

# Technické standardy vytváření datových schémat pro datové sady na stupni otevřenosti 5

Obdobou datového schématu pro datové sady na stupni otevřenosti 5 jsou tzv. ontologie (někdy též zvané slovníky). Zatímco však ve světě tabulkových či XML dat popisuje datové schéma striktní syntaxi např. CSV či XML distribucí datových sad, ontologie popisuje datovou sadu na konceptuální úrovni. Konkrétně definuje třídy a vlastnosti, které lze v RDF distribucích datových sad použít.

## Standardy pro vyjádření a zápis ontologií

Standardem pro zápis ontologií jsou jazyky RDF Schema (RDFS, <http://www.w3.org/TR/rdf-schema/>) a Web Ontology Language (OWL, <http://www.w3.org/TR/owl/>). Oba jazyky předpokládají ontologii reprezentovanou v datovém modelu RDF a umožňují využít pro zápis ontologie libovolný formát pro zápis RDF datového modelu (viz [Otevřená formální norma pro Propojená data.](#)).

## Standardy pro tvorbu ontologií

Primárním cílem při návrhu tříd a vlastností použitých v RDF distribuci datové sady musí být maximální znovupoužitelnost údajů v distribuci napříč různými SW aplikacemi. Proto je standardem co nejvíce využívat tříd a vlastností z již existujících ontologií. Datový model RDF je připraven k tomu, aby bylo možné v jedné RDF distribuci kombinovat třídy a vlastnosti z několika různých ontologií. V rámci různých iniciativ již vznikla v mezinárodním kontextu celá řada ontologií, které jsou katalogizovány např. v projektu Linked Open Vocabularies (LOV, <http://lov.okfn.org>).

Při návrhu tříd a vlastností použitých v RDF distribuci je proto standardem postupovat následovně:

- Identifikujeme typy entit vyskytujících se v datové sadě a jejich atributy a vztahy mezi nimi. (Jinými slovy tvoříme konceptuální schéma datové sady.)
- Vybereme třídy a vlastnosti definované existujícími ontologiemi, jejichž sémantika přímo odpovídá sémantice typů entit a jejich atributům a vztahům mezi nimi, které jsme identifikovali v kroku 1.
- V případě, že se v kroku 2 nepodařilo pro některé typy entit, resp. jejich atributy nebo vztahy mezi nimi nalézt odpovídající třídy či vlastnosti z existujících ontologií, vybereme třídy a vlastnosti definované existujícími ontologiemi, jejichž sémantika je obecnější než sémantika těchto zbylých typů entit, resp. jejich atributů nebo vztahů mezi nimi. Pro tyto zbylé typy entit, resp. jejich atributy a vztahy mezi nimi vytvoříme nové třídy, resp. vlastnosti v naší vlastní ontologii a propojíme je s vybranými třídami a vlastnostmi z existujících ontologií s obecnější sémantikou pomocí ISA vazby (pro technické vyjádření ISA vazby viz dále).
- V případě, že se v kroku 3 nepodařilo pro některé typy entit, resp. jejich atributy nebo vztahy mezi nimi nalézt třídy či vlastnosti z existujících ontologií s obecnější sémantikou, vytvoříme pro ně nové třídy, resp. vlastnosti v naší vlastní ontologii.
- V případě, že po vypublikování RDF distribuce zjistíme, že jsme v kroku 2 udělali chybu a zavedli jsme třídu či vlastnost pro typ entity, resp. atribut či vztah, pro níž existuje sémantický

ekvivalent v existující ontologii, nebo zjistíme, že se objevila nová ontologie s takovým sémantickým ekvivalentem, propojíme naši třídu či vlastnost s jejím nalezeným sémantickým ekvivalentem pomocí ekvivalenční vazby (pro technické vyjádření ekvivalenční vazby viz dále).

- V případě, že po vypublicování RDF distribuce zjistíme, že jsme v kroku 3 udělali chybu a neidentifikovali jsme ISA vazbu, nebo se objevila nová ontologie pro kterou ISA vazba existuje, tuto ISA vazbu vytvoříme (pro technické vyjádření ISA vazby viz dále).

Standardem je využívání především následujících ontologií:

- DCMI Metadata Terms (<http://purl.org/dc/terms/>)
- Simple Knowledge Organization System (<http://www.w3.org/2004/02/skos/core#>)
- Schema.org (<http://schema.org/>)
- Data Cube Vocabulary (<http://www.w3.org/TR/vocab-data-cube/>)
- Friend-of-a-Friend (<http://xmlns.com/foaf/0.1/>)
- The organization ontology (<http://www.w3.org/TR/vocab-org/>)
- FRBR-aligned bibliographic ontology (<http://purl.org/spar/fabio/>)

## Standardy propojování prvků ontologií

Mezi prvky ontologií (tj. mezi třídami nebo mezi vlastnostmi) mohou existovat sémantické vazby. Může se jednat o vazby mezi prvky stejné ontologie nebo mezi prvky dvou různých ontologií. Tyto vazby je nutno v ontologiích vyjádřit. Vazbu vždy vyjádříme v ontologii, která definuje prvek, pro který chceme vazbu vyjádřit. Pro vyjádření vazby postupujeme podle následujícího standardu:

- pro vyjádření ISA vazby mezi třídami A a B, kdy třída A má specifičtější sémantiku a třída B má obecnější sémantiku, vytvoříme v zápisu ontologie následující trojici
  - subjekt : IRI třídy A
  - predikát : <http://www.w3.org/2000/01/rdf-schema#subClassOf>
  - objekt : IRI třídy B
- pro vyjádření ISA vazby mezi vlastnostmi A a B, kdy vlastnost A má specifičtější sémantiku a vlastnost B má obecnější sémantiku, vytvoříme v zápisu ontologie následující trojici
  - subjekt : IRI vlastnosti A
  - predikát : <http://www.w3.org/2000/01/rdf-schema#subPropertyOf>
  - objekt : IRI vlastnosti B
- pro vyjádření ekvivalenční vazby mezi třídami A a B, kdy sémantika třídy A je ekvivalentní sémantice třídy B, vytvoříme v zápisu ontologie následující trojici
  - subjekt : IRI třídy A
  - predikát : <http://www.w3.org/2002/07/owl#equivalentClass>
  - objekt : IRI třídy B
- pro vyjádření ekvivalenční vazby mezi vlastnostmi A a B, kdy sémantika vlastnosti A je ekvivalentní sémantice vlastnosti B, vytvoříme v zápisu ontologie následující trojici
  - subjekt : IRI vlastnosti A
  - predikát : <http://www.w3.org/2002/07/owl#equivalentProperty>
  - objekt : IRI vlastnosti B

# Standardy tvorby IRI tříd a vlastností definovaných v ontologiích

V případě, že poskytovatel dat definuje vlastní predikáty či třídy v podobě ontologie, potom pro ně musí volit IRI ve tvaru URL v následující podobě:

- v případě, že poskytovatel zavádí pouze jednu ontologii, má ontologie IRI ve tvaru {základ-IRI}/ontology, kde {základ-IRI} je základ IRI poskytovatele dat (viz [Otevřená formální norma pro Propojená data](#))
- v případě, že poskytovatel zavádí více ontologií, mají ontologie IRI ve tvaru {základ-IRI}/ontology/{určení-ontologie} kde {určení-ontologie} je validní část IRI, která jednoznačně určuje ontologii v kolekci ontologií definovaných poskytovatelem
- třídy mají IRI ve tvaru {IRI-ontologie}/{NázevTřídy} kde {IRI-ontologie} je IRI ontologie, ve které je třída definována a {NázevTřídy} je název třídy unikátní v rámci kolekce všech tříd zavedených v ontologii zapsaný v CamelCase notaci (tj. jednotlivá slova tvořící název třídy jsou spojena za sebe, první znak každého slova je velký a všechny ostatní znaky jsou malé)
- predikáty mají IRI ve tvaru {IRI-ontologie}/{názevPredikátu} kde {IRI-ontologie} je IRI ontologie, ve které je třída definována a {názevPredikátu} je název predikátu unikátní v rámci kolekce všech predikátů zavedených v ontologii zapsaný v camelCase notaci (tj. jednotlivá slova tvořící název třídy jsou spojena za sebe, první znak každého slova kromě prvního slova je velký a všechny ostatní znaky jsou malé)

From:

<https://opendata.gov.cz/> - **Otevřená data**

Permanent link:

<https://opendata.gov.cz/standardy:technicke-standardy-pro-vytvareni-datovych-schemat-na-stupni-5>

Last update: **2020/06/03 09:37**

