to databáze popisů jednotlivých knih, autorů, atd..., jsou to popisy metadat, standardů, kterými se teprve popisují jednotlivá díla, záznamy a další informace v digitálních knihovnách.

### **Závěrečné doporučení:**

Obecně by knihovny měly začít transformovat jejich současná data tak, aby byla v souladu s Linked Data, stejně tak tvořit nové a zároveň používat data z jiných zdrojů dostupná pomocí Linked Data.

Vedení knihoven by mělo určit přednostně skupinu dat, která mají vysokou prioritu, a na ta postupně začít aplikovat Linked Data, pokud by všechny knihovny přešly najednou, vznikla by spousta duplicitních dat (spíše by nebyly odstraněny) a nebyl by zde takový přínos. Dále by vedení mělo přezkoumat a prodiskutovat práva k datům, u kterých pomocí Linked Data umožňuje jejich další použití a poskytuje je globální síti knihoven.

Standardizační organizace a další zúčastnění by se měli pokusit zvýšit účast knihoven na standardizaci sémantického webu. Měly by spolu spolupracovat komunity sémantického webu a Linked Data, aby spolu oba modely mohly bezproblémově spolupracovat a minimalizovalo se tak množství možných komplikací. Současné standardy používané v digitálních knihovnách vychází ze starších standardů, tudíž nejsou kompatibilní s Linked Data, jelikož Linked Data fungují trochu jinak. Proto je potřeba knihovní standardy vytvářet tak, aby již byly kompatibilní s technologií Linked Data. Standardy je potřeba vytvářet podle toho, co se v praxi osvědčilo a co nikoliv, aby byly úspěšné a mohly se prosadit.

Systémoví dodavatelé a designéři by se měli snažit vytvářet nové služby uživatelům s využitím Linked Data. Experimentální projekty by měly vznikat kvůli prozkoumání nových možností a měly by informovat o svých výsledcích další projekty. Dále by měli tvůrci systémů vytvářet URI pro záznamy v digitálních knihovnách (bez použití URI by neměla Linked Data žádný význam). Oficiální majitelé zdrojových dat a standardů by měli přiřadit URI co nejdříve. Tvůrci katalogů by měli převzít vedoucí roli ohledně tvorby URI adres pro popisování jejich záznamů a zdrojů. Také by měli vypracovat zásady pro tvorbu již zmíněných URI. Například zásadu pro používání URI s protokolem HTTP, závazky k persistenci URI a další. Pro maximální prolínání s dalšími databázemi musí knihovny používat vlastnosti, třídy a instance, které mají výborně definované vztahy s dalšími záznamy v internetu používajícími Linked Data.

### **Zhodnocení projektu:**

Projekty, které se snaží sjednotit standardy, většinou neskončí velkým úspěchem, vytvoří další standard, který má všechny předchozí standardy sjednotit, nakonec se ale nepoužívá, nebo existuje vedle předchozích standardů, ne místo nich. Proto je potřeba projekty zabývající se standardy pečlivě naplánovat, promyslet a dělat je s rozumem. Myslím si, že toto se projektu LLD povedlo, nevytvořil žádný další standard, který by pouze zapadl mezi ostatní, nebo by postupem času dokonce vymizel. Otázka Linked Dat v digitálních knihovnách je během na dlouhou trať, sjednotit všechny knihovny tak, aby používaly Linked Data a používaly je správně, není to otázka, kterou lze vyřešit za jeden den, proto je potřeba, aby organizace, které byly pověřeny úkolem starat se o výsledky tohoto projektu, dělaly svou práci a dělaly ji pečlivě, jinak nikdy nebude web jedna velká dobře strukturovaná sémantická a přehledná síť plná užitečných informací.

## **Zdroje:**

- <http://www.w3.org/2005/Incubator/ld/>
- <http://commonplace.net/2009/06/linked-data-for-libraries>
- <http://acrl.ala.org/techconnect/?p=1208>
- <http://datahub.io/group/ld>
- <http://www.geonames.org>
- <http://viaf.org>

## **Metadata:**

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Library Linked Data Incubator Group" />
<meta name="DC.Creator" content="David Vávra" />
<meta name="DC.Subject" content="Library Linked Data" />
<meta name="DC.Subject" content="Linked Data" />
<meta name="DC.Subject" content="Semantic Web" />
<meta name="DC.Format" scheme="IMT" content="application/pdf" />
<meta name="DC.Description.tableOfContents" content="Tato esej pojednává o projektu
W3C, jehož úkolem je zlepšit interoperabilitu digitálních knihoven a to pomocí Linked
Data." />
<meta name="DC.Date" content="4.12.2012" />
<meta name="DC.Source" scheme="URL" content="http://www.w3.org/2005/Incubator/ld/"
/>
<meta name="DC.Language" scheme="RFC3066" content="cze" />
```