



Jan Zemánek (jan.zemanek@gmail.com)

LINKED | RDF DATA MASHUP

Cvičný mashup

- Přidejte své **jméno, příjmení** a
- **DBpedia URI** místa, kde jste se narodili*
 - * A které má zároveň přiřazeno **owl:sameAs** a **GeoNames URI** (<http://sws.geonames.org/{id}>)
 - jinak přidejte nejbližší větší město (např. okresní, nebo krajské), které splňuje předchozí podmínku
- do Google Spreadsheetu na adrese:
 - <http://is.gd/fXqkh6>

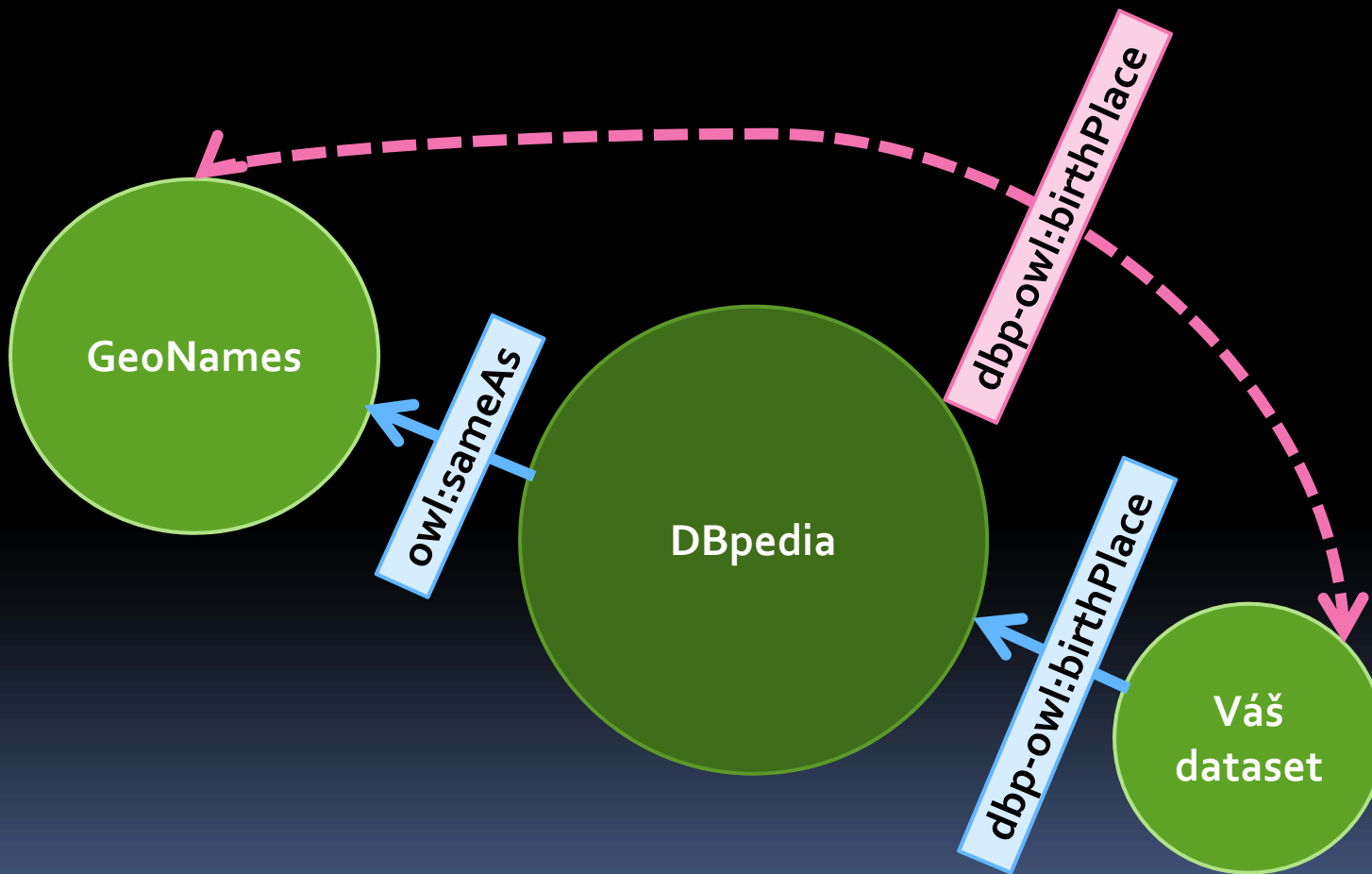
Linked data vs. Linked Data

- Linked **d**ata
 - **Instance, RDF data**
 - Dataset obsahuje odkazy na jiné datasety
 - Datasety sdílí stejné ontologie
- Linked **D**ata
 - Doporučený **způsob vystavování RDF dat** na webu
 - 4 principy
 - **Princip 1+2:** Identifikujte entity pomocí HTTP URI.
 - **Princip 3:** „Dereferencovaná“ HTTP URI vrací RDF|HTML|atd. reprezentaci entity.
 - **Princip 4:** Reprezentace entity obsahuje odkazy na související entity.

Linked Data

- Aplikace LD principů na RDF data na webu umožňuje:
 1. **procházet** RDF data pomocí generických RDF prohlížečů (např. Tabulator, Disco)
 2. **indexovat** RDF data sémantickými vyhledávači (např. Sindice)
 3. **dotazovat** distribuovaná RDF data (např. SQUIN)
- Homogenní globální datové prostředí
 - „Global Data Space“
 - „Giant Global Graph“ (GGG)

Linked data



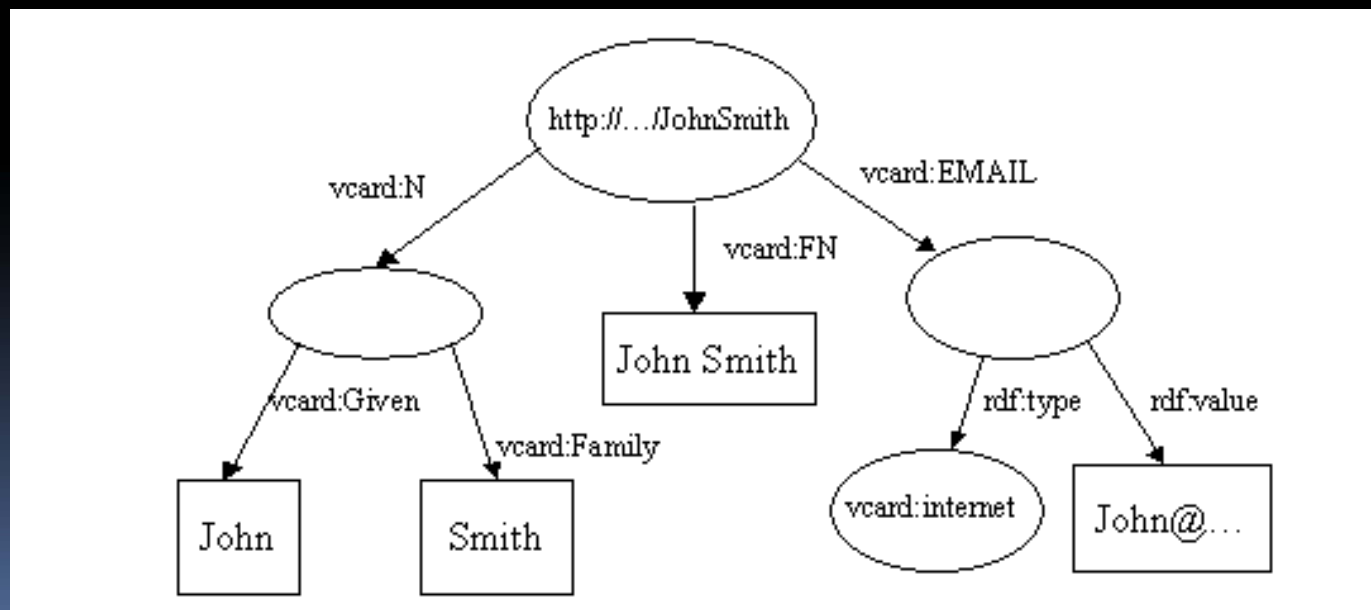
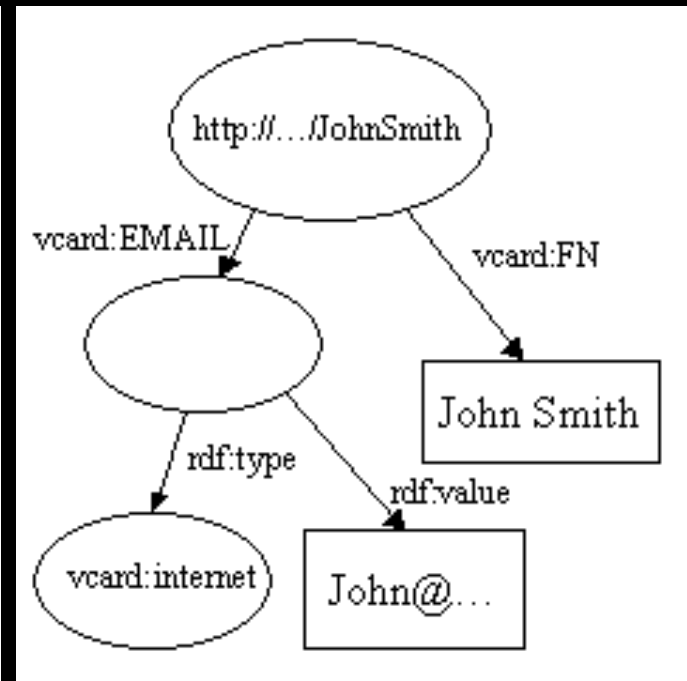
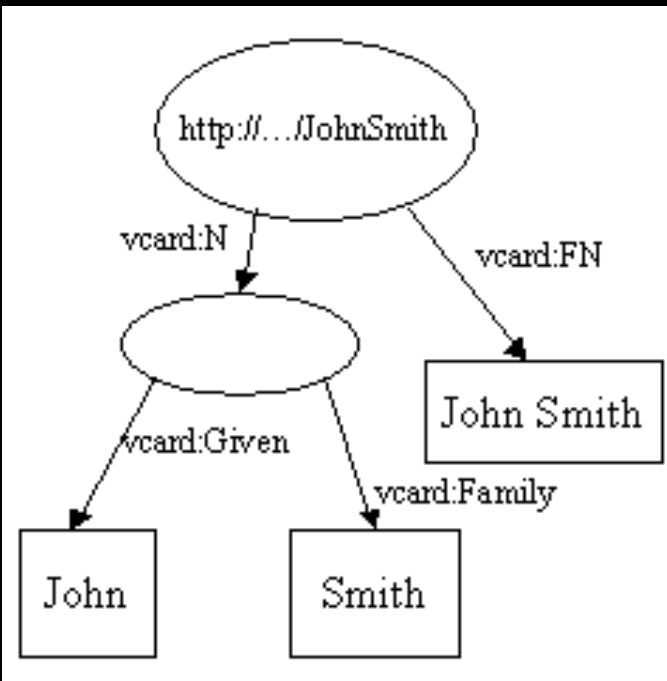
Web API vs. Linked Data

	Webová API	Linked Data
Formát reprezentace	Různé jazyky (XML, JSON, YAML), zachycuje pouze datové struktury	Jednotný jazyk (RDF), tvrzení mají logickou sémantiku
Schémata	Proprietární, specifické pro danou službu	Sdílené slovníky (ontologie), možnost odvození implicitních faktů
Zdroj a jeho reprezentace	Společný identifikátor; content negotiation pro různé reprezentace	Různé identifikátory pro zdroj a jeho reprezentaci, možnost vyjadřovat se zvláště o zdroji a jeho reprezentaci
Odkazy mezi zdroji	Nativně není podporováno	Každou trojici lze chápat jako odkaz mezi zdroji

„Mergeování“ RDF grafů (1)

- Sjednocení („merge“) 2 různých RDF grafů
- Nejdůležitější operace nad RDF daty
- Jena
 - Model model = model₁.union(model₂);
 - (viz. tutoriál)
 - [http://www.openjena.org/tutorial/RDF_API/index.html#ch-Operations on Models](http://www.openjena.org/tutorial/RDF_API/index.html#ch-Operations%20on%20Models)

(2)



Jena – Java SemWeb Framework

- <http://www.openjena.org/>
- Download
 - <http://sourceforge.net/projects/jena/files/Jena/Jena-2.6.4/jena-2.6.4.zip/download>
- Tutorial
 - http://www.openjena.org/tutorial/RDF_API/index.html
- Javadoc
 - <http://www.openjena.org/javadoc/index.html>

DBpedia

- URI zdroje („resource“)
 - <http://dbpedia.org/resource/Přerov>
- URI HTML reprezentace („page“)
 - <http://dbpedia.org/page/Přerov>
- URI RDF reprezentace („data“)
 - <http://dbpedia.org/data/Přerov>
- Wikipedia -> DBpedia
 - Wikipedia: <http://en.wikipedia.org/wiki/{wikiID}>
 - DBpedia: <http://dbpedia.org/resource/{wikiID}>

DBpedia-GeoNames link

- **owl:sameAs**
 - <http://www.w3.org/TR/owl-ref/#sameAs-def>
- Příklad (S-P-O):
 - S: <http://dbpedia.org/resource/Přerov>
 - P: <http://www.w3.org/2002/07/owl#sameAs>
 - O: <http://sws.geonames.org/3067580/>

GeoNames

- URI zdroje
 - <http://sws.geonames.org/3067580/>
- URI **HTML** reprezentace
 - <http://www.geonames.org/3067580/prerov.html>
- URI **RDF** reprezentace
 - <http://sws.geonames.org/3067580/about.rdf>

RDF data

- Vytvořte v Jeně **model**
 - <http://www.openjena.org/javadoc/com/hp/hpl/jena/rdf/model/Model.html>
- (viz. tutoriál)
 - http://www.openjena.org/tutorial/RDF_API/index.html
- a naplňte jej údaji o **vybraných 3 kolezích** z Google Spreadsheetsu (GS)
 - <http://is.gd/fXqkh6>
- (použijte predikáty uvedené v GS)

Naplnění modelu v Jeně

- http://www.openjena.org/tutorial/RDF_API/index.html#ch-Introduction

// some definitions

```
String personURI = "http://somewhere/JohnSmith";
```

```
String fullName = "John Smith";
```

// create an empty Model

```
Model model = ModelFactory.createDefaultModel();
```

// create the resource

```
Resource johnSmith = model.createResource(personURI);
```

// add the property

```
johnSmith.addProperty(VCARD.FN, fullName);
```

„Mergeování“ modelů v Jeně

```
Model model = ModelFactory.createDefaultModel();
```

```
Model m1 = ModelFactory.createDefaultModel();  
m1.read("http://planetrdf.com/index.rdf");
```

```
Model m2 = ModelFactory.createDefaultModel();  
InputStream stream =  
    FileManager.get().open("file://C:/dev/bloggers.rdf");  
m2.read(stream, "");
```

```
model = m1.union(m2);  
model.write(System.out);
```

Vyhodnocování SPARQL dotazů v Jeně

```
String qy = " {SPARQL dotaz}";  
Query query = QueryFactory.create(qy);  
QueryExecution qe =  
    QueryExecutionFactory.create(query,  
    model);  
ResultSet rs = qe.execSelect();  
ResultSetFormatter.out(System.out, rs, query);  
qe.close();
```


Úkol (1)

- V Jeně provedte sjednocení RDF dat z
 - Vámi vytvořeného datasetu 3 kolegů
 - DBpedia (např. <http://dbpedia.org/data/Přerov>)
 - GeoNames (např. <http://sws.geonames.org/3067580/about.rdf>)
- Vymyslete a vyhodnoťte SPARQL dotaz, který vrátí jméno, příjmení kolegů, název místa narození z DBpedia a počet obyvatel z GeoNames
- Stejně informace získajte programově
 - Viz. minulé cvičení
 - http://nb.vse.cz/~svatek/rzzw/RDF_API_Petrak_l11.pdf
 - Resp. viz. tutoriál
 - http://www.openjena.org/tutorial/RDF_API/index.html

Úkol (2)

- Programově získejte absolutní rozdíl mezi počty obyvatel uvedenými v DBpedii a GeoNames

- Příklad:

- DBpedia

`<http://dbpedia.org/page/Přerov>` `<http://dbpedia.org/ontology/populationTotal>` 47373

- GeoNames

`<http://sws.geonames.org/3067580/>` `<http://www.geonames.org/ontology#population>` 47311

- $47373 - 47311 = 62$