

Projekt LOD2 a jeho „Stack“

Doc. Ing. Vojtěch Svátek, Dr.

Zimní semestr 2013

<http://nb.vse.cz/~svatek/rzzw.html>

Projekt LOD2, <http://lod2.eu>

- Kolaborativní projekt v 7. rámcovém programu EU (priorita ICT)
- Délka 4 roky (září 2010 - srpen 2014)
 - VŠE až od září 2011
- 15 partnerů ze 12 zemí
- Koordinátorem je Univerzita v Lipsku (prof. Sören Auer)
- Celkový rozpočet přes 10 mil. eur
 - Z toho VŠE cca 318 tisíc eur

Struktura projektu

- 13 pracovních balíčků („workpackages“, WP), dále rozdělených na úlohy („tasks“)
 - dále **tučně** ty, na kterých se VŠE podílí
- Z toho 6 technologických WPs
 - WP1 – sběr požadavků
 - WP2 – databázová infrastruktura
 - **WP3** - extrakce dat (ze strukturovaných i nestrukturovaných zdrojů) a tvorba datových slovníků
 - **WP4** – propojování a slévání dat
 - WP5 – vizualizace a editování dat
 - **WP6** – integrace nástrojů do **LOD2 Stacku**

Struktura projektu

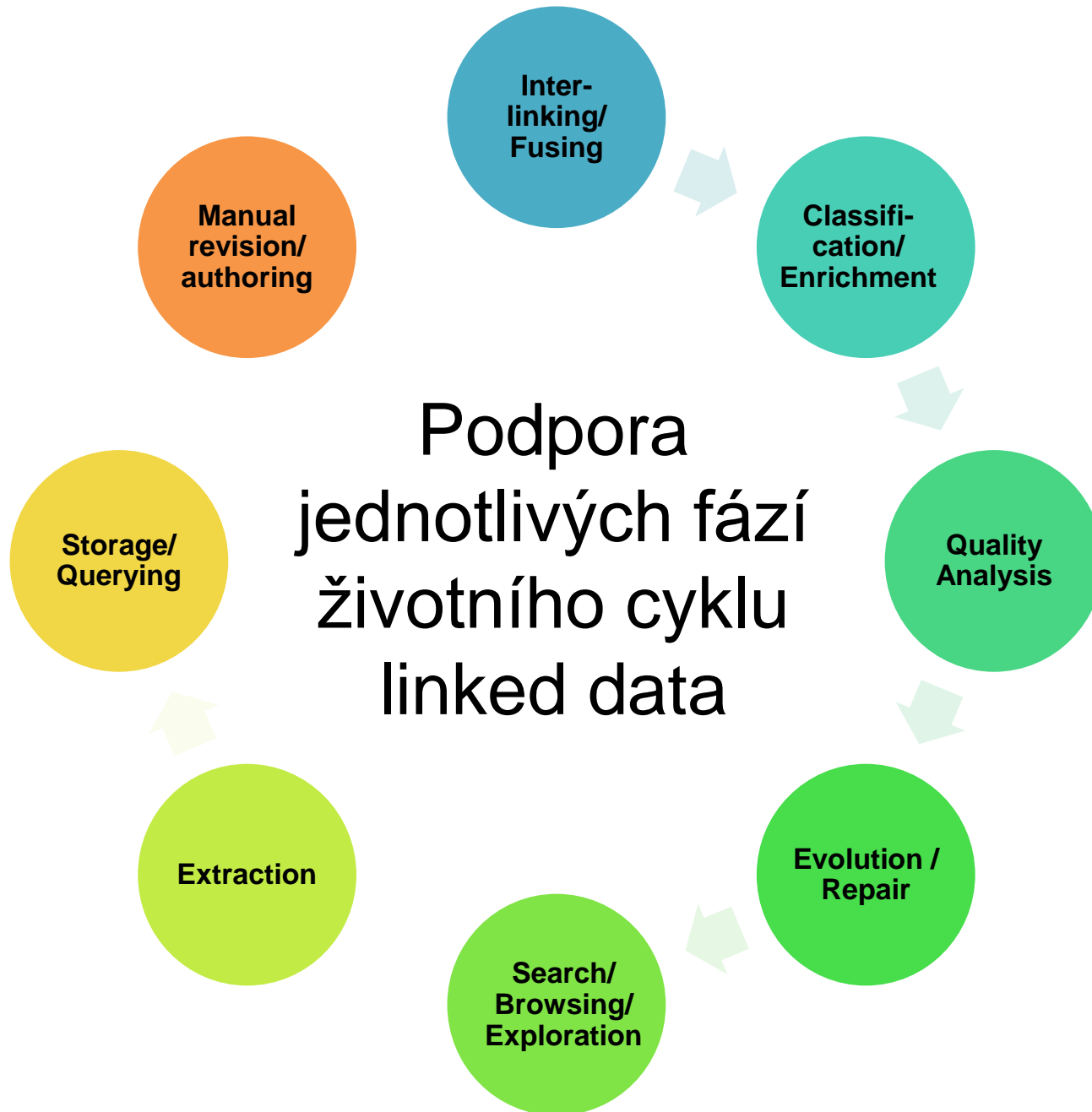
- 4 WPs zaměřených na konkrétní případy užití
 - WP7 – média a publikování (řídí německé vydavatelství Wolters Kluwer)
 - WP8 – podniková data (řídí franc. firma Exalead)
 - **WP9** – data veřejné správy (řídí Open Knowledge Foundation, OKFN, neziskovka s centrálou v Británii)
 - **WP9a** – data o veřejných zakázkách (řídí **VŠE Praha**)
- 3 návazné a podpůrné WPs:
 - **WP10** šíření informací („dissemination“)
 - **WP11** standardizace a uvádění do praxe
 - **WP12** projektový management

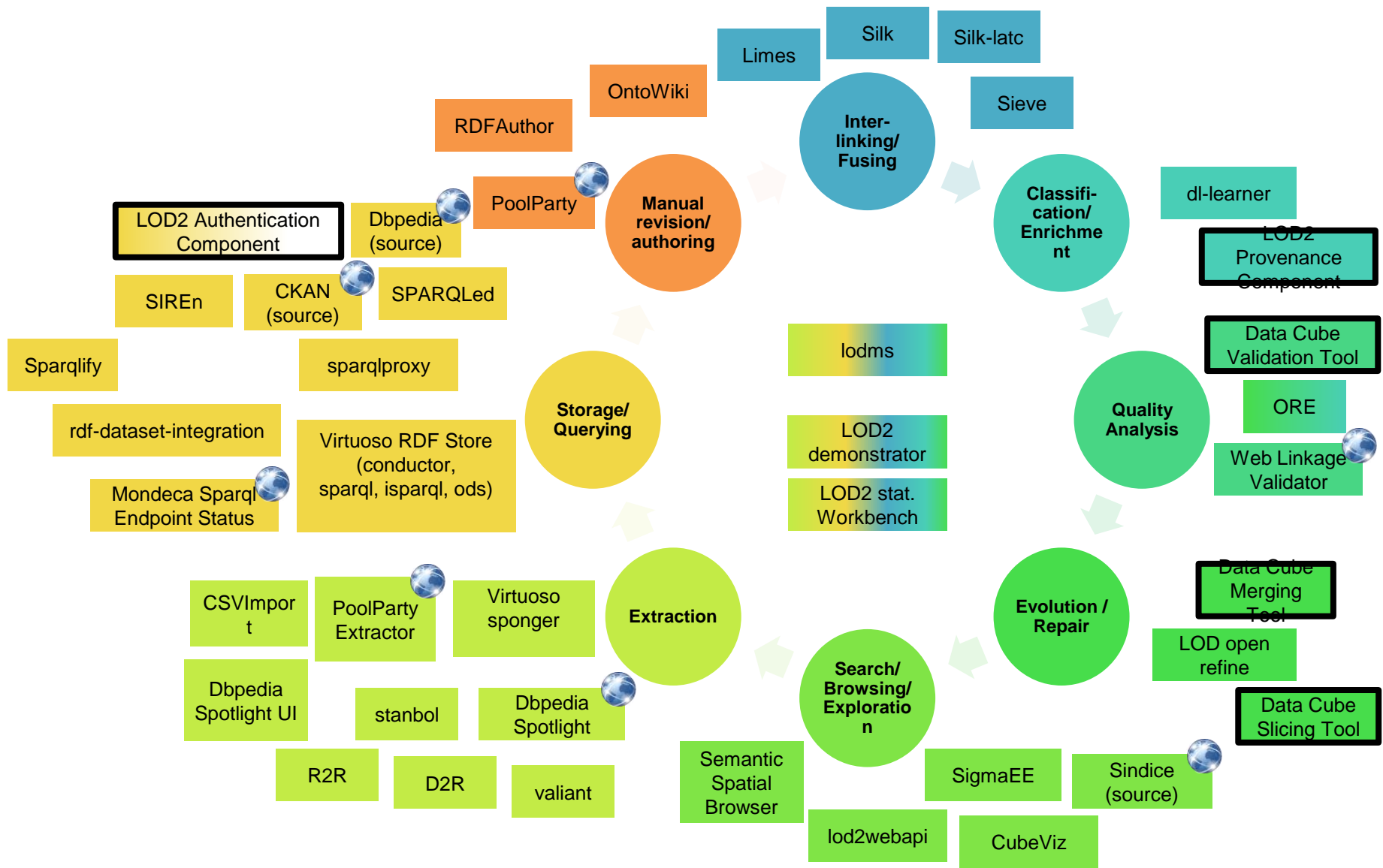
Hodnocení projektu

- Projekt je posuzován zejména na základě oficiálních výstupů („deliverables“)
 - 2 typy: „report“, „prototype“
- Ty procházejí interní oponenturou, a pak se odevzdávají zpravodaji („project officer“, PO) a třem stálým oponentům („reviewers“)
 - PO je Dr. Stefano Bertolo, vedoucí pracovník řízení výzkumu v rámci EK v Lucemburku
- Každoročně probíhá oponentní schůzka za účasti všech partnerů, PO i oponentů
 - Poslední se konala 12.-13.9.2013 v Lucemburku

LOD2 Stack

- Repozitář balíčků pro UNIX (Debian)
 - Aktuálně 47 komponent
 - 24 z nich je přístupných přes některé ze dvou společných grafických rozhraní (generické online demo, „statistický workbench“)
- Lze i jako virtuální stroje do VMware
- Snaha postupně sjednotit autentifikaci apod.
- Online demo na <http://demo.lod2.eu>





Vybrané nástroje

- **Virtuoso** – databázový systém pro RDF, efektivní indexování a zodpovídání SPARQL dotazů
- **Sindice** – vyhledávací index sémantického webu – sbírá a indexuje RDF data v otevřeném prostředí webu
- **SILK** – nástroj pro automatické linkování entit, pomocí vložených nebo natrénovaných pravidel rozpoznává, které entity z různých datasetů jsou totožné
- **SPARQLed** – grafický editor pro podporu tvorby dotazů ve SPARQL
- **CKAN** – nástroj pro správu portálů metadat k otevřeným datům (např. evropský <http://publicdata.eu>, český <http://cz.ckan.net/>)

Vybrané nástroje

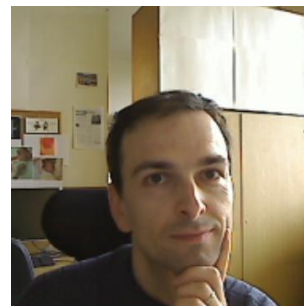
- **DBpedia Spotlight** – automaticky anotuje zadaný volný text entitami z datasetu DBpedia
- **Sparqlify** a **D2R** – umožňují přístup k relačním DB a jiným zdrojům tak, jako by to byly zdroje v RDF
- **OntoWiki** – „sémantický CMS“, místo textu se pohodlně editují přímo data v RDF
- **ORE (Ontology Repair and Enrichment)** – opravuje chyby v datovém slovníku (ontologii) a obohacuje definice pojmů o nové vlastnosti naučené z dat
- **PoolParty** – správa rozsáhlých taxonomií (tezaurů) v podobě linked data, anotování textu pojmy z taxonomie

Tým VŠE v rámci LOD2

- Fakticky spolupráce tří (resp. čtyř) týmů ze dvou (resp. tří) pražských VŠ
 - „Knowledge Engineering Group“ (KEG) na KIZI VŠE
 - „XML and Web Engineering Research Group“ (XRG) na MFF UK
 - Tým zaměřený na otevřená data veřejné správy na KIT VŠE
 - Jednotlivci (doktorandi) z FIT ČVUT

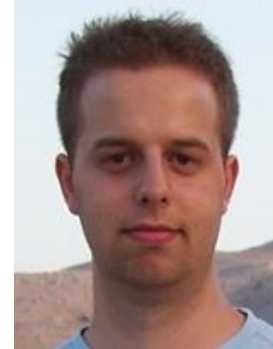
Tým VŠE v rámci LOD2

- **Vojtěch Svátek** – místní koordinátor, aktivity okolo ontologií (WP3, WP4)
- **Jindřich Mynarz** – expert přes linked data, výzkum v oblasti matchmakingu veřejných zakázek (WP9a), organizační a diseminační aktivity
- **Martin Nečaský** (MFF) – vedení týmu MFF, kontakty na českou veřejnou správu
- **Jan Kučera** (KIT) – byznys modely a licenční otázky otevřených dat, kontakty na českou veřejnou správu



Tým VŠE v rámci LOD2

- **Jakub Klímek** (MFF) – matchmaking a vizualizace RDF dat
- **Tomáš Knap** (MFF) – extrakce, čištění a propojování RDF dat
- **Ondřej Zamazal** – technologie transformace a opravování ontologií (WP3, WP4)
- **Marek Dudáš** – datové slovníky pro oblast produktů a služeb, vizualizace ontologií
- **Simone Serra** – anotování datových slovníků



Tým VŠE v rámci LOD2

- **Ondřej Vadinský** – správa serveru s LOD2 Stackem
- **Milan Dojchinovski** (FIT) a **Tomáš Kliegr** – využití zpracování přirozeného jazyka pro obohacování zdrojů linked data
- prof. **Petr Berka** – spolupráce na dataminingu z linked data
- Studenti: Jiří Helmich (MFF), Ján Černý, Michal Pomykacz, ... a v třeba **kdokoli další z Vás!**