

4IZ440 Propojená data na webu

# Datové slovníky pro RDF

Vyučující: prof. Ing. Vojtěch Svátek, Dr.




*Zimní semestr 2019*

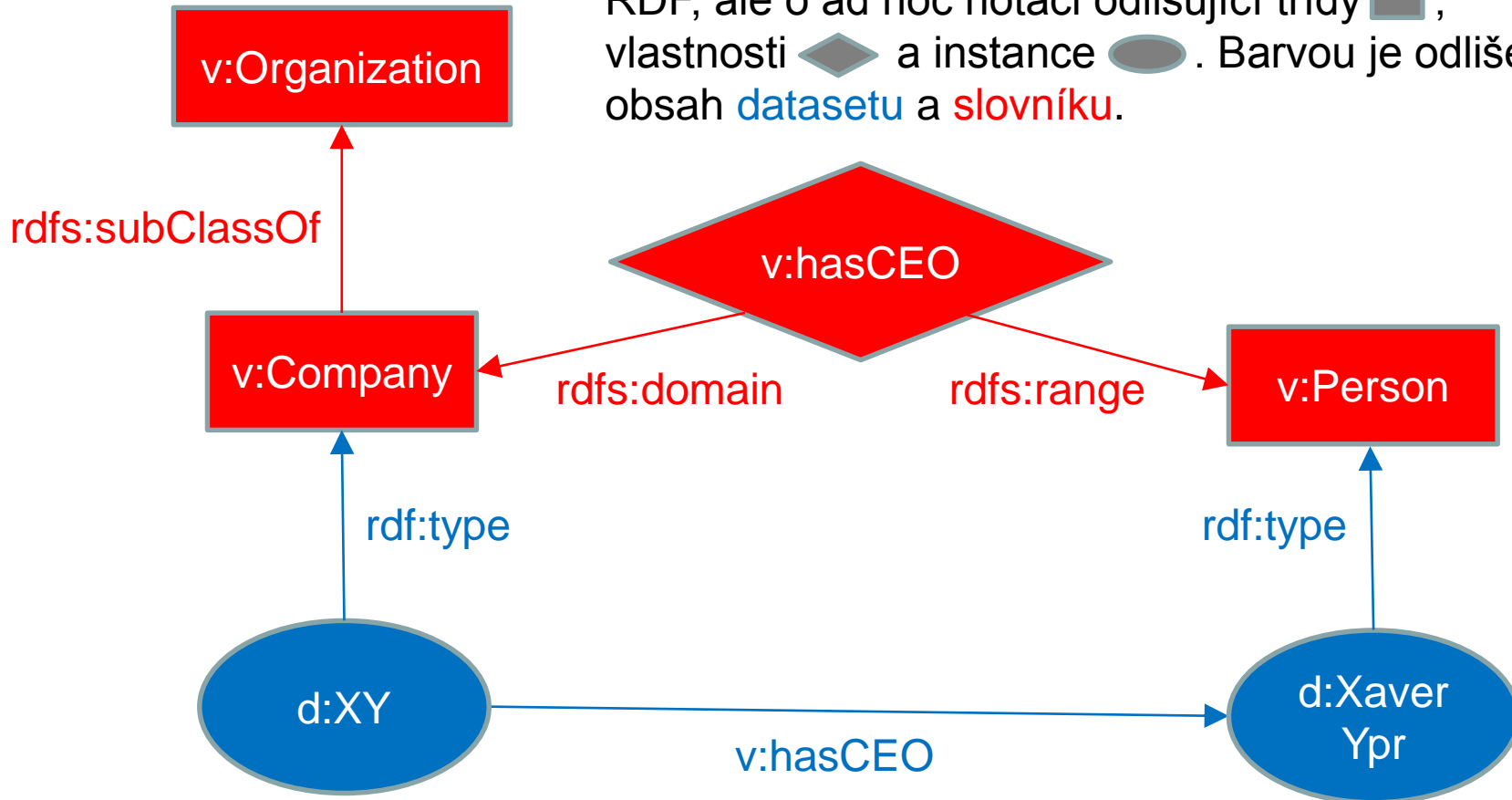
<http://nb.vse.cz/~svatek/rzzw.html>

# Co potřebujeme definovat v (jednoduchém) slovníku

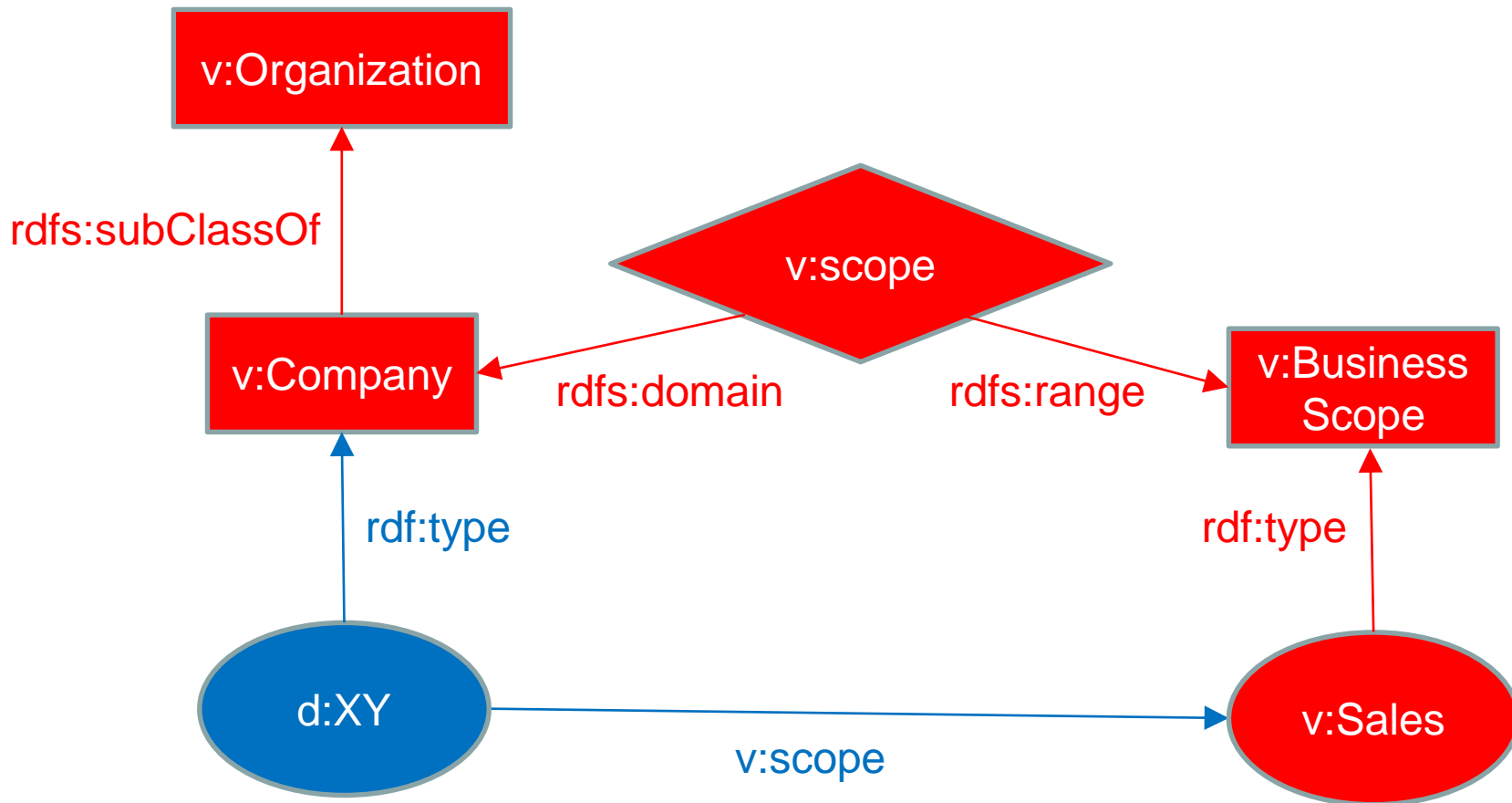
- **Třídy**, kterými lze zdroje RDF typovat
- Typy vztahů, neboli **vlastnosti** (predikáty), kterými lze zdroje RDF propojovat
- „Standardní“ zdroje, které lze jako hodnoty vlastností přiřazovat: označují se jako **individua**
- Tyto tři typy **entit** jsou ve slovníku propojené do logických struktur

# Modelový příklad struktury I.

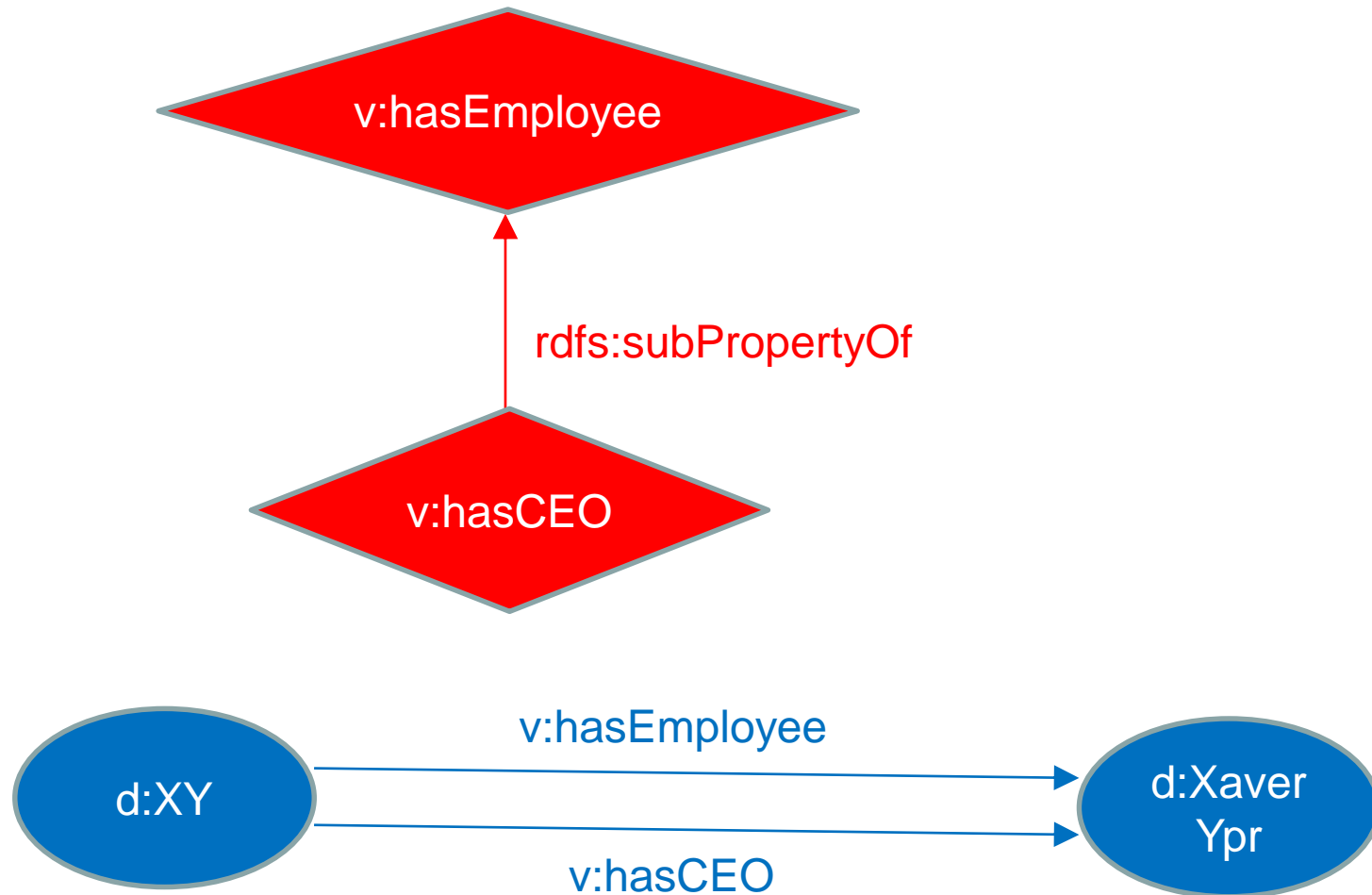
Pozor, nejde o dříve zavedenou grafickou notaci RDF, ale o ad hoc notaci odlišující třídy , vlastnosti  a instance . Barvou je odlišen obsah **datasetu** a **slovníku**.



# Modelový příklad struktury II.



# Modelový příklad struktury III.



# Slovníky a ontologie

- Obecně: jazyky RDF (resp. RDFS jako jeho „schémátová“ část) a OWL tvoří jeden propojený nástroj pro tvorbu slovníků pro RDF data
- Slovníky RDF jsou samy o sobě (syntakticky) RDF daty, zpravidla doplněnými o textově zapsané specifikace
- Mezi pojmy „slovník“ a „ontologie“ neexistuje předěl
  - Pokud slovník obsahuje komplikovanější logické formule definující třídy pomocí jiných tříd, označujeme ho často jako ontologii
  - Ale toto je jen nezávazná konvence; obyčejný vztah *podtřídy* nebo určení *definičního oboru / oboru hodnot* je také logická formule (i když umožňuje jen omezenější okruh odvození)

# Typické modely použití slovníku

- Dataset používá celý slovník nebo jeho významnou část; slovník je pro dataset dominantní
  - Profil určitého objektu: FOAF, SIOC, DOAP... („strom“)
  - Znalostní graf s mnoha propojenými objekty: DBpedia („graf“)
  - Datová kostka: Data Cube Vocabulary („tabulková data“)
- Dataset používá jen několik entit, ve spojení s mnoha jinými jmennými prostory
  - Dublin Core: dc:subject, dc:author...
  - GEO: wgs84\_pos:alt, wgs84\_pos:lat...
- Ale existují i hybridní modely

# Důležité slovníky

- *Meta-slovníky*: RDF, OWL
- Specifičtější slovníky (z pohledu tohoto předmětu; v různých komunitách může být pohled odlišný)
  - Dublin Core; Schema.org; FOAF; SKOS; Data Cube; Prov; Geo; VOID; DCAT; Adms; ... aj., viz též LD Book, kap. 4.4.4-4.4.5
  - Vesměs zkatalogizované na <https://lov.linkeddata.es/>
- Samostatnou kategorií tvoří *biomedicínské* nomenklatury/ontologie
  - Často důraz spíše na systémy pojmů (rozsáhlé taxonomie) než na využití jako schéma pro publikování dat
  - Viz <https://bioportal.bioontology.org/>