

# **Triplifikace**

Reprezentace a zpracování znalostí  
na WWW, 5. 11. 2012

Jindřich Mynarz

# Triplifikace

- Triplifikace = **převod non-RDF dat do RDF**
- Jak se vyrovnat s tím, že ne všechna data jsou k dispozici v RDF?
- Chicken and egg problem (deadlock)

# Nástroje pro zdrojové formáty

- Text
  - Zpracování přirozeného jazyka
- HTML
  - Screen-scraping
- XML
  - XSLT transformace do RDF/XML
- Tabulka
- Relační DB
  - Mapování ze SQL
- Web API
  - Wrappery
- ...a mnoho dalších; viz [1](#) nebo [2](#)

# Příklady

- [DBpedia](#) (zdroj: [Wikipedia](#))
- [World Bank Linked Data](#) ([zdroj](#))
- [LinkedGeoData](#) (zdroj: [OpenStreetMap](#))
- [Linked Data Cup](#) (dříve [LOD Triplification Challenge](#)): [vítěz za rok 2012](#)

# Tipy

- Separation of concerns
  - Nesnažte se během triplifikace vyřešit vše (např. linkování, kvalitu dat).
- Streaming parsery
  - Zvláště pro velká data, nevytěžují tolik paměť.
- ~~RDF model~~
- Turtle (a la NTriples)
  - Výstup: stream 1 řádek/1 trojice
- SPARQL Update
  - Výstup: stream SPARQL Update požadavků

# Zdroje

- [LATC Triplification Experiences](#)
- Heath, Tom; Bizer, Christian. Linked data: evolving the Web into a global data space. (část [5.1.1](#))