



## Zadání bakalářské práce

<b>Název:</b>	Využití ontologické analýzy pro zajištění sémantické interoperability heterogenních dat
<b>Student:</b>	Václav Šír
<b>Vedoucí:</b>	doc. Ing. Robert Pergl, Ph.D.
<b>Studijní program:</b>	Informatika
<b>Obor / specializace:</b>	Informační systémy a management
<b>Katedra:</b>	Katedra softwarového inženýrství
<b>Platnost zadání:</b>	do konce letního semestru 2022/2023

### Pokyny pro vypracování

Téma přispívá k projektu Datového inkubátoru dat. Cílem práce je ontologická analýza klíčových domén a jejich propojení s datovými sadami tak, aby byla umožněna jejich sémantická interoperabilita.

1. Seznamte se s projektem Datového inkubátoru, problematikou sémantické interoperability, Unified Foundational Ontology, jazykem OntoUML a nástrojem OpenPonk.
2. Ve spolupráci s vedoucím vyberte několik klíčových domén a souvisejících datových sad.
3. Vytvořte ontologické konceptuální modely těchto domén.
4. Propojte ontologické konceptuální modely s datovými sadami a vytvořte pravidla pro jejich mapování.
5. Zdokumentujte své řešení a přínos pro zajištění sémantické interoperability.



Bakalářská práce

**VYUŽITÍ ONTOLOGICKÉ  
ANALÝZY PRO  
ZAJIŠTĚNÍ SÉMANTICKÉ  
INTEROPERABILITY  
HETEROGENNÍCH DAT**

Václav Šír

Fakulta informačních technologií  
Katedra softwarového inženýrství  
Vedoucí: doc. Ing. Robert Pergl, Ph.D.  
10. května 2022

České vysoké učení technické v Praze

Fakulta informačních technologií

© 2022 Václav Šír. Odkaz na tuto práci.

*Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení na předchozí straně, je nezbytný souhlas autora.*

Odkaz na tuto práci: Šír Václav. *Využití ontologické analýzy pro zajištění sémantické interoperability heterogenních dat*. Bakalářská práce. České vysoké učení technické v Praze, Fakulta informačních technologií, 2022.

# Obsah

Poděkování	vii
Prohlášení	viii
Abstrakt	ix
Seznam zkratk	x
Úvod	1
1 Cíl práce	3
<b>I Teoretická příprava</b>	<b>5</b>
<b>2 Ontologie</b>	<b>7</b>
2.1 Rozdělení ontologie . . . . .	7
2.1.1 Popisná ontologie . . . . .	7
2.1.2 Formální ontologie . . . . .	7
2.1.3 Formalizovaná ontologie . . . . .	7
2.2 Ontologie v informatice . . . . .	8
2.2.1 Vyšší ontologie . . . . .	8
2.2.2 Doménové ontologie . . . . .	8
2.3 Využití ontologie v praxi . . . . .	9
2.4 Unified Foundational Ontology . . . . .	9
2.4.1 Mikroteorie . . . . .	9
2.4.2 Kategorie UFO . . . . .	10
<b>3 UML</b>	<b>11</b>
3.1 Sémantické kategorie . . . . .	11
3.2 UML diagram tříd . . . . .	11
<b>4 OntoUML</b>	<b>13</b>
4.1 Třídy . . . . .	13
4.2 Princip Identity . . . . .	13
4.2.1 Sortály . . . . .	13
4.3 Relace . . . . .	15
4.3.1 Typy relací . . . . .	16
4.4 Teorie celku a části . . . . .	17
4.4.1 Definice z UML . . . . .	18
4.4.2 Závislosti mezi celkem a částí . . . . .	18
4.4.3 Stereotypy celků a částí . . . . .	20

<b>5</b>	<b>Sémantická interoperabilita</b>	<b>23</b>
5.1	Souvislost s datovými sadami	23
5.2	Sémantická interoperabilita v OntoUML	23
5.3	Sémantická interoperabilita a datový inkubátor	23
5.4	FAIR	24
5.4.1	FAIRifikační proces	24
<b>6</b>	<b>OpeNest</b>	<b>27</b>
6.1	Projekt OpeNest	27
6.1.1	Datová platforma	27
6.1.2	Inkubátor	27
6.2	Otevřené standardy pro data	28
6.2.1	Typy otevřených standardů	28
<b>7</b>	<b>OpenPonk</b>	<b>31</b>
7.1	OntoUML	31
<b>II</b>	<b>Vlastní řešení</b>	<b>33</b>
<b>8</b>	<b>Datové sady</b>	<b>35</b>
<b>9</b>	<b>Ontologické konceptuální modely</b>	<b>37</b>
9.1	Šablony	37
9.2	Tvorba konceptuálního modelu	38
9.3	Praktický příklad	42
9.3.1	Telefonní síť a operátoři	42
9.3.2	SIM karta	44
9.3.3	Hlasový provoz	44
9.3.4	Propojení domén	44
<b>10</b>	<b>Datový model</b>	<b>47</b>
10.1	Datové entity	47
10.1.1	Standard názvu	47
10.1.2	Vlastnosti entit	48
10.2	Mapovací pravidla	48
10.2.1	Předpisy mapovacích pravidel	48
10.3	Praktický příklad	49
10.3.1	Mapování podnikatelů	50
10.3.2	Mapování kategorie a hodnoty	50
<b>11</b>	<b>Zajištění sémantické interoperability</b>	<b>53</b>
	<b>Závěr</b>	<b>55</b>
	<b>A Příklady sémantické interoperability</b>	<b>57</b>
	<b>Obsah přiloženého média</b>	<b>67</b>

## Seznam obrázků

2.1	Příklad vyšší ontologie [9, překlad vlastní]	8
2.2	Příklad doménové ontologie [9, překlad vlastní]	9
4.1	Ukázka vazby 1:1	16
4.2	Ukázka vazby 1:N	16
4.3	Ukázka vazby M:N	16
4.4	Ukázka dekomponované materiální relace [21]	17
4.5	Ukázka generalizace	18
5.1	Proces FAIRifikace [25, překlad vlastní]	25
9.1	Koncové entity dle struktury	39
9.2	Namodelované koncové entity e-komunikace	40
9.3	Ukázka propojení mateřské a koncové entity	41
9.4	Původní šablona pro podnikatele	43
10.1	Ukázka datové entity <i>Označení podnikatele</i>	50
10.2	Datová entita pro kategorie a hodnoty	51
A.1	Entita SIM karta v modelu trh-mobilnich-siti	57
A.2	Entita SIM karta v modelu trh-mobilnich-siti	58
A.3	Entita telefonní hovor v modelu trh-mobilnich-siti	58
A.4	Entita telefonní hovor v modelu MML-mobilni-telefon	59
A.5	Entita SMS zpráva v modelu trh-mobilnich-siti	60
A.6	Entita SMS zpráva v modelu MML-mobilni-telefon	60
A.7	Entita mobilní operátor v modelu trh-mobilnich-siti	61
A.8	Entita mobilní operátor v modelu MML-mobilni-telefon	62

## Seznam tabulek

10.1	Popis struktury datové sady po transformaci	49
------	---	----

## Seznam výpisů kódu

10.1	Základní mapovací pravidlo . . . . .	48
10.2	Entita podmíněná obsahem vlastnosti . . . . .	48
10.3	Mapování na více entit . . . . .	49
10.4	Mapování se složitější podmínkou . . . . .	49
10.5	Podmíněné pravidlo . . . . .	49
10.6	Mapovací pravidlo pro označení podnikatele . . . . .	50
10.7	Ukázka mapovacích pravidel pro kategorii a hodnotu . . . . .	52



*Chtěl bych především poděkovat doc. Ing. Robertu Perglovi, Ph.D. za příležitost vytvořit tuto bakalářskou práci, za odborné konzultace a vedení bakalářské práce. Dále bych chtěl poděkovat Bc. Tereze Macháčové a Bc. Janě Martínkové za vedení modelovacího týmu, za cenné rady a konzultace. Naposled bych chtěl poděkovat rodině za veškerou poskytnutou podporu a pomoc.*

## Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 2373 odst. 2 zákona č. 89/2012 Sb., občanský zákoník, ve znění pozdějších předpisů, tímto uděluji nevýhradní oprávnění (licenci) k užití této mojí práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užít. Tyto osoby jsou oprávněny Dílo užít jakýmkoli způsobem, který nesnižuje hodnotu Díla, avšak pouze k nevýdělečným účelům. Toto oprávnění je časově, teritoriálně i množstevně neomezené.

V Praze dne 10. května 2022

.....

## Abstrakt

Tato bakalářská práce se zabývá využitím ontologické analýzy k propojení dat mezi různými klíčovými doménami.

Součástí práce byla ontologická analýza vybraných klíčových domén, které jsou zpracovávány v rámci projektu OpeNest. Pro vizuální zachycení analýzy byly využity ontologické konceptuální modely vytvořené v modelovacím jazyce OntoUML. Na základě konceptuálních modelů jsou poté vytvořeny datové modely, ve kterých je zachyceno mapování dat, obsažených v datové sadě klíčové domény.

V rámci práce byly vytvořeny dva ontologické konceptuální modely s mapováním dat a dvě šablony pro modelování. Vypracované modely i šablony jsou součástí datové platformy, která je jednou z komponent projektu OpeNest.

Závěr práce obsahuje analýzu splnění sémantické interoperability mezi vypracovanými modely a již existujícími modely.

**Klíčová slova** sémantická interoperabilita, ontologická analýza, konceptuální model, datový inkubátor, heterogenní data, FAIR data, UFO, Remmark, a.s., OpeNest, OntoUML, OpenPonk

## Abstract

This bachelor thesis deals with the usage of ontological analysis for connecting data between various vital domains.

This thesis includes an ontological analysis of chosen vital domains processed within project OpeNest. For visual capture of the analysis were used ontological conceptual models. Models were made using the modeling language OntoUML. On top of the conceptual models were created data models containing a mapping of data defined in a dataset in the vital domain.

Two ontological conceptual models with data mapping and two modeling templates were made during this thesis's making. Models and templates are part of a data platform that is one of the components of project OpeNest.

The end of the thesis contains the analysis of semantic interoperability between created models and existing models.

**Keywords** semantic interoperability, ontological analysis, conceptual model, data incubator, heterogeneous data, FAIR data, UFO, Remmark, a.s., OpeNest, OntoUML, OpenPonk

## Seznam zkratk

AACR	American Association for Cancer Research
BORM	Business Objects Relation Modelling
BPMN	Business Process Model and Notation
CCMi	Centrum pro Konceptuální Modelování a implementace
ČR	Česká republika
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
FAIR	Findability, Accessibility, Interoperability, and Reuse of digital assets
GFO	General Formalized Ontology
MML	Model Marketing a Lifestyle
MMS	Multimedia Messaging Service
ODI	Open Data Institute
PNG	Portable Network Graphics
PR	Public Relations
SIM	Subscriber Identity Module
SMS	Short Message Service
SVG	Scalable Vector Graphics
UFO	Unified Foundational Ontology
UML	Unified Modeling Language

# Úvod

Zpracování dat se stává v dnešní době rozsáhlým problémem. Množství dat se neustále rozrůstá a s tím počet kolizí v datech, proto je nutné data nejprve důkladně analyzovat a najít tak případné souvislosti. Pomocí analýzy je snahou dospět k sémantické interoperabilitě mezi daty. Ta zaručí efektivní zpracování dat, a to díky eliminaci duplikací dat se sémanticky stejným významem či sdružování s rozdílným významem, ale se stejným názvem. Jelikož zveřejňování dat není plně standardizované, je použití různých datových sad velmi náročné. Tuto skutečnost si uvědomili lidé z Open data institutu a vydali takzvané otevřené standardy pro data. Bohužel i s těmito standardy se vyskytl problém a tím je jejich dodržování. Jelikož nejsou standardy ještě plně zavedené, tak společnosti často nesplňují všechny požadavky standardů. Často o nesplňování požadavků ani nevědí, protože standardům plně nerozumí.

Na tento problém navazuje projekt OpeNest od společnosti Remmark, a.s. Jedním z cílů projektu je osvěta v sekci sběru a poskytování dat. Záměrem osvěty je dodržování zmíněných standardů zveřejněných dat. Jejich dodržení poté usnadní integraci dalších datových sad do projektu. Součástí projektu je datový inkubátor, který obsahuje zpracované datové sady, otevřené i placené, jež jsou dostupné pro užití. Datové sady jsou přístupné v datové platformě, kde jsou shromážděna na jednom místě a mají zajištěnou sémantickou interoperabilitu, to zaručuje uživatelům příjemnější práci s daty a odstranění nutnosti vlastního skladování a zpracování dat.

Jádrem závěrečné práce je zajištění sémantické interoperability mezi vybranými datovými sadami. Té je dosaženo pomocí ontologické analýzy obsahu datové sady a porovnáním s již zpracovanými sadami. Z analýzy vyplynou entity, které se namodelují. Model se poté obohatí o potřebné detaily a propojení, aby v dostatečné míře odpovídal skutečnosti. Jednotlivé modely datových sad se po vložení do datové platformy naváží na sebe pomocí jmen entit. Obsah sad se navazuje na ontologické modely pomocí mapovacích pravidel. Ty se vážou k datovým entitám, které jsou napojené na své ontologické protějšky.

K usnadnění modelování se používají šablony. Jedná se o samostatné modely, bez napojení na data, které obsahují často používané entity (např. problematika právní osobnosti). V modelu datové sady se pak využívá jen podmnožina šablony, které je společná s modelovanou klíčovou doménou. Využívání šablon zrychluje proces modelování domén a díky jednotnému pojmenování entit splňuje sémantickou interoperabilitu mezi modely.





## Kapitola 1

# Cíl práce

Ústředním cílem bakalářské práce je ontologická analýza klíčových domén a jejich propojení s datovými sadami tak, aby byla umožněna jejich sémantická interoperabilita. Celá bakalářská práce spadá pod projekt OpeNest realizovaný společností Remmark, a.s.

Prvním cílem je průzkum transformovaných datových sad, dodaných společností Remmark, a.s. a jejich následná ontologická analýza. Analýza slouží jako základ pro vytváření konceptuálních modelů.

Dalším cílem je vypracování ontologických konceptuálních modelů z jednotlivých zanalyzovaných datových sad. V konceptuálních modelech je zachycen výstup ontologické analýzy.

Konečným cílem je realizace propojení obsahu datových sad a jim příslušných ontologických modelů pomocí mapovacích pravidel navázaných na datové entity v modelech. Realizace mapování je závislá na transformacích. Transformace předchází ontologické analýze a nejsou součástí této bakalářské práce.

Vypracované modely budou využívány v rámci datové platformy společnosti Remmark, a.s. Projekt OpeNest, jenž platformu zaštiťuje, má za cíl vytvoření inkubátoru firem. Dalším přínosem bakalářské práce je zlepšování platformy OpenPonk, ve které jsou vytvářeny modely.





Část I

**Teoretická příprava**



# Ontologie

Ontologie vychází ze stejnojmenné filosofické disciplíny, která se zabývá zkoumáním bytí. Slovo ontologie pochází ze spojení řeckých slov *ontos* (jsoucí) a *logos* (výklad). Důkazy o její existenci sahají až do 17. století. [1] Snaží se o vytvoření pouhé abstrakce reality, to znamená určení *vlastnosti* a *činností*, které objekty reality provádí, či vykazují.

## 2.1 Rozdělení ontologie

Ontologie se rozděluje na tři hlavní kategorie. Stejně jako obecná ontologie mají i její kategorie základ ve filosofii. Kategorie nejsou všeobecně uznávané. Často se formální ontologie zahrnuje buď do popisné, nebo do formalizované. [2, překlad vlastní]

### 2.1.1 Popisná ontologie

Zabývá se sběrem informací obecných či ze specifické domény, které mohou být jak *reálné*, tak *ideální* (imaginární). [2, 3, překlad vlastní]

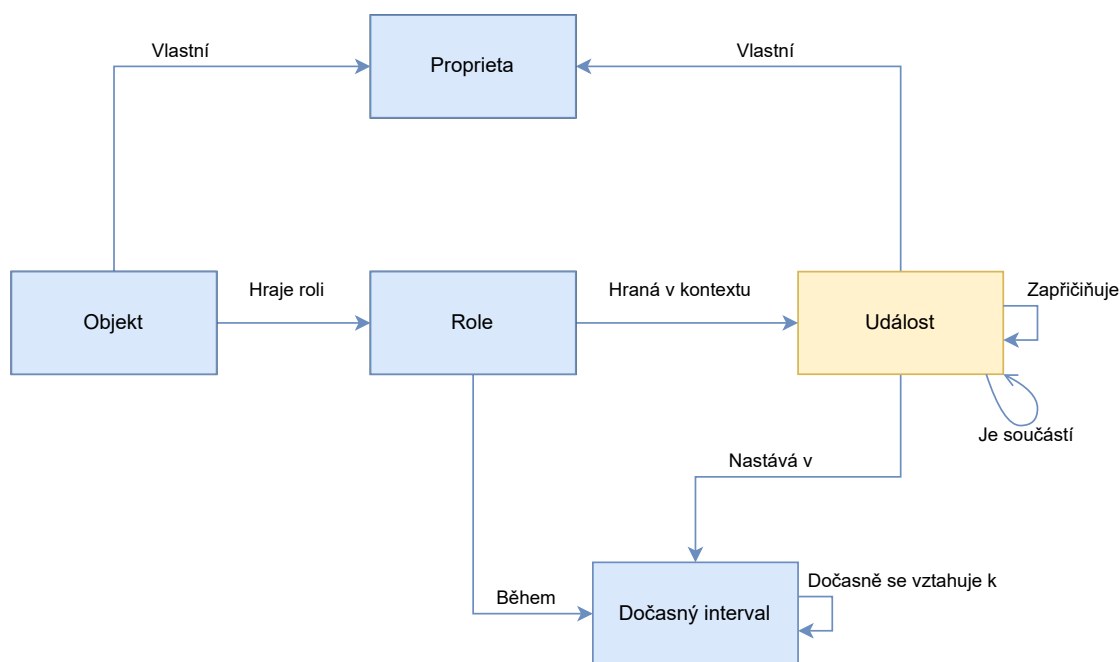
### 2.1.2 Formální ontologie

Formální ontologie se stará o zpracování výstupu popisné ontologie. Informace obohacuje o popis jejich vlastností, kategorizaci, díky kterým se mohou informace filtrovat a organizovat. [2, překlad vlastní]

Původní specifikaci Formální ontologie definoval filosof Edmund Husserl. Ten ve svém díle *Logická zkoumání* jako první rozlišil formální logiku a formální ontologii. Formální logika se zabývá zkoumáním pravdy a jejími atributy (validita, konzistence, důkazy). Na druhou stranu formální ontologie se zabývá věcmi a propojováním s jejich konkrétními vlastnostmi, ale i příbuznými objekty. V neposlední řadě definuje způsob a vlastnosti propojení neboli relace. [2, 4, 3, 5, překlad vlastní]

### 2.1.3 Formalizovaná ontologie

Vytvoření formální kodifikace pro data, která vzešla z předchozích druhů ontologie. [2, 3, překlad vlastní] Z podstaty formální kodifikace vyplývá, že se při jejím použití může dosáhnout sémantické interoperability (více v kapitole 5), jelikož je snahou, aby byl obsah strukturovaný a jednoznačný. [6, překlad vlastní]



■ **Obrázek 2.1** Příklad vyšší ontologie [9, překlad vlastní]

## 2.2 Ontologie v informatice

Ontologie se v informatice používá pro zjištění a zachycení vlastností objektů a konceptů. Při ontologické analýze daného problému jsou zjišťovány vnitřní a vnější souvislosti mezi objekty. Dále se používá pro kategorizaci věcí, určení jejich vlastností, ať už vrozených, nebo časem získaných. Ontologická analýza má za úkol dostatečně abstraktně popsat realitu k úspěšnému vyřešení daného problému.

### 2.2.1 Vyšší ontologie

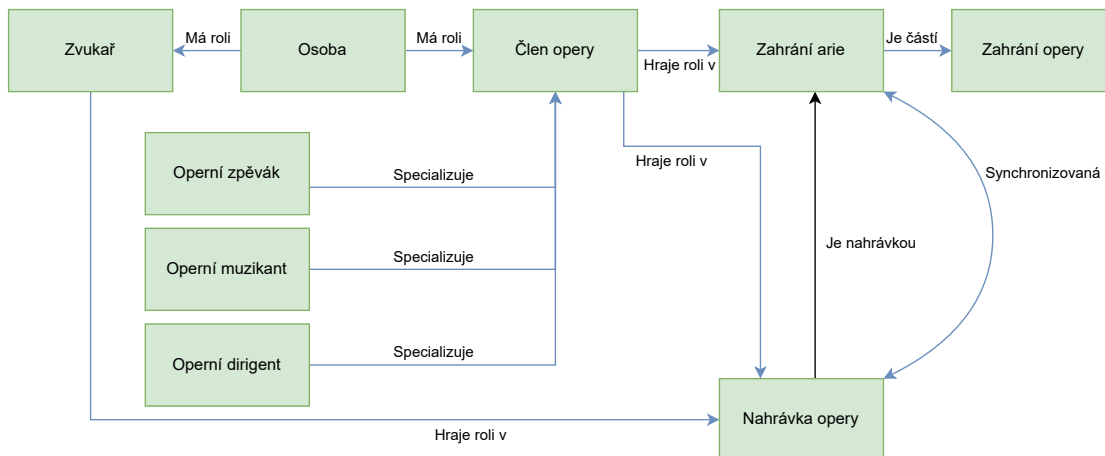
Popisují vybranou problematiku ve velmi obecné abstraktní rovině. Pro popis problematiky se používají nekonkrétní pojmy (např. *objekt*) a vztahy (např. *vlastní*) (viz obrázek 2.1). Zachycují obecné problémy. Využívají se pro mapování doménových ontologií. [7, 8, překlad vlastní]

### 2.2.2 Doménové ontologie

Charakterizují problematiku určité domény. K popisu používá určité pojmy z dané domény. [10, 11, překlad vlastní] Doménu popisuje do dostatečného detailu abstrakce reality relativního k vybranému účelu (viz obrázek 2.2).

#### 2.2.2.1 Klíčové ontologie

„Klíčové ontologie poskytují precizní definici strukturálních vědomostí v určitém oboru, který se rozpíná přes několik aplikačních domén.“ [12, překlad vlastní]



■ Obrázek 2.2 Příklad doménové ontologie [9, překlad vlastní]

## 2.3 Využití ontologie v praxi

V současnosti se více rozšiřuje použití ontologie v informatice. Významnými obory, kde se tento fenomén objevuje, jsou zdravotnictví a ekonomika.

V odvětví ekonomiky se pomocí ontologie řeší například problematika identity peněz [13, překlad vlastní], peněžních transakcí či komunikace mezi subjekty. [14, 15, překlad vlastní]

Ve zdravotnictví je možné se setkat s využitím ontologie pro vylepšení konceptuálního schématu lidského genomu [16, překlad vlastní] nebo, vzhledem na stav pandemie SARS-CoV-2 ve světě [17, překlad vlastní], s využitím pro lepší zjišťování stavu pacientů a dalších složek zdravotnictví.

## 2.4 Unified Foundational Ontology

UFO je jednou z několika existujících vyšších ontologií (sekce 2.2.1) (např. *DOLCE*, *GFO*). Byla vytvořena Giancarlem Guizzardim a dalšími před dvaceti lety a je stále používána a upravována dle potřeb vzhledem k evoluci konceptuálního modelování. UFO byla vytvořena spojením dvou vyšších ontologií *DOLCE* a *DFO*. Důvod k tomuto spojení byla snaha o zhotovení referenční vyšší ontologie pro potřeby konceptuálního modelování. [18, překlad vlastní]

### 2.4.1 Mikroteorie

Z důvodu vytvoření UFO pro účely konceptuálního modelování bylo nutné jeho principy zachytit a vydefinovat. Za tímto účelem bylo zformulováno několik mikroteorií, které na principy odkazují. [18, překlad vlastní]

- „teorie typů a taxonomických struktur, které jsou propojeny s teorií identifikátorů objektů, včetně formální sémantiky v sortálně kvantifikované modální logice“,
- „teorie relací částí a celků“,
- „teorie konkrétních vnitřních vlastností, atributů a jejich hodnot, do kterých patří pohledy na databáze jako sémantické referenční struktury“,
- „teorie konkrétních relačních vlastností a relací včetně návrhu na slabé pravdomlvnosti propojující konkrétní vlastnosti s tvrzeními“,

- „teorie rolí“,
- „teorie událostí včetně aspektů jako jsou například mereologie událostí, dočasné uspořádání událostí, účastnění se objektů na událostech, příčinná souvislost, změna a propojení mezi událostmi a enduranty přes dispozice“,
- „teorie víceúrovňového modelování“. [18, překlad vlastní]

## 2.4.2 Kategorie UFO

### 2.4.2.1 UFO-A

„Ontologie endurantů jednajících s aspekty strukturálního konceptuálního modelování. Je organizována do čtyř kategorií, zahrnující teorie typů a taxonimických struktur napojených na teorii objektových identifikátorů, relací částí a celků, konkrétních vnitřních vlastností, atributů a jejich hodnot, konkrétních relačních vlastností a relací a rolí.“ [11, překlad vlastní]

### 2.4.2.2 UFO-B

„Ontologie perdurantů jednajících s aspekty jako jsou perdurantská morologie, dočasné uspořádání perdurantů, účastnění se objektů na perdurantech, příčinná souvislost, změna a propojení mezi událostmi a enduranty přes dispozice.“ [11, překlad vlastní]

### 2.4.2.3 UFO-C

„Ontologie sociálních a záměrných entit, založená na podkladech UFO-A a UFO-B, popisuje představy jako jsou víra, touhy, úmysly, cíle, činnosti, závazky a tvrzení, sociální role a konkrétní sociální relační komplexy a další.“ [11, překlad vlastní]

## Kapitola 3

# UML

Unified Modeling Language je jedním z modelovacích jazyků. Využívá se pro tvorbu a dokumentaci softwarových systémů včetně jejich analýzy, návrhu a implementace. Zahrnuje i podporu modelování byznys vrstvy, například pomocí diagramů aktivit. [19, překlad vlastní] Modely se skládají z elementů, které se rozdělují do tří kategorií:

### 1. Klasifikátory

Je uskupení objektů, které mají své stavy, založené na hodnotách jejich vlastností, a relace na jiné objekty. [19, překlad vlastní]

### 2. Události

Je seskupení různých výskytů, které mohou ovlivnit popisovaný systém. Dopad výskytu na systém není rozdělován na pozitivní či negativní. Jedná se spíše například o časové intervaly, které po uplynutí spouští vybranou funkcionalitu systému. [19, překlad vlastní]

### 3. Chování

Je skupina provedení, následujících na událost a odpovídajících na něj, s možností úpravy stavu objektů. [19, překlad vlastní]

## 3.1 Sémantické kategorie

Modely se dále rozdělují podle následujících sémantických kategorií:

### 1. Strukturální sémantika

Definuje elementy strukturálních modelů, které odpovídají *jednotlivcům* v určité fázi jejich cyklu života. [19, překlad vlastní]

### 2. Sémantika chování

Definuje elementy modelů chování, jež vypovídají o změnách *jednotlivců* v průběhu jejich existence. [19, překlad vlastní]

## 3.2 UML diagram tříd

Diagram tříd je jedním ze *strukturálních* diagramů obsažených v UML. Popisuje objekty, jejich vlastnosti, metody a relace v rámci objektově orientovaných systémů. Na UML diagramu tříd je postaven ontologický diagram v OntoUML.





## Kapitola 4

# OntoUML

OntoUML je ontologický modelovací jazyk, založený na *vyšší ontologii* UFO (sekce 2.4). OntoUML, stejně jako UFO, pochází z dílny Giancarla Guizzardiho, který je významným průkopníkem ontologické analýzy.

### 4.1 Třídy

Představují definice objektů. Obsahují vlastnosti, které budou mít jednotlivé instance, jenž jsou na základě třídy vytvořeny. Instance definují obsah vlastností třídy. Existují *třídní atributy*, které definují obsah pro všechny instance dané třídy.

Třídy se rozdělují na dva typy: *konkrétní* a *abstraktní*. Rozdíl mezi nimi je ve vytváření instancí. Abstraktní třídy nemohou mít vlastní instance. Používají se pro definici kategorických vlastností, např. nemovitost. [20]

Na rozdíl od UML modelu tříd neobsahují třídy definici metod. [20] Metody jsou funkce, které dle vstupních parametrů mění obsah třídy, ve které jsou definovány. Jelikož je OntoUML zaměřeno na *strukturální modelování*, tak pro ně v modelech není místo, protože vykazují vlastnosti *behaviorální sémantiky* (sekce 3.1).

### 4.2 Princip Identity

Jedná se o jednoznačné určení jednotlivce v rámci celého jeho životního cyklu nehladě na změny, kterými si projde. [20] Objekty se dělí na dvě kategorie podle principu identity:

#### 4.2.1 Sortály

Mají vlastní *identitu* a slouží jako kategorizace pro třídy, které z nich dědí. Nejběžnějším případem je osoba. Každou instanci osoby lze jednoznačně určit dle její identity a to i v případě, že se jedná o dvojčata či víceročata. Každý sortál má **pouze jednu** identitu. Sortály se rozdělují na dva podtypy na základě *rigidity*, ta popisuje neměnnost objektu v průběhu života. [20]

##### 4.2.1.1 Rigidní sortály

► **Definice 4.1** (Rigidní typ [20]). *Typ  $T$  je rigidní pro každou instanci  $x$ , právě tehdy, když je nutně  $\square$  instancí  $T$ .*

*Pokud je tedy  $x$  instancí  $T$  v nějakém světě  $w$ , tak musí být  $x$  instancí  $T$  v každém možném světě  $w'$ .*

$$R^+(T) =_{def} \Box(\forall x: T(x) \implies \Box(T(x)))$$

- $\Box$  – ve všech světech platí
- $T(x)$  –  $x$  je instancí  $T$

Podskupinou rigidních sortálů jsou takzvané *Ultimátní sortály*. To je skupina sortálů, které jsou schopni poskytnout *identitu* při dědění (popis relace dědění, sekce 4.3.1.2). [20] Do této skupiny patří:

- «Kind»  
Základní entita představující objekt, poskytující identitu. [20]
- «Relator»  
Objekt sloužící jako pečetidlo materiální relace (sekce 4.3.1.1) mezi dvěma entitami. [21]
- «Quality»  
Popisuje *měřitelnou* vlastnost. Patří mezi **aspekty**<sup>1</sup>. [22]
- «Mode»  
Popisují neměřitelné vlastnosti. Patří mezi **aspekty**<sup>1</sup>. [22]
- «Quantity», «Collective»  
Stereotypy využívané v teorii *celku a části* (sekce 4.4).

Jediným rigidním sortálem nepatřícím do této skupiny je stereotyp «SubKind». Od *ultimátních sortálů* se liší schopností poskytnout **vlastní** identitu. «SubKind» může získat identitu pouze děděním od některého z prvků ze skupiny *ultimátních sortálů*. [20]

#### 4.2.1.2 Anti-rigidní sortály

► **Definice 4.2** (Anti-rigidní typ [20]). *Typ  $T$  je **anti-rigidní** pro každou instanci  $x$ , právě tehdy, když je **možné**  $\Diamond$ , že  $x$  nemusí být instancí  $T$ .*

*Pokud je tedy  $x$  instancí  $T$  v nějakém světě  $w$ , tak je možné, že existuje nějaký svět  $w'$ , kde  $x$  **není** instancí  $T$ .*

$$R^-(T) =_{def} \Box(\forall x: T(x) \implies \Diamond(\neg T(x)))$$

- $\Box$  – ve všech světech platí
- $\Diamond$  – v alespoň jednom světě platí
- $T(x)$  –  $x$  je instancí  $T$

Mezi anti-rigidní sortály patří stereotypy «Role» a «Phase». Vyznačují se svojí závislostí na vlastnostech svých *předků*. [20]

- «Phase»  
Jsou závislé na vnitřní vlastnosti *předka*. Všechny generalizující relace mezi fázemi a *předkem* musí mít přidělené klíčové vlastnosti *{disjoint, complete}*. [20]
- «Role»  
Jsou závislé na vnější podmínce, tzn. jsou **relačně závislé**. [20]

<sup>1</sup>Entita existenčně závislá na objektu, respektive aspektu (**nositeli**); při zániku nositele zaniká i aspekt

► **Definice 4.3** (Relační závislost [20]). *Typ  $T$  je relačně závislý na typu  $P$  skrze relaci  $R$ , právě tehdy, když pro každou instanci  $x$  typu  $T$  existuje instance  $y$  typu  $P$  taková, že  $x$  a  $y$  jsou v relaci  $R$ .*

$$D^+(T, P, R) =_{def} \Box(\forall x : T(x) \implies \exists y : P(y) \wedge R(x, y))$$

- $\Box$  – ve všech světech platí
- $\Diamond$  – v alespoň jednom světě platí
- $T(x)$  –  $x$  je instancí  $T$
- $P(y)$  –  $y$  je instancí  $P$
- $R(x, y)$  –  $x$  je v relaci s  $y$

### 4.2.1.3 Non-sortály

Nemají vlastní identitu. Z názvu a vlastností jedince nelze jednoznačně určit, o který objekt se jedná. Typickým příkladem jsou nemovitosti, kdy nelze jednoznačně říct, jestli se jedná o pozemek, dům, či garáž. [21]

Svým *potomkům* poskytují své (kategorické) vlastnosti. Tím, že neposkytují princip identity, nezamezují svým *potomkům* mít odlišné identity. [21] Mezi non-sortály patří tyto stereotypy:

- «*Category*»

Reprezentuje *rigidní non-sortály*, kvůli této skutečnosti může mít za *potomky* pouze *rigidní sortály*<sup>2</sup> nebo jinou «*Category*». Celá podmnožina *potomků* **musí** být *disjunktní*. [21]

- «*RoleMixin*»

Je jedním z *anti-rigidních non-sortálů*. *Potomci* jsou omezeni na stereotyp «*Role*», a tudíž je stejně jako «*Role*» existenčně závislý. [21]

- «*PhaseMixin*»

Je druhým z *anti-rigidních non-sortálů*. *Potomci* jsou omezeni na stereotyp «*Phase*». [21]

- «*Mixin*»

Je zástupcem *semi-rigidních non-sortálů*. [21]

► **Definice 4.4** (Semi-rigidní typ [21]).  $R^\sim(T) =_{def} \neg R^+(T) \wedge \neg R^-(T)$

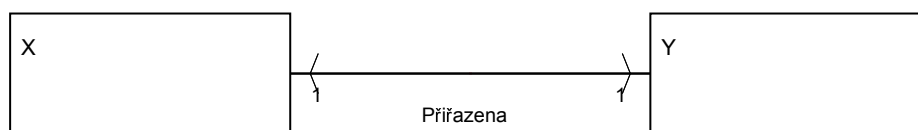
- $R^+(T)$  –  $T$  je *rigidní typ*
- $R^-(T)$  –  $T$  je *anti-rigidní typ*

«*Mixin*» **musí** mít za *potomky* jak jeden *rigidní typ*, tak jeden *anti-rigidní typ*. [21]

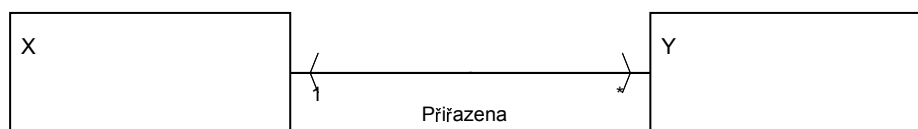
## 4.3 Relace

Relace jsou vztahy mezi dvěma třídami. Počet relací mezi vybranými dvěma třídami je omezen na jednu, avšak počet relací, které má jedna třída, není omezen. Všem relacím, vyjma *generalizace*, definujeme *násobnost* a *název*. Pro zlepšení čitelnosti může být definován i *směr* (čtení) relace. [21]

Násobnost je definována na obou koncích relace. Vypovídá o počtu instancí svázaných s entitou. Násobnost má tyto definované tvary  $0..1$ ,  $1..1$ ,  $0..*$ ,  $1..*$ , přičemž  $*$  může být nahrazena za jakékoliv přirozené číslo. [21] Z těchto tvarů násobností vyplývají tři typy relací:



■ **Obrázek 4.1** Ukázka vazby 1:1



■ **Obrázek 4.2** Ukázka vazby 1:N

### 1. 1:1

Na jednu instanci třídy  $X$  připadá pouze jedna instance třídy  $Y$  (viz obrázek 4.1).

### 2. 1:N

Instance třídy  $X$  je v relaci s až  $N$  instancemi třídy  $Y$  (viz obrázek 4.2).

### 3. M:N

Instance třídy  $X$  je v relaci s až  $N$  instancemi třídy  $Y$  a instance třídy  $Y$  je v relaci s až  $M$  instancemi třídy  $X$  (viz obrázek 4.3).

## 4.3.1 Typy relací

Základní typy relací jsou opět odvozeny z notace UML, avšak stejně jako ve třídách, tak i u relací nestačí pouhá UML notace pro potřeby ontologického modelování. Z tohoto důvodu byly definovány stereotypy relací. *Asociace* získala čtyři vlastní stereotypy (*formální*, *«Material»*, *«Mediation»*, *«Characterization»*). [21] Dále přibýly stereotypy pro popis vztahu *celku a části* (sekce 4.4.3).

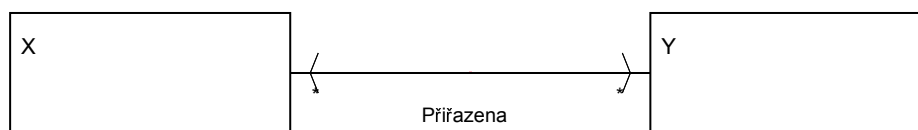
### 4.3.1.1 Asociace

Propojuje dvě entity jednoduchým vztahem. Jedná se o nejjednodušší typ relace. Musí splňovat všechny definované požadavky (sekce 4.3).

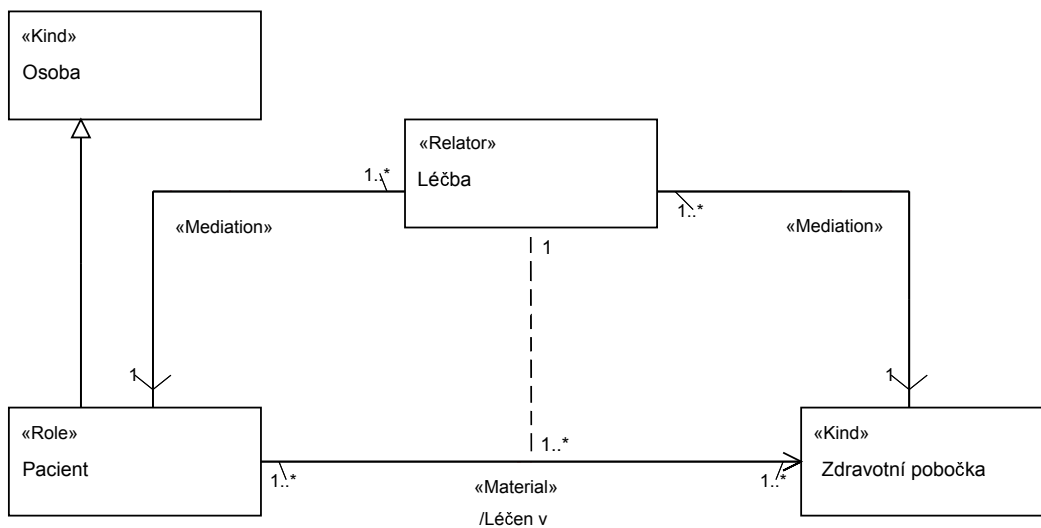
#### 1. Formální relace

Pro svoji funkci nepotřebují žádné pečetící individuuum. Váže se na vnitřní vlastnosti entit. Relace značíme jako *odvozené*, tzn. jejich výsledek lze dopočítat z ostatních objektů modelu. [21]

<sup>2</sup> «Kind», «SubKind», «Relator», «Mode», «Quality»



■ **Obrázek 4.3** Ukázka vazby M:N



■ **Obrázek 4.4** Ukázka dekomponované materiální relace [21]

## 2. Materiální relace

Pro svoji funkci potřebují pečetidlo (**truthmaker**) (popis objektu *«Relator»*, sekce 4.2.1.1). Pečetidlo slouží jako potvrzení existence relace, proto když zanikne instance pečetidla, tak se přeruší relace mezi subjekty, které byly pečetidlem spojeny. Pečetidlo lze chápat jako smlouvu, může být jak hmotná, tak i imaginární (společenská úmluva). Materiální relace jsou označovány stereotypem *«material»* a jsou značeny jako *odvozené*, jelikož jsou odvozené od upřesňující relace s *relatorem*. [21]

Materiální relace hrají důležitou roli v relacích s násobností **M:N**. Slouží k dekomponování vztahu na dva vztahy, jeden vztah **1:N** a jeden vztah **M:1**. Tyto dvě relace se označují stereotypem *«mediation»* (viz obrázek 4.4). [21]

## 3. Charakterizace

Slouží k propojení **nositele** s **aspektem** (*«Quality»*, *«Mode»*).

### 4.3.1.2 Generalizace

Třída *předek* definuje vlastnosti všech svých *potomků*. Potomci se od sebe mohou lišit, např. dalšími novými soukromými vlastnostmi. Z matematického hlediska je generalizace antisymetrická, tranzitivní relace. [21] Vazbu generalizace lze dále upřesnit pomocí kombinace sémantických meta-atributů (viz obrázek 4.5):

- *complete*

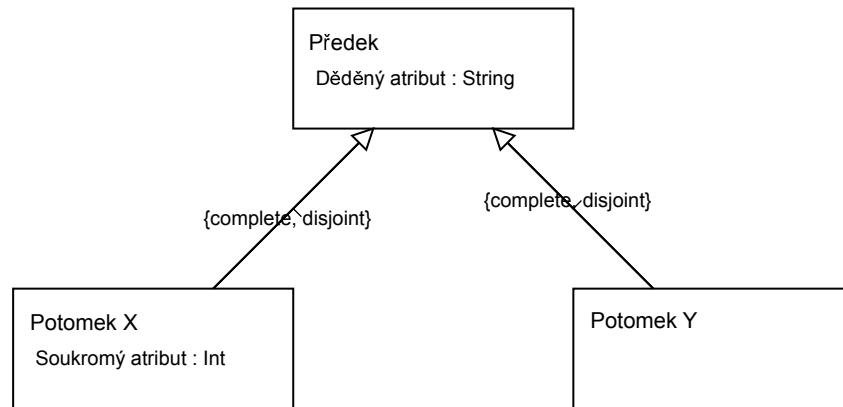
Instance *předka* je nutně instancí alespoň jednoho *potomka*. [21]

- *disjoint*

Instance je instancí maximálně jednoho *potomka*. [21]

## 4.4 Teorie celku a části

Je zásadní pro objektové programování. Pro ontologické modelování jsou podstatné z důvodu podpory kognitivního vnímání. [22]



■ **Obrázek 4.5** Ukázka generalizace

### 4.4.1 Definice z UML

Teorie celku a části vychází z notace UML. Avšak ta je pro potřeby ontologického konceptuálního modelování nedostatečná. V UML je teorie založena na pouze dvou druhích relací.

► **Definice 4.5** (Agregace). *Entita  $X$  je nedílnou součástí entity  $Y$ , tzn. instance třídy  $X$  nemůže existovat, pokud již neexistuje instance třídy  $Y$ , která je s ní ve vztahu.*

► **Definice 4.6** (Kompozice). *Instance třídy  $X$  může existovat nezávisle na existenci instance třídy  $Y$  k ní vztahené.*

Z hlediska OntoUML se definice *Agregace* a *Kompozice* mění. Oba stereotypy v OntoUML fungují na bázi *možnosti sdílení* části, dle použité relace se mění hodnota vlastnosti *isShareable*. [22]

Při použití *agregace* je nastavena proměnná *isShareable* na *false*. Část nemá možnost sdílení, což znamená, že multiplicita relace u části je **maximálně 1**. [22]

Při použití *kompozice* je nastavena proměnná *isShareable* na *true*. Část má možnost sdílení, multiplicita relace u části je tedy **minimálně 1**. [22]

### 4.4.2 Závislosti mezi celkem a částí

Pro potřeby ontologického modelování byly definovány dva druhy závislostí mezi celkem a částí.

► **Definice 4.7** (Generická závislost [22]). *Instance  $y$  je genericky závislá (generically dependent) na typu  $T$ , pokud pro existenci  $y$  je **nutné**, aby existovala i instance  $T$ .*

$$GD(y, T) =_{def} \Box(\varepsilon(y) \implies \exists x : T(x) \wedge \varepsilon(x))$$

- $\varepsilon(x)$  – unární predikát „existovat v doméně“
- $\Box$  – ve všech světech platí
- $T(x)$  –  $x$  je instancí  $T$

► **Definice 4.8** (Existenční závislost [22]). *Instance  $x$  je existenčně závislá (existentially dependent) na jiné instanci  $y$ , pokud existence instance  $x$  vyžaduje existenci určité specifické instance  $y$ .*

$$ed(x, y) =_{def} \Box(\varepsilon(x) \implies \varepsilon(y))$$

- $\varepsilon(x)$  – unární predikát „existovat v doméně“

- $\square$  – ve všech světech platí

Z důvodu nedostatečné definice, nacházející se v UML, pro ontologické modelování, je OntoUML rozšířeno o vlastnosti relací, závisejících na rigiditě entit, se kterými je relace spojena.

#### 4.4.2.1 Obecné

- Povinný celek

► **Definice 4.9** (Povinný celek [22]). *Instance  $y$  je povinný Celek (mandatory whole) pro jinou instanci  $x$ , pokud  $x$  je genericky závislá na typu  $T$ , který  $y$  instancuje, a instance  $x$  je nezbytně Částí instance typu  $T$ .*

$$MW(T, x) =_{def} \square(\varepsilon(x) \implies (\exists y : T(y) \wedge (x \triangleleft y)))$$

- $\varepsilon(y)$  – unární predikát „existovat v doméně“
- $x \triangleleft y$  – binární predikát „ $x$  je Částí  $y$ “
- $\square$  – ve všech světech platí
- $T(x)$  –  $x$  je instancí  $T$

- Nepovinný celek

Část může existovat nezávisle na existenci celku. [22]

- Povinná část

Část, která nemůže být celkem postradatelná, aniž by ovlivnila jeho *klasifikaci*, či *identitu*. [22]

► **Definice 4.10** (Povinná část [22]). *Instance  $x$  je povinná Část (mandatory part) jiné instance  $y$ , pokud  $y$  je genericky závislá na typu  $T$ , jež  $x$  instancuje a instance  $x$  je nezbytně Částí instance  $y$ .*

$$MP(T, y) =_{def} \square(\varepsilon(y) \implies \exists x : T(x) \wedge (x \triangleleft y))$$

- $\varepsilon(y)$  – unární predikát „existovat v doméně“
- $x \triangleleft y$  – binární predikát „ $x$  je Částí  $y$ “
- $\square$  – ve všech světech platí
- $T(x)$  –  $x$  je instancí  $T$

- Volitelná část

Část, která může být celkem postradatelná bez změny na jeho *klasifikaci*, či *identitě*. [22]

#### 4.4.2.2 Rigidní

- Esenciální část

Značí se klíčovým slovem *essential*.

► **Definice 4.11** (Esenciální část [22]). *Instance  $x$  je esenciální Část (essential part) jiné instance  $y$ , pokud  $y$  je existenčně závislá na instanci  $x$  a instance  $x$  je nezbytně Částí instance  $y$ .*

$$EP(x, y) =_{def} \square(\varepsilon(y) \implies x \triangleleft y)$$

- $\varepsilon(y)$  – unární predikát „existovat v doméně“
- $x \triangleleft y$  – binární predikát „ $x$  je Částí  $y$ “

- $\square$  – ve všech světech platí

#### ■ Neoddělitelná část

Značí se klíčovým slovem *inseparable*.

► **Definice 4.12** (Neoddělitelná část [22]). *Instance  $x$  je neoddělitelná Část (inseparable part) jiné instance  $y$ , pokud  $x$  je existenčně závislá na instanci  $y$  a instance  $y$  je nezbytně Částí instance  $x$ .*

$$IP(x, y) =_{def} \square(\varepsilon(x) \implies (x \triangleleft y))$$

- $\varepsilon(y)$  – unární predikát „existovat v doméně“
- $x \triangleleft y$  – binární predikát „ $x$  je Částí  $y$ “
- $\square$  – ve všech světech platí

### 4.4.2.3 Anti-rigidní

#### ■ Neměnitelná část

Značí se klíčovým slovem *immutablePart*.

Funkcionalita je analogická k esenciální části, z důvodu odlišné rigidity je potřeba mezi pojmy rozlišovat, jinak by byla porušena definice. [22]

#### ■ Neměnitelný celek

Značí se klíčovým slovem *immutableWhole*.

Funkcionalita je analogická k neoddělitelné části, z důvodu odlišné rigidity je potřeba mezi pojmy rozlišovat, jinak by byla porušena definice. [22]

## 4.4.3 Stereotypy celků a částí

### 4.4.3.1 «Quantity»

Je **maximální**<sup>3</sup> **spojitý**<sup>4</sup> objekt složený z částí stejného typu, jako je on sám, což je příčinou, že je **topologicky vymezen**. Tento kontinuální kompaktní prostor, který okupuje, může být hmotný i nehmotný (čas). [22] K entitě typu «Quantity» se vážou dvě relace, podporující její funkcionalitu:

#### 1. «Containment»

Propojuje *kvantitu* s objektem, jehož tvaru *kvantita* nabývá. [22]

#### 2. «SubQuantityOf»

Propojuje mezi sebou dvě *kvantity* a vyjadřuje, že jedna z *kvantit* se skládá z druhé. Jelikož je *kvantita* **maximální** objekt, musí být na obou stranách relace *multiplicita* **právě jedna**. Jelikož je pro *kvantitu*, ze které vede vazba «SubQuantityOf», subkvantita povinná, musí být relaci definována klíčová vlastnost *essential*. Relace je **anti-symetrická**, **tranzitivní** a **irreflexivní**. [22]

<sup>3</sup>Není pododdílem žádného dalšího objektu stejného typu

<sup>4</sup>Části jsou mezi sebou spojeny v prostoru



#### 4.4.3.2 «Collective»

Je složen z částí odlišného typu než je sám *kolektiv*, avšak všechny jeho části plní stejnou roli. *Kolektiv* je **maximální**<sup>3</sup> a **konečně dělitelný**, tzn. obsahuje **atomickou** entitu s identitou. [22] Pro práci s *kolektivem* byly definovány tyto relace:

1. «*MemberOf*»

Je relace mezi *kolektivem* a *atomem*. Na straně *atomu* je multiplicita **minimálně 2**, jelikož jeden prvek není sám o sobě kolektiv. Relace je **irreflexivní** a **anti-symetrická**. [22]

2. «*SubCollectionOf*»

Je relace mezi dvěma *kolektiv*y. Vazba je **tranzitivní**, **irreflexivní** a **anti-symetrická**. Jelikož jsou subkolektivy zaměřením určitých kolektivů, je nutné definovat relaci vlastnost *in-seperable*. [22]

#### 4.4.3.3 Funkční celek

Nejedná se o stereotyp, ale o vlastnost entity. Skládá se z prvků, které mají různé role v rámci *funkčního celku*. [22] Pro svou funkčnost potřebuje jednu relaci:

1. «*ComponentOf*»

Propojuje *funkční celek* se svou částí. Relace je **irreflexivní**, **anti-symetrická** a v závislosti na kontextu může být **tranzitivní**. [22]



## Sémantická interoperabilita

Řešení problému záměny významu slov v jazyce. Záměna může probíhat dvěma způsoby. První je záměna rozdílných slov stejného významu (např. adresát a příjemce). Druhý je záměna stejných slov jiného významu (např. fotbalová branka a vstupní branka). Tyto záměny se poté projevují v sémantických modelech, kde se projevují kolizemi mezi objekty nebo zbytečným nabobtnáním počtu entit.

Toto dále může vyústit v problémy s databázemi, které byly navrženy z těchto chybných modelů. Duplikace entit s vnitřně stejným významem zbytečně zkomplikuje vazby mezi dalšími entitami a tím se výrazně zkomplikuje mapování dat na dané entity. Může také dojít k duplikaci dat, které prodraží provoz z důvodu nadměrného využívání kapacitních zdrojů.

K správnému použití sémantické interoperability mohou dopomoci různé standardy (např. AACR, Dublin Core). [23]

### 5.1 Souvislost s datovými sadami

Datové sady poskytnuté společností Remmark, a.s. obsahují různé statistiky provedené v České republice. Jelikož statistiky nejsou standardizované, tak se velmi často objevují termíny popisující jednu a tu samou věc napříč více datovými sadami. To znamená, že kolize mezi entitami by byly nevyhnutelné, pokud bychom se nesnažili o sémantickou interoperabilitu v jednotlivých modelech. Toto je však velice pracné kvůli expresivitě modelů v OntoUML.

### 5.2 Sémantická interoperabilita v OntoUML

Jelikož OntoUML definuje ve svých modelech povinné stereotypy entit, tak se výrazně komplikuje vytváření modelů. Aby byla zachována sémantická interoperabilita mezi všemi modely, je nutné sjednotit název i stereotyp entity. Se stereotypy je ale problém, že se do modelu vůbec nemusí hodit. Pak je v krajním případě nutné předělat všechny existující modely s danou entitou, aby odpovídaly modelu s entitou s novým stereotypem.

### 5.3 Sémantická interoperabilita a datový inkubátor

Datový inkubátor je napojen na jeden velký sémantický model. Tento model se skládá z menších modelů, které jsou spolu provázány při nahrání na datovou platformu společnosti Remmark, a.s. Z důvodu sjednocování modelů je nutné dodržet sémantickou interoperabilitu skrze všechny modely. Pokud by nebyla dodržena, tak na tento problém datová platforma upozorní.

Kdyby se v modelech neřešila sémantická interoperabilita, byl by projekt datového inkubátoru téměř nereálný. Ve statistikách obsažených v datových sadách se nachází takové množství kolizí mezi termíny, že by vytvoření jednotlivých modelů zabralo enormní množství času. Jejich spojení by bylo prakticky nemožné. A i pokud by se podařilo vytvořit finální model pro datový inkubátor, tak by byl finančně velmi náročný na provoz. Ať už z hlediska nadměrného využívání úložného prostoru z důvodu duplikace dat či času stráveného nad změnami nebo rozšířeními datového inkubátoru.

## 5.4 FAIR

Z důvodu rychle narůstajícího objemu dat je prakticky nutné data zpracovávat počítačově, což není vždy jednoduchý úkol. Jednou z reakcí na tuto skutečnost bylo definování FAIR principů. Akronym FAIR znamená *Findable, Accessible, Interoperable, Reusable*. [24, překlad vlastní]

### 1. Findable

Je obohacení dat o jednoznačný identifikátor a metadata, díky kterým je jednoduché data vyhledat jak pro stroje, tak pro lidi. [24, překlad vlastní]

### 2. Accessible

Určuje standardizaci přístupu k datům. [24, překlad vlastní]

### 3. Interoperable

Definuje požadavky na integraci s dalšími daty. *Sémantická interoperabilita* (více v sekci 5) je jedním z požadavků FAIR interoperability dat. [24, překlad vlastní]

### 4. Reusable

Metadata jsou navržena tak, aby byla znovupoužitelnost dat co nejjednodušší. Jako znovupoužitelnost se počítá i kombinace a replikace dat. [24, překlad vlastní]

### 5.4.1 FAIRifikační proces

Popis procesu FAIRifikace dat (viz obrázek 5.1).

#### 1. Získání neFAIRifikovaných dat

Spočívá v získání přístupu k datům. V rámci FAIRifikačního procesu se nebere v potaz, jestli jsou data otevřená, nebo soukromá. Z tohoto důvodu je součástí procesu i definice licence. [25, překlad vlastní]

#### 2. Analyzování získaných dat

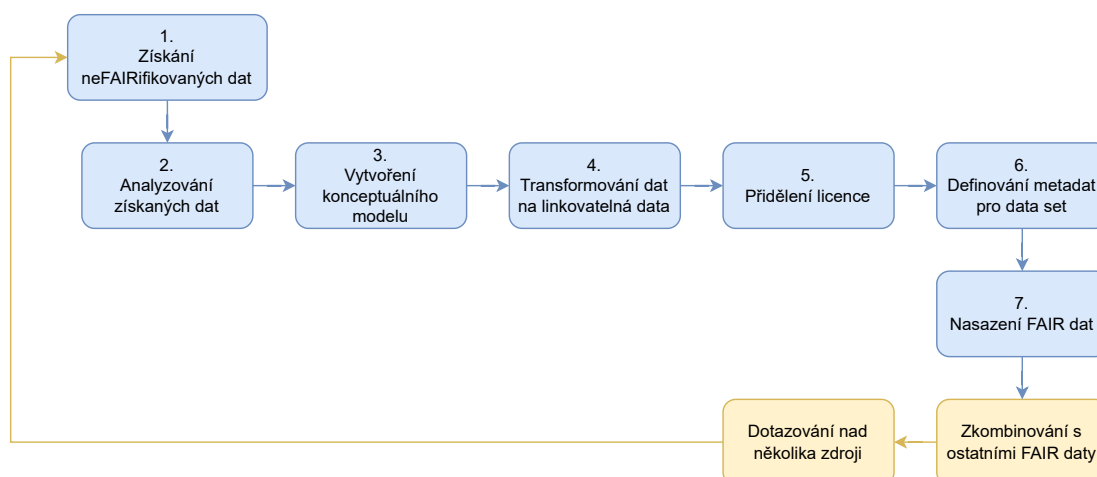
Nad daty se provede (ontologická) analýza, která spočívá v zjištění typu obsahu datové sady, definování struktury dat, vlastností a propojení dat mezi sebou. [25, překlad vlastní]

#### 3. Vytvoření konceptuálního modelu

Jedná se o vytvoření sémantického konceptuálního modelu, popisujícího vlastnosti a relace entit, které vyplynuly z analýzy obdržených dat. [25, překlad vlastní]

#### 4. Transformování dat na linkovatelná data

Pomocí vzniklého sémantického modelu se jeho aplikováním vytvoří linkovaná data. Tento krok je možné přeskočit, pokud z analýzy dat vyplyne, že se to nevyplatí, např. audiovizuální data. [25, překlad vlastní]



■ **Obrázek 5.1** Proces FAIRifikace [25, překlad vlastní]

#### 5. Přidělení licence

Definuje použití nové FAIR datové sady. Licenci je důležité definovat, kvůli možnosti komerčního využití dat. [25, překlad vlastní]

#### 6. Definování metadat pro data set

Jedná se o přidání metadat k datasetu pro zlepšení hledání datových sad, možnost rekonstrukce dat. Dále zlepšují čitelnost, když jsou v metadatech obsaženy popisy obsahu sad. [25, překlad vlastní]

#### 7. Nasazení FAIR dat

Pojednává o vydání licencovaných FAIRifikovaných dat. Data mohou být vydána soukromě, či veřejně pod otevřenou, nebo uzavřenou licenci. [25, překlad vlastní]



## Kapitola 6

# OpeNest

### 6.1 Projekt OpeNest

OpeNest je projekt společnosti Remmark, a.s., což je full-service agentura, která se na trhu pohybuje od roku 1998.

OpeNest navazuje na projekt NEST BIG DATA ARENA, který byl dotován v roce 2020 z Evropských strukturálních a investičních fondů. Výsledkem bylo vytvoření datové platformy pro shromažďování marketingových dat. [26]

#### 6.1.1 Datová platforma

Datová platforma slouží jako centrum pro datové sady. Umožňuje přístup k mnoha datovým sadám z oblasti marketingu a jejich propojení za pomoci dotazovacího jazyka, který dokáže propojit dvě různé entity na základě souvislostí nalezených v ontologických modelech jednotlivých datových sad obsažených v platformě.

Další důležitou funkcionalitou je možnost přidat vlastní již zpracovanou datovou sadu. To ulehčuje přechod na platformu pro společnosti, které již mají vlastní uložené zdroje a nechtějí je mít rozprostřené přes několik úložišť.

Datová platforma je tak určena jak pro nové uživatele datových sad, tak i pro ty již zaběhlé v oboru. Obzvláště pro společnosti přicházející do oblasti sběru a zpracování dat je datová platforma zakomponována do balíčků služeb pro pomoc při rozvoji firmy.

#### 6.1.2 Inkubátor

Projekt OpeNest není jen o poskytnutí datové platformy na trh. Datová platforma je jenom částí velkého celku, který symbolizuje inkubátor. V rámci inkubátoru bude Remmark, a.s., jako full-service agentura, poskytovat služby pro rozvoj firem. Jedná se o pomoc s marketingovou kampaní, poskytnutí kancelářských prostorů i expertů z oborů marketingu, PR a dalších oborů v oblasti komunikace se zákazníkem. [26]

Jednou z hlavních částí marketingu je šíření takzvané osvěty v odvětví poskytování datových sad. Cílem osvěty je přimět společnosti k dodržování otevřených standardů pro data. Pokud se tento cíl splní, tak by se výrazně urychlilo rozšíření obsahu datové platformy, jelikož by se zefektivnilo zpracování nových standardizovaných datových sad a jejich následné zakomponování do platformy.

## 6.2 Otevřené standardy pro data

Otevřené standardy pro data jsou dohody, které zajišťují, aby data byla kvalitní a lépe se s nimi pracovalo. Normy jsou veřejné a zdokumentované, aby se dosáhlo co největšího pokrytí a znovupoužitelnosti. Otevřených standardů je mnoho, v rámci tisíců, kdokoli si může vytvořit svůj vlastní nebo se může podílet na vývoji již existujícího. Jelikož jsou standardy veřejné pro všechny, tak využívat či měnit normy může jakákoliv právnická, či fyzická osoba. Díky této skutečnosti je výhodné standardy využívat, protože se tím přesouvá práce se zpracováním dat na vlastní společnost a ne na odběratelé, čímž se znatelně zvýší množství potenciálních uživatelů. [27, překlad vlastní]

Nevýhodou využívání standardů je právě navýšení množství práce nutné k poskytnutí dat a tím množství spotřebovaných zdrojů. Avšak pokud se zveřejní nějaký obsah, je záměrem vydavatele, aby se co nejvíce využíval. K tomu ony normy dopomáhají, jelikož se odstraní nutnost odběratele vytvářet nové transformace pro data. Transformace přetváří strukturu dat do podoby vhodné pro účely uživatele. Pokud uživatel již má existující transformace pro daný standard, nemusí tak vynakládat zdroje navíc pro zpracování datových sad, čímž se navyšuje její přitažlivost využití.

Použití otevřeného standardu ale neznamená, že datové sady, které ho využívají, musí nutně patřit mezi otevřená data. Otevřená data jsou pouze jednou ze tří kategorií datového spektra definovanými ODI. [28, překlad vlastní] Normy obecně neomezují pro jakou kategorii dat jsou používány. Díky tomu jsou díky otevřeným standardům zlepšovány data ve všech sektorech trhu od soukromých až po veřejné. [27, překlad vlastní]

### 1. Otevřená data

- Dostupná bez vyšší ceny než je cena reprodukce, bez limitací na identitu uživatele či jeho záměr.
- V digitálním, počítačově zpracovatelném formátu pro interoperabilitu s jinými daty.
- Bez restrikcí na užití nebo redistribuci v jejich licenčních podmínkách. [29, 30, překlad vlastní]

### 2. Sdílená data

- Named access - „Data jsou sdílena jen s jmenovanými lidmi a organizacemi.“
- Group-based access - „Data která jsou dostupná specifickým skupinám, které splňují určená kritéria.“
- Public access - „Data, která jsou dostupná komukoliv, pod podmínkami užití, které nejsou otevřené.“

### 3. Uzavřená data

- „Data, která jsou přístupná jen jejich subjektu, vlastníkovvi nebo držitelvi.“ [29, překlad vlastní]

#### 6.2.1 Typy otevřených standardů

Standardy lze používat prakticky kdekoli, kde se pracuje s daty, i kdyby se neposkytovali na veřejnost. Norem je však veliké množství, a proto je nutné je kategorizovat. ODI definovalo tři základní skupiny otevřených standardů pro lepší orientaci mezi nimi. Standardy nejsou omezené jen na jednu skupinu. Mohou mít vlastnosti klidně všech tří.



### 6.2.1.1 Standardy pro sdílený slovník

Standardizování slovníku má za cíl co nejvíce zjednodušit čtení dat. Při použití normy je snaha zamezit používání nejednoznačných pojmů a obsahu. Složitost slovníku závisí na datovém obsahu.

Standardy sdíleného slovníku se využívají v ontologii, jelikož s jejich pomocí dosahujeme sémantické interoperability (kapitola 5). [31, překlad vlastní]

### 6.2.1.2 Standardy pro výměnu dat

Definují strukturu dat a práci s nimi. Jedná se o upřesnění formátování, datových typů, pravidel sdílení a mapování dat na ontologické modely. [31, překlad vlastní]

### 6.2.1.3 Standardy pro řízení

Určují procesy práce s daty. Mohou obsahovat definice jednotek pro data, či proces sběru určitých dat a efektivního nakládání s nimi. [31, překlad vlastní]



# OpenPonk

OpenPonk je meta-modelovací platforma vytvořená Centrem pro konceptuální modelování a implementaci. Platforma je open-source, a tudíž je možné ji stáhnout zdarma. Platforma je podporována na operačních systémech Windows, Linux a macOS. OpenPonk je postaven na prostředí Pharo a je vyvíjen v jazyce Smalltalk. [32, 33, překlad vlastní] Pharo je založené na objektově orientovaném programování. [34, překlad vlastní]

Platforma obsahuje několik modulů pro konceptuální modelování a simulaci. Mezi moduly patří:

- UML,
- OntoUML,
- BORM,
- Konečné automaty,
- Petriho síť,
- BPMN. [32, 33, překlad vlastní]

### 7.1 OntoUML

Platforma OpenPonk poskytuje rozšířenou podporu pro modelování v jazyce OntoUML.

Modul poskytuje možnost vytváření entit a relací s již předdefinovaným stereotypem. V rámci vytváření a úpravy entit či relací umožňuje nastavit různé vlastnosti definované v metodice OntoUML, např. *static* nebo *derived*.

Dále modul obsahuje kontrolu správnosti modelu dle definic obsažených v OntoUML. Tato funkcionality umožňuje rychlou a efektivní kontrolu modelu bez možnosti lidského pochybení. Kontrola modelu také obsahuje detekci podezřelých struktur (antipattern). Podezřelá struktura indikuje možný špatný návrh modelu či nedostatečný detail zpracování.

OpenPonk dále umožňuje exportování OntoUML modelu do PNG či SVG formátu. V této chvíli je export do formátu SVG pouze v experimentální verzi.

Kvůli těmto funkcionalitám, byl OpenPonk vybrán jako nástroj pro vytváření ontologických konceptuálních modelů v rámci této bakalářské práce.



Část II

Vlastní řešení



# Datové sady

Bakalářská práce byla realizována v ontologickém týmu pod vedením Bc. Jany Martínkové a Bc. Terezy Macháčové. Jednotlivé části práce byly vypracovány samostatně a pouze konzultovány s vedoucími. V rámci práce byly zanalyzovány, namodelovány a poté namapovány dvě klíčové domény a dále byly vypracovány dvě šablony pro modelování. Obě zpracované datové sady jsou licencovány jako otevřené, z toho důvodu mohou být v bakalářské práci zveřejněny v celém svém rozsahu. Vydavatelem obou datových sad je Český telekomunikační úřad. Na šabloně *sablona-osoby* se podílel celý tým. Pro kontrolu validity šablony byla mou osobou zařízena konzultace s Ing. Miroslavou Šírovou, vedoucí Oddělení daňového procesu, Finančního úřadu pro Liberecký kraj, která dopomohla k lepšímu zachycení reality v konceptuálním modelu.

Datové sady byly poskytnuty společností Remmark, a.s. Mezi poskytovanými sadami se nacházeli jak otevřené, tak soukromé zpeněžené datové sady. Data měla různé poskytovatele, např. Eurostat, Český statistický úřad, či samotný Remmark, a.s. Tato skutečnost, že datové sady mají rozdílné poskytovatele, vede k zpomalení ontologické analýzy. Důvodem tohoto zpomalení je nejednotný formát výstupu statistik. Tato nejednotnost zapříčiňuje výskyt nutnosti znovu zanalyzovat již kompletně namodelovanou a namapovanou doménu, jelikož se zjevil konflikt s jinou doménou. Tyto konflikty mohou být ontologického rázu, např. nevyhovující stereotyp entity, nebo v horším případě, nedostatečně, či dokonce chybně zanalyzovanou a v modelu zachycenou realitou. V případě chybné ontologické analýzy je potřeba analýzu provést znovu, s případnou vyšší mírou detailu, pokud si to konfliktní doména vyžaduje a následně, dle nové analýzy, upravit konceptuální model. Mapování se v tomto případě musí předělávat jen zřídka, protože na koncové entity se tyto konflikty většinou nevztahují.

Před zpracováním ontologické analýzy je ve většině případů nutné provést transformaci dat. Problematikou transformací dat se zabývá odlišný specializovaný tým, a tudíž je popis jejich funkcionality mimo rozsah této bakalářské práce. Z hlediska transformací je relevantní jejich výstup, a to konkrétně tedy struktura výstupních dat. Výsledek transformace může výrazně ovlivnit způsob mapování obsahu datových sad na konceptuální modely. Mohou například ovlivnit počet datových entit, či počet mapovaných atributů. Transformace dat nemusí být nutně jednorázovou akcí. V případě nalezení chyby je možné je předělat a následovně upravit mapování na datové entity v konceptuálních modelech.

Jednotlivé klíčové domény byly vybírány na základě priority určené společností Remmark, a.s. Tato priorita je založena na výčtu potřebných domén k uvedení první produkční verze Inkubátoru OpeNest na trh. Součástí inkubátoru je datová platforma, jež bude v době vydání obsahovat přes dvě stě různých zpracovaných modelů, mezi které patří i modely zpracované v rámci této bakalářské práce.





# Ontologické konceptuální modely

Jedním z výstupů bakalářské práce byly ontologické konceptuální modely. Modely slouží k zachycení struktury a ontologických vlastností dat, které vyplynuly z předcházející ontologické analýzy klíčových domén.

Konceptuální modely jsou vytvořeny za pomoci modelovacího jazyka OntoUML. Jazyk byl vybrán z důvodu jeho specializace na ontologické modelování. Tento jazyk je doporučován iniciativou GoFAIR, která se zabývá zlepšením oblasti poskytování dat. Součástí tohoto zlepšení je i sémantická interoperabilita dat, jež je zároveň i jedním z cílů této bakalářské práce. Zároveň na vylepšování a provozu dokumentace OntoUML spolupracuje *Centrum pro konceptuální modelování a implementaci* (CCMi), jež vede doc. Ing. Robert Pergl, Ph.D., vedoucí této bakalářské práce. Právě jedním z bonusových přínosů byla úprava dokumentace, jelikož v průběhu modelování byly nalezeny nesrovnalosti s výukovými materiály bakalářského předmětu konceptuálního modelování BI-KOM, ve kterém je OntoUML vyučováno.

Notace OntoUML obsahuje detekci podezřelých struktur, takzvaných *anti-patternů*. Z důvodu obsáhlosti jejich řešení jsou, po dohodě s vedoucím, v modelech opomenuty. Důvodem vynechání kontroly *anti-patternů* je jejich časová náročnost řešení a s tím svázaný pravděpodobný mnohonásobný nárůst velikosti modelů. Tento nárůst je zapříčiněn nutným doplněním pomocných entit pro správné řešení *anti-patternů*. Zároveň s časovou náročností by byly zvýšeny i nefunkční požadavky na datovou platformu, jelikož by bylo nutné ukládat a provázat větší struktury.

Pro modelování byla využita platforma OpenPonk, jež je také pod správou CCMi a aktivně se vyvíjí v reakci na požadavky, případně na nalezené chyby, které vyplynuly v průběhu modelování. Díky tomu je jedním z pasivních přínosů této práce také vylepšování platformy OpenPonk, zejména nástroje pro modelování OntoUML modelů.

## 9.1 Šablony

Šablony jsou konceptuální modely, bez datových entit a mapování, jež popisují danou doménu do potřebného detailu. Je v nich zachycena struktura domény se základními vlastnostmi, které jsou běžně využívány, nebo je vysoká pravděpodobnost, že využívány budou. Šablony se od konceptuálních modelů domén nijak neliší. Platí pro ně stejná pravidla notace OntoUML a neplatí pro ně žádné speciální předpisy navíc.

### 9.1.0.1 Princip funkce šablon

Využití šablon spočívá v ulehčení vytváření konceptuálních modelů klíčových domén tím, že není potřeba ve všech modelech vyobrazit problematiku celé domény včetně částí, jež jsou danou

datovou sadou nevyužité.

Šablony fungují na bázi odkazu. Z šablony se vybere skupina entit (*indukovaný podgraf*), obsahující potřebné informace z domény. Tato skupina se poté znovu vymodeluje do konceptuálního modelu pro danou datovou sadu. K namodelovanému podgrafu jsou poté připojeny entity z řešené klíčové domény.

Celá tato funkcionalita je možná díky vlastnostem Datové platformy. Ta při propojování datových sad přes dotazovací jazyk dokáže spojovat modely podle jmen a stereotypů entit. Díky tomu stačí namodelovat pouze indukovaný podgraf šablony, protože ten se při propojování domén v platformě naváže na stejný indukovaný podgraf, který obsahuje šablona.

### 9.1.0.2 Dopad využití šablon

Šablony výrazně ovlivňují rychlost vytváření modelů klíčových domén. Zároveň se snížením časové náročnosti modelování domén dopomáhají k lepší čitelnosti modelů a snížení náročnosti na úložiště Datové platformy.

Proti vytváření a využívání šablon je možné argumentovat ztrátou času na nerelevantních modelech. Avšak opak je pravdou. Čas využitý na tvorbě šablony, která je často využívána, se mnohonásobně vrátí při tvorbě konceptuálních modelů, jenž tuto šablonu využívají. S každým modelem, používajícím šablonu, je ušetřen čas strávený zkoumáním domény šablony a jejího následného modelování.

## 9.2 Tvorba konceptuálního modelu

Tvorba ontologického konceptuálního modelu počíná vykonáním ontologické analýzy vybrané klíčové domény. Cílem této analýzy je návrh struktury dat obsažených v datové sadě tak, aby byla zajištěna sémantická interoperabilita s ostatními zpracovanými datovými sadami a sadami, které budou v budoucnu zpracovány.

Součástí ontologické analýzy je určení konstruktů reality a jejich vlastností a vazeb. Takovými konstrukty mohou být reálné objekty, např. osoba, automobil, nebo abstraktní kategorické vlastnosti, např. léčivý přípravek. V rámci analýzy je potřeba prostudovat věrohodné zdroje popisující danou doménu a nevycházet pouze z vlastních domněnek a zkušeností. Nedostatečné prozkoumání domény může vést ke špatně navrženému konceptuálnímu modelu, který nebude odpovídat všem datovým sadám vázaným na tuto doménu. Poté s narůstajícím počtem modelů, založených na prvním z domény, narůstá i skrytý *technický dluh*, který může mít velmi závažné následky až bude objeven.

Ontologická analýza klíčových domén je rozdělena do několika po sobě jdoucích kroků:

1. Určení koncových entit,
2. Rozdělení koncových entit do souvisejících skupin,
3. Propojení jednotlivých skupin do celistvého modelu.

### 9.2.0.1 Určení koncových entit

Počátečním krokem ontologické analýzy je určení koncových entit. To jsou ty entity, které mají relaci s datovými entitami (sekce 10.1). Definování těchto entit je klíčovým bodem, jelikož koncové entity tvoří základ, od kterého je možné postupovat s analýzou celé domény.

Koncové entity většinou bývají různými vlastnostmi dalších entit. Pomocí koncových entit určíme tyto *mateřské entity*, jenž utvoří první vrstvu detailu modelu. *Mateřské entity* poskytují základy pro hledání informačních zdrojů o doméně.

Označení podnikatele
Průměrná doba realizace připojení /zřízení + zprovoznění/ služby přístupu účastníkovi - hlasové služby
Průměrná doba realizace připojení /zřízení + zprovoznění/ služby přístupu účastníkovi - internet
Četnost poruch služeb přístupu - hlasové služby
Četnost poruch služeb přístupu - internet

■ **Obrázek 9.1** Koncové entity dle struktury

Na obrázku 9.1 lze vidět výňatek ze struktury datové sady *kvalita-sluzeb-v-ekomunikaci-2018* dodané firmou Remmark, a.s. Dle výše popsaného postupu se provede první krok ontologické analýzy, jímž je určení koncových entit datové sady.

Definované koncové entity se namodelují v OntoUML (viz obrázek 9.2). Součástí tohoto kroku je i analýza ontologických vlastností entit. Do té spadá určení jejich stereotypů, které jsou v obrázku také znázorněny.

Po tomto kroku následuje určení *mateřských entit*. Entity lze jednoduše určit z popisu koncových entit, např. *Označení podnikatele* → *Podnikatel*, *Průměrná doba realizace připojení hlasové služby* → *Hlasová služba*. Ovšem koncová a mateřská entita nemusí být nutně propojeny přímo mezi sebou. Mezi entitami se může nacházet několik pomocných entit, které upřesňují jejich reálný vztah (viz obrázek 9.3).

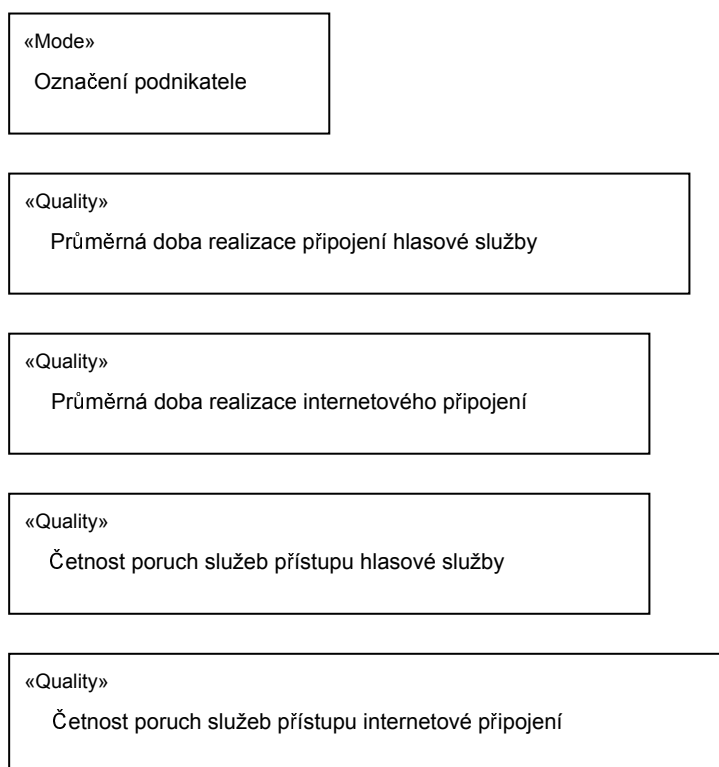
Výsledkem tohoto postupu jsou uskupení mateřských entit s nimi svázaných koncových entit. Tyto uskupení tvoří základy pro ontologickou analýzu. Mateřské entity jsou klíčová slova, podle kterých se hledají informace a koncové entity poskytují záchytné body pro hledání podrobností domény.

### 9.2.0.2 Zachycení podrobností domény

Provedením počáteční analýzy koncových entit se podařilo získat uskupení entit, které je potřeba obohatit o podrobnosti. Důvodem tohoto obohacení je snaha přiblížit modely realitě v dostatečném měřítku, aby byl zachycen její princip, ale model neobsahoval nadbytečné podrobnosti. Je na ontologovi, aby zvážil do jaké míry detailu ztvárnit realitu.

Proces zachycení podrobností počíná výběrem určité skupiny entit, obsahující mateřskou entitu. Mateřská entita je použita jako klíč pro hledání zdrojů popisujících doménu, ve které se entita vyskytuje. Z nalezených zdrojů se poté čerpají informace pro návrh nových entit, jež se poté vloží do konceptuálního modelu domény. Při modelování entit se ontolog řídí pravidly definovanými konceptuálním jazykem OntoUML (viz kapitola 4).

Při ontologické analýze se postupuje dle doporučeného postupu modelování OntoUML modelů. Dle postupu je prvním krokem určení rigidních sortálů. Důvodem je jejich vlastnost poskytování *principu identity*, jež je pro je pro OntoUML velmi významná. Po definování rigidních sortálů následuje vymezení anti-rigidních sortálů, kvůli jejich závislosti na získání *identity* od rigidních sortálů. Poté přichází na řadu určení non-sortálů, kam spadají různé kategorické vlastnosti z domény jako je například *právní osobnost*. Na konec se určí ontologické *aspekty*, kterými



■ **Obrázek 9.2** Namodelované koncové entity e-komunikace

většinou bývají i koncové entity.

Při vytváření konceptuálního modelu je možné narazit na entity, které už jsou obsaženy v modelech jiných klíčových domén. To může vyústit v kolize mezi jmény, či stereotypy entit. Reakcí na tento problém bylo vytvoření *Indexu*, který obsahuje tabulku všech použitých entit včetně jejich jmen, stereotypů a modelů, ve kterých byly využity. Zároveň s *Indexem* byla do *datové platformy* přidána funkcionality kontroly kolizí, která funguje na principu analýzy jazyka a stereotypu. Kontrola kolizí v *datové platformě* dokáže najít i podobné názvy a překlepy. Díky těmto dvěma aplikacím lze udržovat *sémantickou interoperabilitu* mezi modely.

Tento proces je proveden pro všechna uskupení koncových a mateřských entit, čímž se získá fragmentovaný podrobný popis domény. Posledním krokem ontologické analýzy klíčové domény je sjednocení fragmentů do jednoho celistvého konceptuálního modelu.

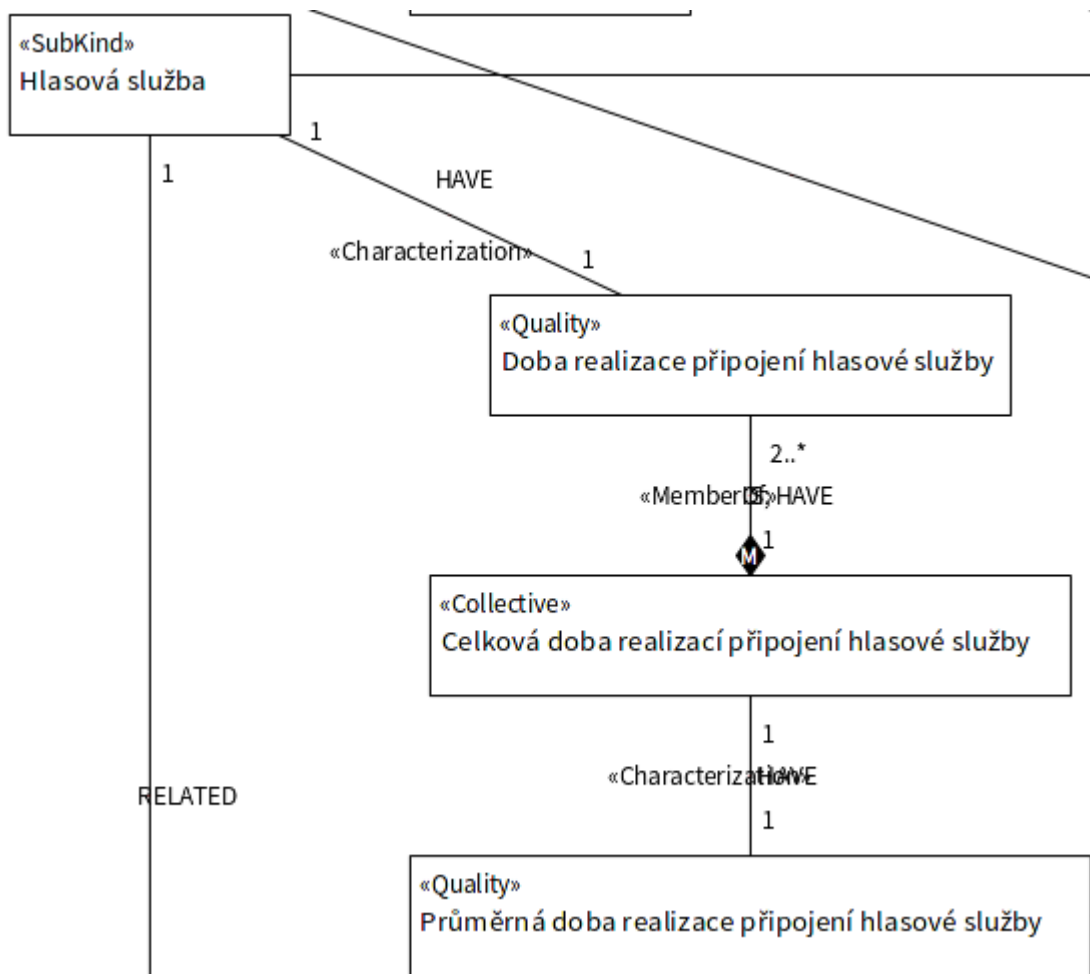
### 9.2.0.3 Propojení fragmentů modelu

Nyní model obsahuje samostatné fragmenty, které je nutno propojit za účelem vytvoření jednotného modelu, připraveného pro mapování obsahu datové sady.

Nejprve je potřeba získat informace. Hlavním cílem sběru informací je nalézt souvislosti mezi jednotlivými fragmenty. Jednou z efektivních cest je postupování po dvojicích. Tento přístup funguje na bázi postupného hledání relací každého s každým a dosažení úplného modelu.

Při hledání souvislostí mezi fragmenty je možné narazit na novou doménu, která ještě nebyla v modelu zachycena. Poté je nutné se vrátit k druhému kroku vytváření konceptuálního modelu (viz 9.2.0.1) a doménu zanalyzovat a vytvořit její náležitý fragment. Po zdokumentování nové domény se lze navrátit k spojování fragmentů.

Po dokončení propojení všech uskupení entit vzniká celistvý ontologický konceptuální model splňující zásady OntoUML, který lze použít pro funkci mapování obsahu datové sady.



■ **Obrázek 9.3** Ukázka propojení mateřské a koncové entity

## 9.3 Praktický příklad

V následujícím příkladu je popsán postup modelování pro klíčovou doménu **Ukazatele o trhu elektronických komunikací (mobilních sítí) v ČR** (zkrácené kódové označení *trh-mobilnich-siti*) včetně popisu problémů, které analýzu a modelování provázely už od počátku ontologické analýzy.

Počátkem bylo zanalyzování struktury datové sady. Ze struktury vyplynuly *koncové entity* domény a jejich *mateřské entity*. Z vybraných koncových entit byly vydefinovány tyto mateřské entity:

- SIM karta,
- Hlasový provoz,
- Mobilní síť,
- Služby v mobilních sítích,
- SMS/MMS,
- Volání v mobilních sítích.

Tento výčet entit tvoří základy ontologické analýzy domény trhů mobilních sítí. Před začátkem vyhledávání informací byla každá z entit zadána do aplikace *Index* za účelem zjištění, jestli se už neobjevuje v některém z hotových modelů. Průnik entit byl nalezen v modelu *MML-mobilni-telefon*, jenž popisuje doménu vlastnění a používání mobilního telefonu, včetně výčtu jeho funkcí a napojení na mobilního operátora. Z důvodu zpeněžení datové sady nemohl být zveřejněn konceptuální model marketingové datové sady *MML-mobilni-telefon*, proto konflikty mezi modely budou popsány pouze slovně. Po zmapování možných konfliktních entit přišla na řadu ontologická analýza domény, při které byly tyto entity brány v potaz.

V rámci ontologické analýzy byly definovány tyto domény, které je potřeba detailně namodelovat:

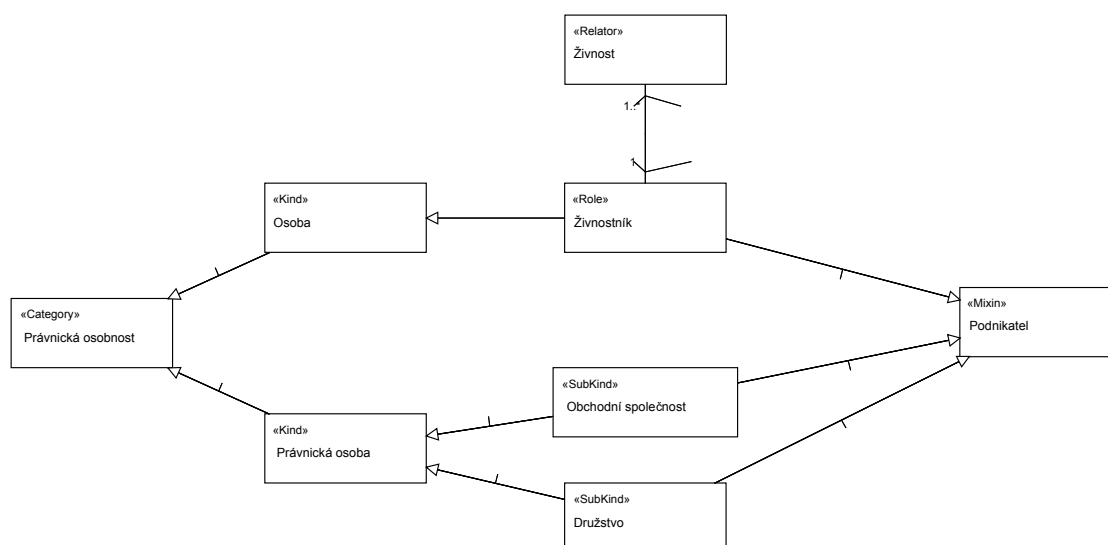
1. Hierarchie a provoz telefonních sítí,
2. Telefonní a mobilní operátoři,
3. SIM karta,
4. Mobilní provoz.

### 9.3.1 Telefonní sítě a operátoři

Domény telefonních sítí a operátorů byly zpracovány společně z důvodu jejich vysoké propojenosti. Při analýze byla definována hierarchická struktura druhů telefonní sítí včetně vlastností jako jsou *veřejné/soukromé*, či *národní/mezinárodní*. Dále byly definovány typy telefonní operátorů s jejich závislostí na získání licence od telekomunikačního úřadu. Právě při analýze telefonních operátorů se objevily hned dva problémy.

#### 9.3.1.1 Kolize mezi modely

Prvním problémem byla kolize s modelem *MML-mobilni-telefon*. Kolize spočívala v nedostatečném detailu popisu domény. V modelu byla zakreslena nutnost získání telekomunikační licence pro každého mobilního operátora a dědičnost mobilního operátora od obchodní společnosti. Oba tyto zakreslené abstrakty reality jsou jen z části správné.



■ **Obrázek 9.4** Původní šablona pro podnikatele

Mobilní operátor je definován poskytováním služeb na mobilní síti. Telekomunikační licence je však potřeba k **provozování** mobilní sítě, ne k poskytování služeb na ní. Tato skutečnost tedy udává, že mobilní operátoři nemusí nutně vlastnit telekomunikační licenci, aby poskytovali své služby na mobilní síti. Tito operátoři se nazývají *virtuální mobilní operátoři* a fungují na bázi pronajímání mobilních sítí od provozovatelů s licenci a následného poskytování služeb na pronajaté síti. Kolizi lze tedy opravit detailnějším zpracováním reality.

Dědičnost mezi mobilní operátorem a obchodní společností také není zcela přesná. Vydání telekomunikační licence není omezeno na právnické osoby. Ve všech oprávněních vydaných Telekomunikačním úřadem je jasně definováno „fyzické a právnické osoby“ [35, 36], respektive „podnikatel“ [37]. Tuto skutečnost zároveň potvrzuje i existence *virtuálních mobilních operátorů*. Kolizi lze opravit vhodným napojením na existující šablonu domény podnikatelů.

### 9.3.1.2 Šablona podnikatelů

Následujícím problémem, který bezprostředně navazoval na předchozí problém s dědičností, byl nedostatečný detail šablony pro podnikatele. Šablona obsahovala pouze základní strukturu pro fyzické a právnické osoby a jejich vztahy s podnikatelem (viz obrázek 9.4). V šabloně lze vidět absenci rozdělení podnikatele od osob provádějících nevýdělečné činnosti, či detailnější zpracování *Osoby samostatně výdělečně činné*.

Kolize mezi modelem a šablonou spočívala v nekompatibilní definici podnikatele. Pro účely modelu bylo nutné definovat novou entitu *Poskytovatel*, jež není závislá na podnikateli. Tato entita je navázána na službu pomocí *relátoru* poskytování služby. Oddělení podnikatele a poskytovatele bylo nezbytné z důvodu možnosti poskytování služeb bez výtěžku.

V rámci úpravy šablony byly doplněny i další podrobnosti, které vzešly z dalších modelů vyhotovených ostatními členy modelovacího týmu. Celkově se šablona rozšířila o problematiku poskytovatele, zápisu podnikatele do rejstříků, zaměstnance, pracovně právního vztahu a nevýdělečné činnosti (viz model *šablona-osoby* v příloženém médiu A).

Tvorbou šablony se zabýval celý tým konceptuálních modelářů. Část modelu obsahující podnikatele byla konzultována s Ing. Miroslavou Šírovou, jež má dlouhou praxi v oboru, jakožto vedoucí Oddělení daňového procesu, Finančního úřadu pro Liberecký kraj.

Šablona se v budoucnu bude ještě dále rozšiřovat s přibývajícím požadavky na detaily. Nejbližším rozšířením bude zahrnutí problematiky provozovatele, které taktéž vzniklo ze zapro-

vávaného modelu *trh-mobilních-sítí* (viz model *trh-mobilních-sítí* v příloženém médiu A). Než proběhne rozšíření šablony, bude problematika dočasně celá vyobrazena v modelu mobilních sítí.

### 9.3.2 SIM karta

Doména SIM karty byla též v kolizi s modelem *MML-mobilní-telefon*. Byl zde chybně identifikovaný problém se stereotypem entity SIM karta, který měl hodnotu «*Kind*». Domněnkou bylo, že by SIM karta měla být spíše «*Relator*». Po konzultaci s vedoucími modelovacího týmu vzešlo najevo, že o kolizi se nejedná a model *trh-mobilních-sítí* bude expandovat doménu SIM karty.

V rámci modelu domény byly identifikovány a namodelovány druhy SIM karet (předplacená, s vyúčtováním, M2M), závislost SIM karty na mobilním operátorovi a podrobnosti o aktivních SIM kartách.

### 9.3.3 Hlasový provoz

Doména hlasového provozu je úzce spjatá s doménou mobilních sítí. Toto spojení vyplývá z pojmenování různých druhů hlasového provozu podle druhů mobilních sítí. Jelikož se v datové sadě využívají všechny druhy hlasových provozů, tak byly rovnou definovány ve struktuře datové sady. Definování potřebné struktury provozů výrazně urychlilo ontologickou analýzu této domény.

Hlavním cílem ontologické analýzy této domény bylo určit rozdíl mezi hlasovým provozem a hlasovým hovorem. Na první pohled se zdá, že by mohlo jít o jednu a tu samou věc. Tuto domněnku lze vyvrátit analýzou struktury klíčové domény, ve které lze najít čísta struktury, jenž obsahují popis vlastností hlasového provozu. Jedná se o části *tržby za provolané minuty* a *počet provolaných minut*. Z jmenovaných částí lze vydedukovat, že hlasový provoz je «*Collective*» (popis stereotypu, sekce 4) skládající se z mnoha hlasových hovorů, jenž dohromady udávají hodnoty vlastností.

### 9.3.4 Propojení domén

Posledním krokem vytváření konceptuálního modelu je propojení všech fragmentů do jednotného modelu. V tomto případě jsou fragmenty v modelu *telefonní sítě a operátoři*, *SIM karta*, *Hlasový provoz* a *SMS a MMS*, který je společný s modelem *MML-mobilní-telefon*.

#### 9.3.4.1 Telefonní síť s operátory a SIM karta

Mezi doménami je potřeba zobrazit závislost SIM karet na mobilním operátorovi. Mobilní operátor poskytuje své mobilní služby za pomoci *aktivované* SIM karty.

Z podstaty akce aktivace karty je možné rozdělit karty na *aktivované* a *neaktivované*. Jelikož je tento stav proměnlivý v průběhu života instance SIM karty, lze tyto vlastnosti zakreslit do modelu pomocí anti-rigidních sortálů «*Role*», nebo «*Phase*» (popis stereotypů, sekce 4). Ve výběru stereotypů rozhoduje existenční závislost entity. Protože stav SIM karty závisí na akci mobilního operátora, tak lze konstatovat, že existuje *existenční závislost* mezi aktivovanou SIM kartou a akcí aktivace.

Existenční závislost potřebuje nějaké pečetiďlo, jenž při zániku instance bude indikovat i zánik relace. Pečetidlem pro závislost mezi SIM kartou a mobilním operátorem je akce *Aktivace SIM karty* se stereotypem «*Relator*» (viditelná v modelu *trh-mobilních-sítí* v příloženém médiu A).

Pro potřeby závislosti je však ještě potřeba definovat anti-rigidní entitu pro mobilního operátora. Důvodem pro existenci této entity je fakt, že instance obyčejného mobilního operátora není existenčně závislá na aktivaci SIM karty. Proto byla definována entita *Mobilní operátor poskytující SIM karty*. Tato entita typu «*Role*» dědí od mobilního operátora a je existenčně závislá na aktivaci karet.



Poslední akcí je propojení všech tří nových entit. Pro jejich propojení je použita relace typu «*Mediation*».

#### 9.3.4.2 Hlasový provoz a mobilní síť

Entity hlasový provoz a mobilní síť jsou mezi sebou propojeny nepřímo. Jejich závislost závisí na telefonní hovoru. Volaná mobilní síť určuje druh hlasového hovoru, a tím pádem i do kterého druhu kolektivu bude hovor zařazen.

#### 9.3.4.3 SMS/MMS s hlasovým provozem

Doména SMS/MMS využívá výňatek struktury modelu *MML-mobilni-telefon*. Ve struktuře zmíněného modelu je definován i telefonní hovor. Díky vyřešení konfliktů mezi modely sítě a mobilního telefonu, je možné použít entity struktury a považovat model sítě jako jejich rozšíření.



# Datový model

Datový model je konceptuální model, jehož účelem je specifikovat na jaké entity se budou mapovat data z datové sady. Aby byl model vhodný pro mapování, musí obsahovat *datové entity* a *pravidla mapování*.

V případě realizace této bakalářské práce a všech konceptuálních modelů pro projekt Datového inkubátoru je datový model pouze rozšíření ontologického konceptuálního modelu o datové entity a mapování, tzn. že například model *trh-mobilních-síti* je ontologický model (obsahuje ontologický popis domény) a zároveň je i datovým modelem (obsahuje datové entity a mapovací pravidla).

Datový model sám o sobě nevyužívá sémantická pravidla OntoUML, a tudíž je ani nesplňuje. Při modelování datového modelu se využívají principy obvyčejného UML.

## 10.1 Datové entity

Datové entity využívají ke své definici sémantiku UML (kapitola 3). Ta definuje třídu jako entitu s názvem a vlastnostmi bez stereotypů určujících typ třídy.

Pro relaci mezi *datovou entitou* a *koncovou entitou* se používá asociace. Asociaci se nedefinují stereotypy, jelikož nevyužívá sémantiku OntoUML (kapitola 4). Relace má definované dvě multiplicity, které **musí** být totožné na obou koncích relace. Tyto multiplicity jsou **1** a **0..1**. Multiplicita **rovna 1** se využívá, pokud je *datová entita* napojena na **pouze jednu** *koncovou entitu*. Násobnost **0..1** se využívá, když je *datová entita* v relaci s **více koncovými entitami**. Znázorňuje se tím možnost výběru *koncové entity* pro mapování.

### 10.1.1 Standard názvu

Pro účely datového mapování byl vymyšlen standard pro pojmenovávání datových entit. Standard definuje jméno entity jako textový řetězec skládající se ze jmen všech koncových entit, na které je datová entita napojena, poskládaných za sebou. Řetězec je ukončen klíčovým slovem *Data*.

Do standardu bylo přidáno i pravidlo, které upravuje strukturu názvu, pokud je poskládaný řetězec jmen nadměrně dlouhý. Toto pravidlo definuje řetězec jako klíčová slova, jenž jednoznačně popisují data, která daná datová entita mapuje. Nebo je řetězec definován jako název datové sady, pokud entita slouží k mapování většiny obsahu sady.

## 10.1.2 Vlastnosti entit

Vlastnosti datových entit jsou definovány názvem a datovým typem. Datové typy jsou obvyčejné datové typy definované UML, např. *String*, nebo *Int*.

Vlastnosti určují jaký obsah se mapuje pomocí datové entity, ke které patří. Proto název musí odpovídat názvu *ukazatele* v datové sadě, popřípadě názvu definovaného transformací dat. Jedna datová entita může obsahovat více vlastností, tudíž přes jednu entitu lze mapovat více částí datové sady. Týmový konsenzus však definuje, že každá *koncová entita* by měla mít vlastní *datovou entitu*, která mapuje její data. Tento konsenzus neplatí, pokud transformace výrazně upravuje strukturu obsahu datové sady, např. sjednocení všech hodnot ukazatelů pod jeden sloupec *ukazatel*.

Všechny datové entity **musí** obsahovat vlastnost s jménem *identifier* s datovým typem *String*, jenž obsahuje jedinečný identifikátor datové sady, pro kterou se mají použít *mapovací pravidla* obsažené v entitě. Zároveň musí být vlastnost *identifier* nastavena jako *static*.

## 10.2 Mapovací pravidla

Mapovací pravidla fungují na bázi přiřazení vlastnosti *datové entity*, jenž obsahuje mapovací pravidla, na *koncovou entitu*, se kterou je v relaci. Pravidla mají podobu komentáře datové entity.

### 10.2.1 Předpisy mapovacích pravidel

Pro mapovací pravidla byl navržen jazyk, jehož cílem je usnadnění a zefektivnění práce s mapováním. Jazyk obsahuje tři definice pravidel, které slouží jako šablony.

Při používání pravidel je **nutné** dodržovat syntaktické předpisy do detailu. Překladač pravidel je závislý i na mezerách mezi příklady, např. „->“ musí mít kolem sebe mezery, jinak pravidlo nebude validní.

Obsah datové sady lze chápat jako tabulku, kde vlastnosti jsou názvy sloupců. Tyto sloupce poté obsahují konkrétní data.

#### 10.2.1.1 Základní mapovací pravidlo

Předpis funguje na principu přiřazení sloupce z tabulky dat přímo na vybranou entitu. *Vlastnost* definuje jméno sloupce, ze kterého bude brán obsah. *Entita* definuje *koncovou entitu*, jenž je v relaci s *datovou entitou*.

##### ■ Výpis kódu 10.1 Základní mapovací pravidlo

```
"vlastnost" -> "entita";
```

#### 10.2.1.2 Entita podmíněná obsahem vlastnosti

Jedná se o rozšíření základního pravidla o možnost přiřazení dat entitě na základě obsahu vybraného sloupce. Jelikož se jedná o rozšíření, tak *vlastnost* i *entita* mají stejný význam jako v základním pravidlu. *Hodnota* odpovídá hodnotě, která se nachází ve vybraném sloupci.

##### ■ Výpis kódu 10.2 Entita podmíněná obsahem vlastnosti

```
"vlastnost" = "hodnota" -> "entita";
```

Obsahem řetězce *hodnota* může být:

Název	Datový typ	Popis	Formát
oznaceni_podnikatele	Název podnikatele		text
kategorie	kategorie	kategorie	text
hodnota	hodnota	hodnota	text

■ **Tabulka 10.1** Popis struktury datové sady po transformaci

### 1. Textový řetězec

Píše se do anglických uvozovek, např. *"text"*.

### 2. Číselná hodnota

Píše se **bez** uvozovek, např. *587*.

### 3. Regulární výraz

Píše se mezi dvě lomítka, např. */tel(efon)?/*.

V podmínce lze využívat i logické operátory jako jsou například *AND* či *OR*. Možnost jejich využití dovoluje mapování na více entit zároveň.

#### ■ Výpis kódu 10.3 Mapování na více entit

```
"vlastnost" -> "entita1" AND "entita2";
```

Nebo mapování na entitu za složitější podmínky.

#### ■ Výpis kódu 10.4 Mapování se složitější podmínkou

```
"sloupec1" = "hodnota1" AND "sloupec2" = "hodnota2" -> "entita";
```

### 10.2.1.3 Podmíněné pravidlo

Rozšíření předchozích dvou pravidel o možnost podmínit celé pravidlo mapování. Pravidlo se skládá z *podmínky* (před *->*) a *podmíněného pravidla* (za *->*). Při tvorbě pravidla je možné využívat logické operace a obsah pole *hodnota* má stejná pravidla jako *entita s podmíněným obsahem vlastnosti*.

Využívá se, pokud datová sada obsahuje sloupec, který ovlivňuje mapování dalších entit. Jedná se většinou o změnu hodnotek v závislosti na datech.

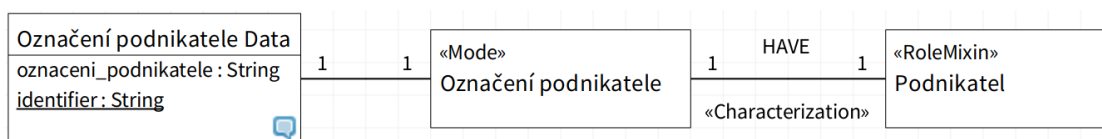
#### ■ Výpis kódu 10.5 Podmíněné pravidlo

```
("vlastnost" = "hodnota") -> ("vlastnost" -> "entita");
```

## 10.3 Praktický příklad

Následující příklad popisuje postup mapování pro datovou sadu **Mapování kvality poskytovaných služeb e-komunikací 2018** (zkrácené kódové označení *kvalita-sluzeb-v-ekomunikaci-2018*).

Mapování datové sady e-komunikací vychází z transformací dat provedených nad sadou. Transformace upravují strukturu datové sady. Jejich výstupem jsou tři sloupce, jež obsahují všechna data (viz tabulka 10.1).



■ **Obrázek 10.1** Ukázka datové entity *Označení podnikatele*

## 10.3.1 Mapování podnikatelů

### 10.3.1.1 Datová entita

Sloupec **oznaceni\_podnikatele** obsahuje názvy podnikatelů v textové podobě (viz tabulka 10.1). Z podstaty obsahu sloupce vyplývá, že *koncovou entitou*, jenž bude napojena na *datovou entitu*, bude entita **Označení podnikatele**. *Datová entita* tedy bude pojmenována **Označení podnikatele Data**.

Jelikož datová entita obsluhuje sloupec s názvem **oznaceni\_podnikatele**, tak musí mít entita stejnojmennou vlastnost. Sloupec má definovaný datový formát „text“, tudíž musí být vlastnosti definován datový typ *String*.

Každá datová entita musí mít definovanou *statickou* vlastnost *identifier* s typem *String*, která určuje, k jaké datové sadě entita patří. Vlastnost byla inicializována na hodnotu *kvalita-sluzeb-v-ekomunikaci-2018*, což je kódové označení zpracovávané datové sady.

Posledním krokem bylo napojit datovou entitu s koncovou entitou. Vazbou je *asociace* s *multiplicitou rovnou 1* na obou stranách, jelikož je datová entita v relaci pouze s jednou koncovou entitou (viz obrázek 10.1).

### 10.3.1.2 Mapovací pravidla

Datová entita *Označení podnikatele* obsluhuje jeden řádek v tabulce 10.1 a je napojena na pouze jednu entitu, tudíž je potřeba vytvořit jen jedno mapovací pravidlo, jenž pokryje celý sloupec s daty.

Pro pravidlo byla využita šablona pro základní pravidlo (sekce 10.2.1.1). Do pole *vlastnost* bylo doplněno *oznaceni\_podnikatele*, jakožto sloupec jenž je snaha namapovat, a do pole *entita* byl vyplněn název koncové entity, na kterou je cílem data namapovat.

■ **Výpis kódu 10.6** Mapovací pravidlo pro označení podnikatele

```
"oznaceni_podnikatele" -> "Označení podnikatele";
```

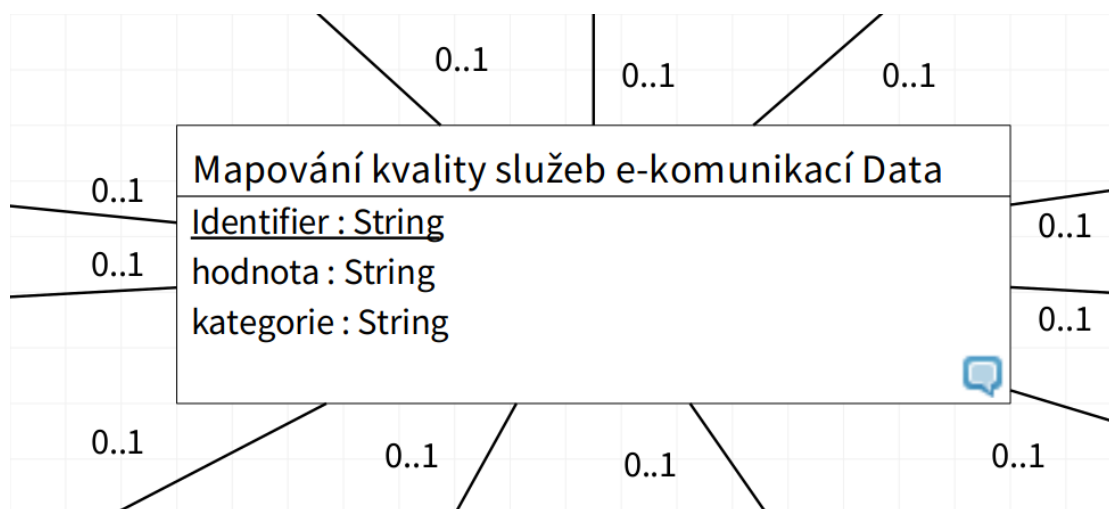
## 10.3.2 Mapování kategorie a hodnoty

Sloupce **kategorie** a **hodnota** se musí mapovat společně, jelikož obsah pole **kategorie** určuje, na kterou koncovou entitu se bude mapovat obsah pole **hodnota**.

### 10.3.2.1 Datová entita

Jelikož spolu sloupce **kategorie** a **hodnota** úzce souvisí, je nutné je oba zahrnout do jedné datové entity, která je oba obsluží (viz obrázek 10.2).

Nejdříve bylo potřeba pojmenovat entitu. Zde se objevil problém se standardem pojmenování, který definoval název datové entity jako seskupení názvů koncových entit postupně za sebou. V tomto případě je ale entita v relaci s jedenácti koncovými entitami, jenž všechny mají více než pět slov v názvu. Výsledkem poskládání názvů za sebe byla entita, jejíž název ji dělal širší



■ **Obrázek 10.2** Datová entita pro kategorie a hodnoty

než celý konceptuální model. To z hlediska přehlednosti není kvalitní řešení, proto byl upraven standard a entita byla pojmenována tak, aby výstižně definovala svůj obsah.

Dalším krokem bylo definování vlastností datové entity. Protože entita obsluhuje dva sloupce, tak jsou potřeba dvě entity, jež jim odpovídají. Byly tedy vytvořeny dvě vlastnosti. Jednou vlastností je *kategorie* a druhou vlastností je *hodnota*. Oběma byl definován datový typ *String*, jelikož mají v tabulce struktury přidělený formát „text“ (viz tabulka 10.1).

Jako všem datovým entitám, tak i této bylo nutné vytvořit statickou vlastnost *identifíer* s datovým typem *String*, jež byla inicializována identifikátorem datové sady (*kvalita-sluzeb-v-ekomunikaci-2018*).

Naposled bylo potřeba navázat datovou entitu na koncové entity. Relace jsou typu *asociace* s multiplicitou **0..1** na obou koncích. Důvodem pro nepovinnou relaci určenou multiplicitou je *podmíněné mapování*.

### 10.3.2.2 Mapovací pravidla

Jelikož datová entita obsluhuje jedenáct koncových entit, které jsou definované sloupcem *kategorie*, je nutné vytvořit pro každou z nich odpovídající mapovací pravidlo. Zároveň je potřeba v pravidlech definovat závislost mezi *hodnotou* a *kategorií*. K tomu dopomůže použití *podmíněného pravidla*.

Pro použití *podmíněného pravidla* byla využita šablona (sekce 10.2.1.3). Jako podmínka slouží obsah sloupce *kategorie*. Podle obsahu sloupce se určí, k jaké entitě se namapuje sloupec *hodnota* (viz ukázka kódu 10.3.2.2).

**■ Výpis kódu 10.7** Ukázka mapovacích pravidel pro kategorii a hodnotu

```
//Ukázka má upravené formátování, aby byla lépe čitelná

("kategorie" = "Průměrná doba přihlášení obsluhy pracoviště
informačních služeb přijímající hlášení účastnických poruch")
-> ("hodnota" -> "Průměrná doba přihlášení obsluhy pracoviště
informačních služeb přijímající hlášení účastnických poruch");

("kategorie" = "Průměrná doba odstranění poruchy
na účastnickém přípojném vedení")
-> ("hodnota" -> "Průměrná doba odstranění poruchy
na účastnickém přípojném vedení");

("kategorie" = "Průměrná doba odstranění poruchy pro přístup
k Internetu /s výjimkou poruch na účastnickém přípojném vedení
/ poruch na účastnickém přípojném vedení")
-> ("hodnota"-> "Průměrná doba odstranění poruchy přístupu
k Internetu s výjimkou poruch na účastnickém přípojném vedení");
```



# Zajištění sémantické interoperability

Všechny klíčové domény zpracovávané v rámci projektu OpeNest (sekce 6.1) jsou heterogenní datové sady. Jejich heterogenita vychází z podstaty sběru informací o populaci a vytváření statistik. Pro efektivní práci s těmito sadami je nutné zajistit sémantickou interoperabilitu. Sémantická interoperabilita mezi ontologickými konceptuálními modely vytvořenými pro klíčové domény je zaručena jednotným pojmenováním entit se stejným významem a shodnými OntoUML stereotypy.

Cílem této bakalářské práce není zajistit sémantickou interoperabilitu mezi ontologickými modely vytvořenými pouze v rámci práce, ale mezi všemi modely, které byly dosud vytvořeny v rámci projektu OpeNest.

K interoperabilitě mezi modely významně dopomohly šablony. Konceptuální model *kvalita-sluzeb-v-ekomunikaci-2018* využívá dvě šablony. Těmito šablonami jsou *sablona-osoby* a *sablona-sluzby*. Model *trh-mobilnich-siti* využívá pouze šablonu *sablona-osoby*.

Jelikož jsou doména e-komunikací a doména mobilních trhů velmi specifické, tak se v nich nenachází žádný průnik entit, kromě šablony právnické osobnosti. Sémantická interoperabilita je tudíž mezi modely vyřešena.

Průnik entit byl nalezen mezi zpracovávaným modelem *trh-mobilnich-siti* a již hotovým modelem *MML-mobilni-telefon*. Domény mají mezi sebou společnou problematiku SIM karet, telefonních hovorů, SMS/MMS zpráv a mobilního operátora.

Doména trhů mobilních sítí se zabývá různými druhy SIM karet (viz obrázek A.1). Doména mobilního telefonu popisuje závislost SIM karty na mobilní telefon (viz obrázek A.2). Aby byla zachována interoperabilita bylo potřeba zanalyzovat entitu SIM karty v modelu mobilního telefonu, jestli vyhovuje modelu trhů.

Při zajišťování interoperability mezi modely trhu mobilních sítí a mobilního telefonu bylo nutné definovat rozdíly mezi telefonním hovorem a hlasovým provozem. Nakonec bylo zjištěno, že hlasový provoz je kolektiv hlasových hovorů (viz obrázky A.3 a A.4).

Díky jednotnému pojmenování entit bylo možné využít sémantické interoperability při modelování entit SMS, resp. MMS zprávy. Stejně pojmenování umožnilo využití entit, které již byly definovány v modelu *MML-mobilni-telefon* (viz obrázek A.6) k potřebám domény trhů mobilních sítí (viz obrázek A.5).

Poslední problémovou entitou byla entita *mobilní operátor* (viz obrázek A.8). Při zajišťování sémantické interoperability byly nalezeny nesrovnalosti v modelu *MML-mobilni-telefon*. V něm se nacházelo zobrazení získávání *telekomunikační licence*. Toto vyobrazení však nebylo správné. Chyba byla prodiskutována s vedoucími na pravidelné konzultaci. Z diskuze vyplynulo, že v modelu *trh-mobilnich-siti* bude vyobrazena nová detailní definice mobilního operátora a získání

telekomunikační licence. Model *MML-mobilni-telefon* se následně upraví dle modelu trhů mobilních sítí, aby byla zajištěna sémantická interoperabilita.

Sémantická interoperabilita mezi modely byla zařízena pomocí jednotného pojmenování entit a definování jejich ontologických vlastností. Interoperabilita byla zajištěna jak mezi modely zpracovávanými v rámci bakalářské práce, tak mezi ostatními modely v Datové platformě.

## Závěr

Cílem bakalářské práce bylo ontologicky zanalyzovat klíčové domény a propojit je s datovými sadami tak, aby byla umožněna jejich sémantická interoperabilita. Potřebné datové sady byly vybrány experty ze společnosti Remmark, a.s., pro kterou byly zároveň datové sady analyzovány v rámci subprojektu Datový inkubátor, který je začleněn do projektu OpeNest. Analýza a tvorba modelů probíhala v týmu pod vedením Bc. Terezy Macháčové a Bc. Jany Martínkové.

V rámci práce byly vytvořeny ontologické konceptuální modely vybraných domén, jenž byly následně propojeny s příslušnými datovými sadami. Při analýze domén byl nejdříve proveden rozbor transformací dat, aby odpovídal popisu datové sady. Dále byla prozkoumána struktura dat, ze které byly určeny koncové entity, na které se navázaly jim příslušné datové entity.

Pro vytvoření kompletního konceptuálního modelu bylo nutné obohatit samotné koncové entity o další s nimi související entity. V rámci rozšiřování modelu byla snaha o propojení všech koncových entit dané datovou sadou s dostatečnou mírou detailu, aby zhruba odpovídala skutečnosti. Díky tomuto zpestření modelů se může nad modely používat speciální dotazovací jazyk, vytvořený společností Remmark, a.s., v rámci datové platformy. Jazyk dokáže najít souvislosti mezi zadanými entitami pomocí jejich názvů a relací mezi nimi a entitami s nimi svázanými.

Dále byly vytvořeny šablony. To jsou konceptuální modely bez vazby na konkrétní doménu, zachycující určitou část skutečnosti, např. šablona zachycující problematiku podnikatele. Šablony usnadňují práci s vytvářením nových konceptuálních modelů klíčových domén, jelikož umožňují vynechat „zbytečnou“ část skutečnosti, která by jinak byla v modelu redundantní.

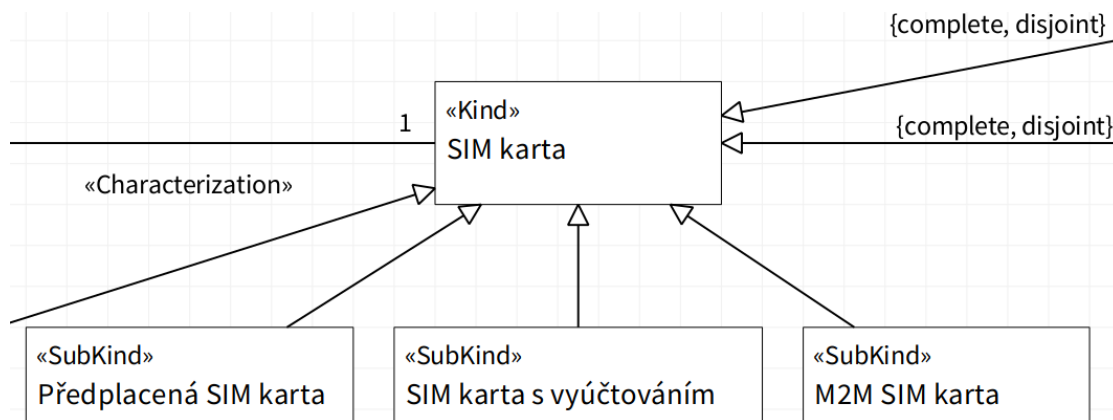
Sémantické interoperability mezi jednotlivými modely bylo dosaženo konvencemi v pojmenovávání entit, verifikací duplikací názvů, verifikací kolizí stereotypů entit v datové platformě a pravidelnými konzultacemi. Díky dosažení sémantické interoperability mezi modely dokáže datová platforma modely provázat a provádět nad nimi operace.

V projektu OpeNest se do budoucna počítá se zpracováním stovek datových sad, které budou tvořit základ datové platformy než vejde v produkci. V rámci projektu se také dále rozvíjí platforma OpenPonk, do které bude v budoucnu přidána možnost modularizace modelů a tím bude výrazně zefektivněn proces tvorby dalších konceptuálních modelů.

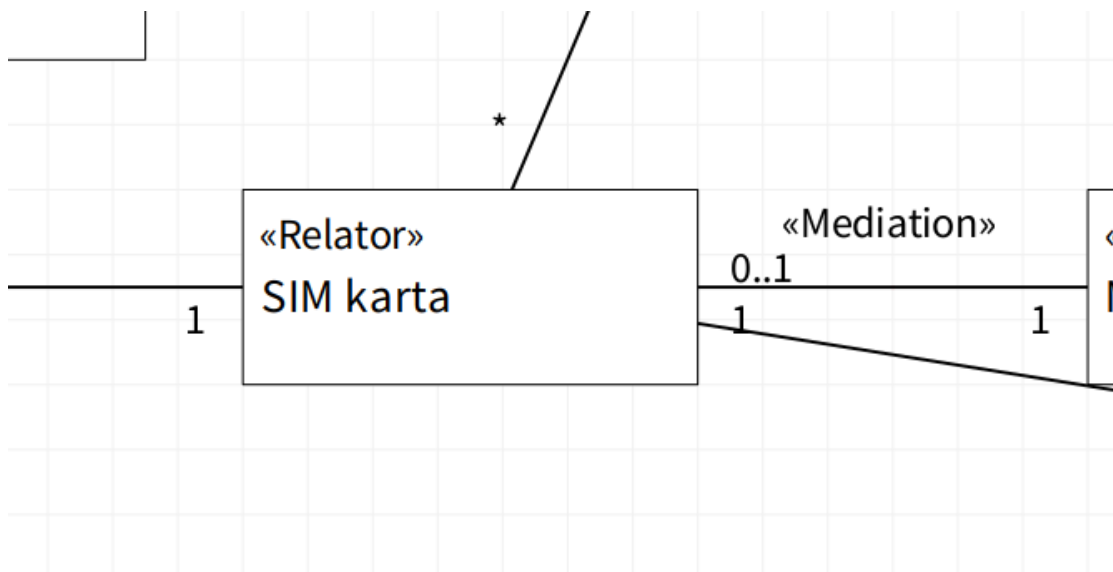


# Příklady sémantické interoperability

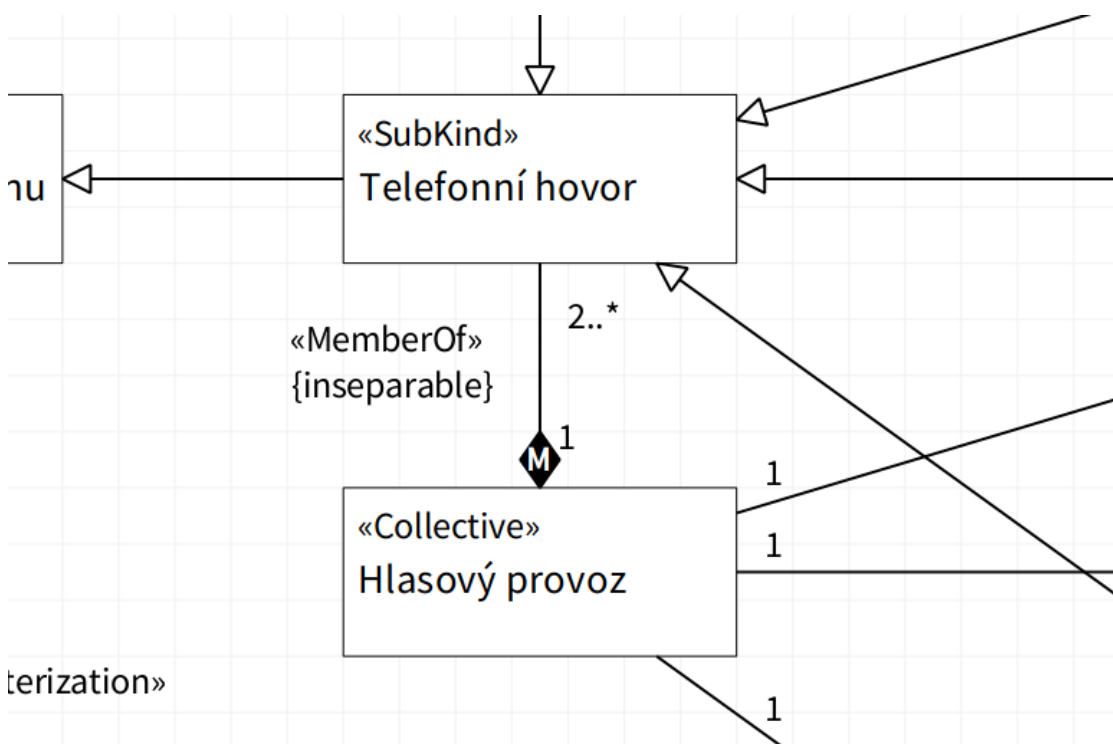
Obrázky zobrazující zajištění sémantické interoperability mezi heterogenními daty pomocí ontologických konceptuálních modelů.



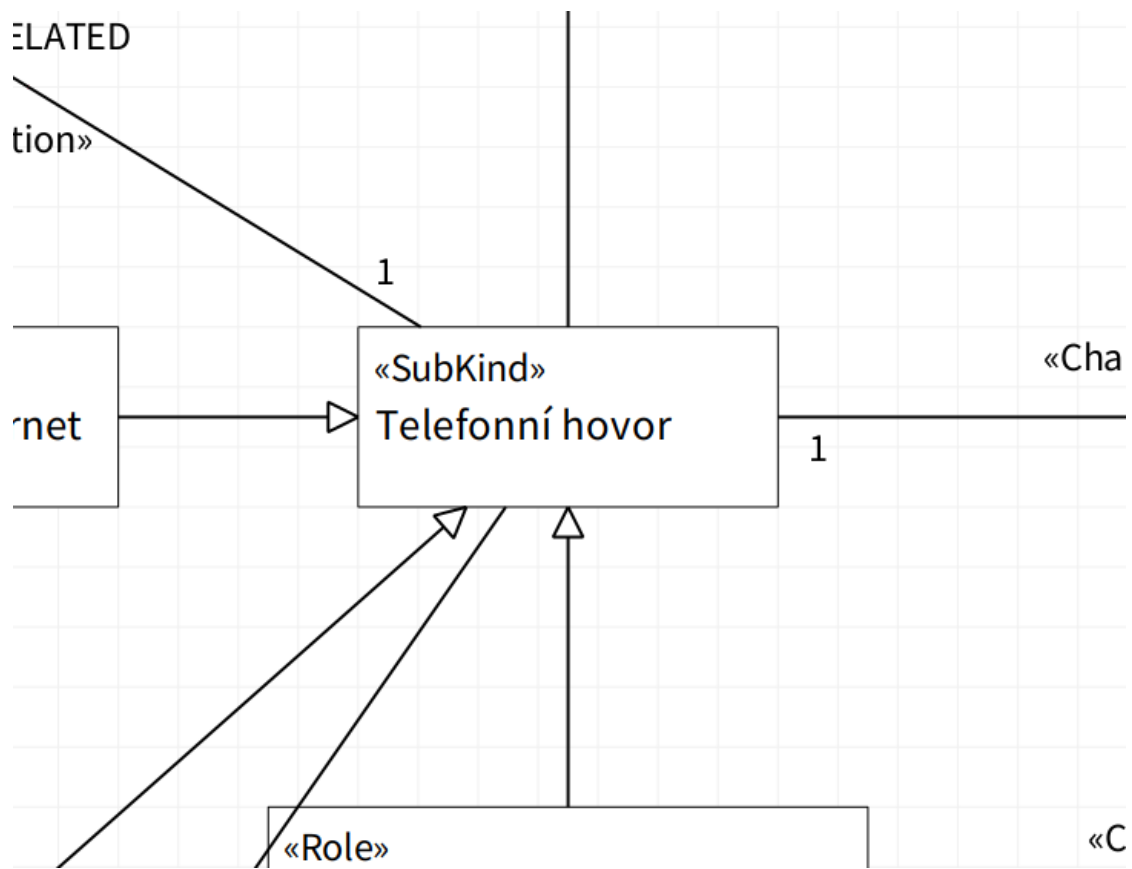
■ **Obrázek A.1** Entita SIM karta v modelu trh-mobilnich-siti



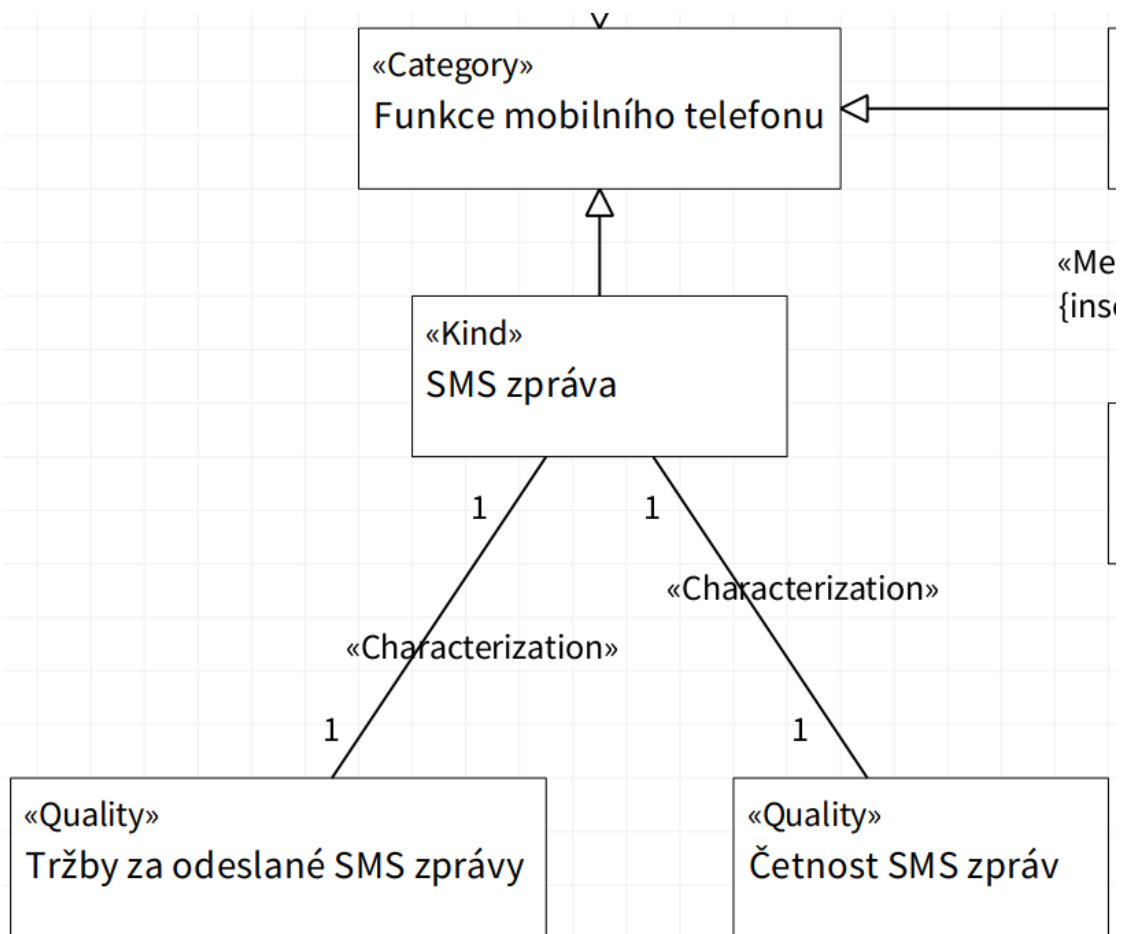
■ Obrázek A.2 Entita SIM karta v modelu trh-mobilnich-siti



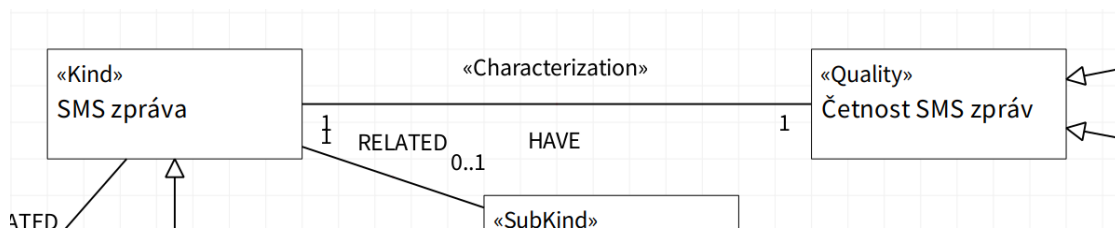
■ Obrázek A.3 Entita telefonní hovor v modelu trh-mobilnich-siti



■ **Obrázek A.4** Entita telefonní hovor v modelu MML-mobilni-telefon

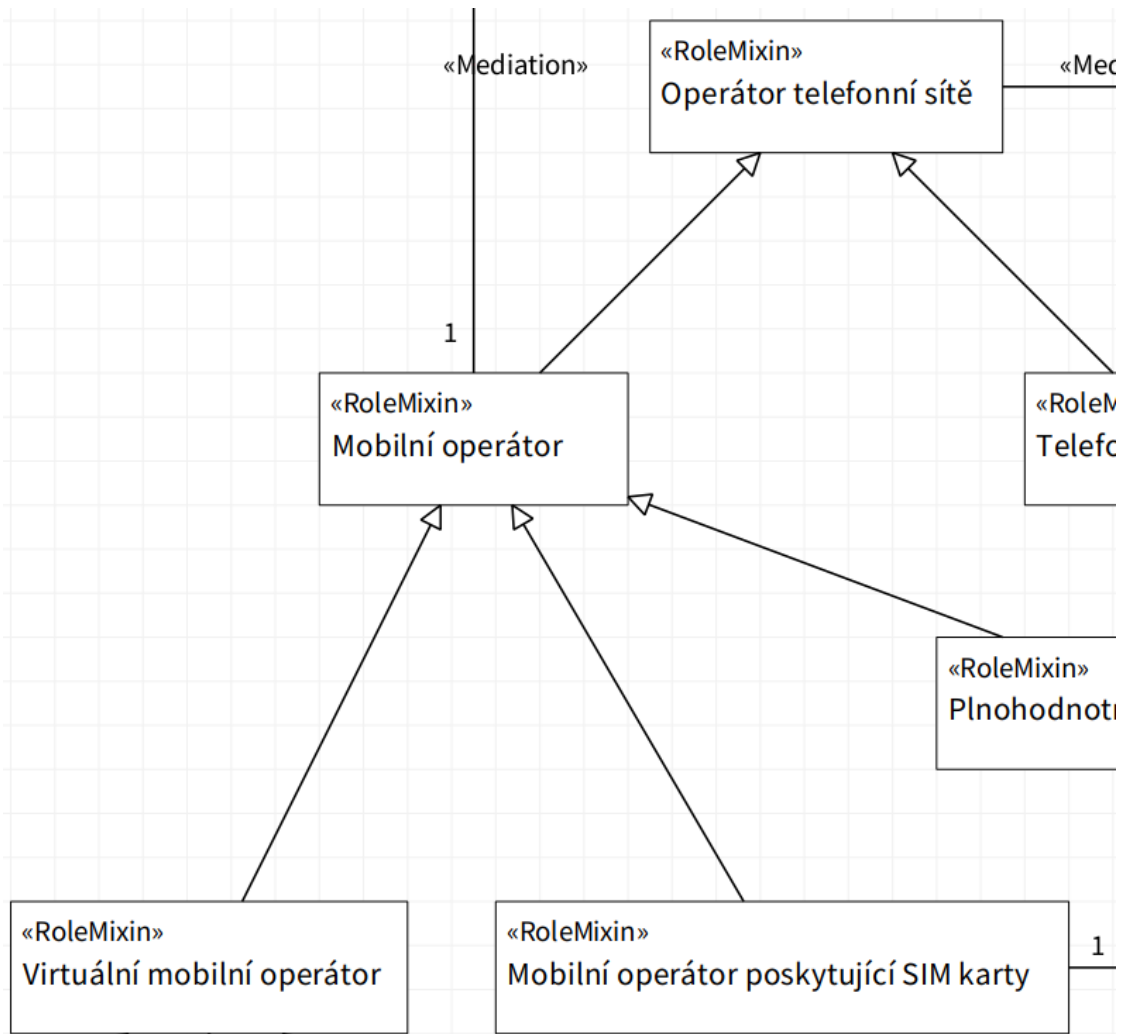


■ Obrázek A.5 Entita SMS zpráva v modelu trh-mobilnich-siti

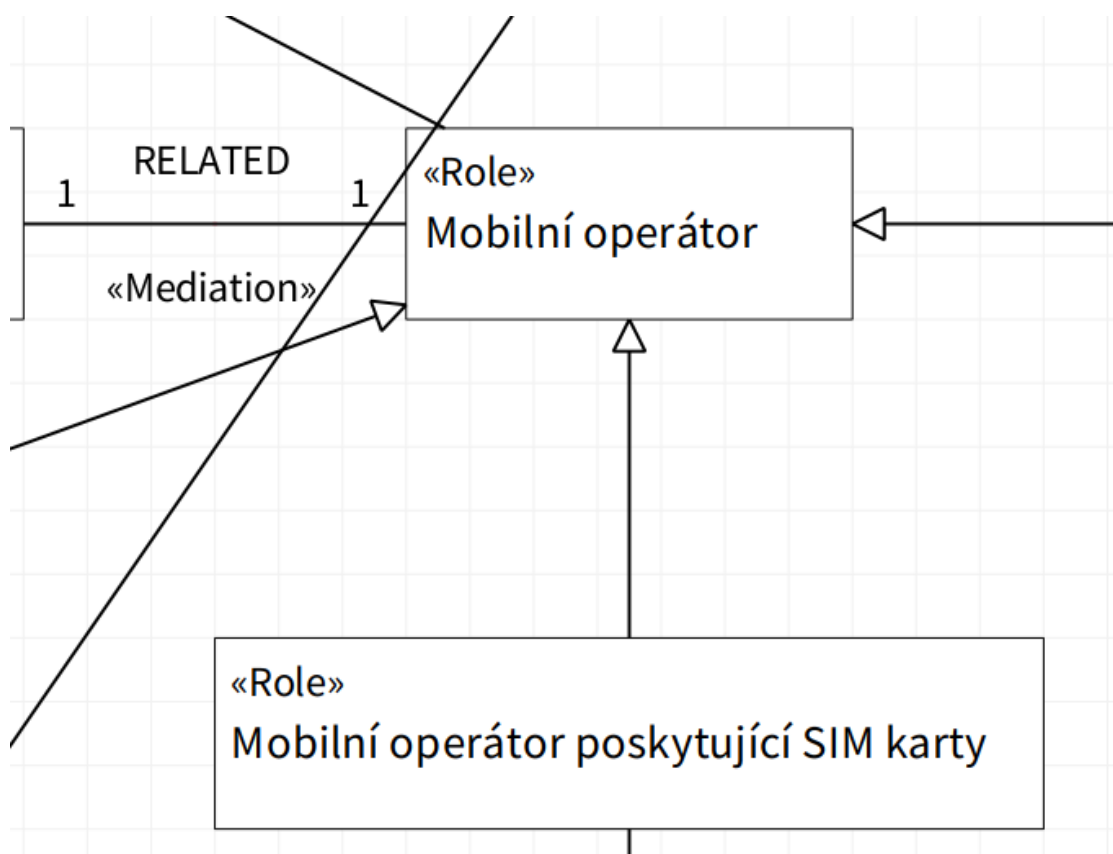


■ Obrázek A.6 Entita SMS zpráva v modelu MML-mobilni-telefon





■ **Obrázek A.7** Entita mobilní operátor v modelu trh-mobilnich-siti



■ **Obrázek A.8** Entita mobilní operátor v modelu MML-mobilni-telefon

# Bibliografie

1. ŠMAJS, Josef a Josef Krob. *Úvod do ontologie: skriptum filoz. fak. Masarykovy univ. 2.*, opr. a rozš. vyd. Brno: Masarykova univerzita, 1994. Dostupné také z: [https://www.phil.muni.cz/fil/eo/skripta/uvod\\_do\\_ontologie.pdf](https://www.phil.muni.cz/fil/eo/skripta/uvod_do_ontologie.pdf).
2. POLI, Roberto. Descriptive, Formal and Formalized Ontologies. In: *Husserl's Logical Investigations Reconsidered*. Dordrecht: Springer Netherlands, 2003, s. 183–210. ISBN 978-94-017-0207-2. Dostupné z DOI: 10.1007/978-94-017-0207-2\_12.
3. CORAZZON, Raul. *Ontology: Its Role in Modern Philosophy* [online] [cit. 2022-04-25]. Dostupné z: <https://www.ontology.co/>.
4. SMITH, Barry. Basic Concepts of Formal Ontology. In: 1998, s. 19–28.
5. CORAZZON, Raul. *Edmund Husserl: Formal Ontology and Transcendental Logic* [online] [cit. 2022-04-25]. Dostupné z: <https://www.ontology.co/husserle.htm>.
6. EPPLER, M.J. Knowledge Communication. In: *Encyclopedia of Knowledge Management, Second Edition*. IGI Global, [b.r.], s. 515–526. Dostupné z DOI: 10.4018/978-1-59904-931-1.ch050.
7. BABOVIC, Zoran; MILUTINOVIC, Veljko. Chapter 2 - Novel System Architectures for Semantic-Based Integration of Sensor Networks. In: HURSON, Ali (ed.). *Connected Computing Environment*. Elsevier, 2013, sv. 90, s. 91–183. Advances in Computers. ISSN 0065-2458. Dostupné z DOI: <https://doi.org/10.1016/B978-0-12-408091-1.00002-6>.
8. HUTH, M.; VISHIK, C.; MASUCCI, R. 8 - From Risk Management to Risk Engineering: Challenges in Future ICT Systems. In: GRIFFOR, Edward (ed.). *Handbook of System Safety and Security*. Boston: Syngress, 2017, s. 131–174. ISBN 978-0-12-803773-7. Dostupné z DOI: <https://doi.org/10.1016/B978-0-12-803773-7.00008-5>.
9. PERGL, Robert. *Motivace a úvod do konceptuálního modelování* [online] [cit. 2021-05-02]. Dostupné z: <https://courses.fit.cvut.cz/BI-KOM/slides/html/lectures-czech/01-intro.htm>.
10. TARI, Furkan. A Review of Emerging Technological Trends in E-Learning. In: *Handbook of Research on Instructional Systems and Technology*. IGI Global, 2008, s. 729–740. Dostupné z DOI: 10.4018/978-1-59904-865-9.ch052.
11. SANTOS, Neide; CAMPOS, Fernanda C.A.; VILLELA, Regina M.M. Braga. Digital Libraries and Ontology. In: *Handbook of Research on Digital Libraries*. IGI Global, 2009, s. 206–215. Dostupné z DOI: 10.4018/978-1-59904-879-6.ch020.
12. SCHERP, Ansgar; SAATHOFF, Carsten; FRANZ, Thomas; STAAB, Steffen. Designing Core Ontologies. *Applied Ontology*. 2011, roč. 6, s. 4. Dostupné z DOI: 10.3233/A0-2011-0096.

13. BIRGER, Andersson; JOHANNESSON, Paul. Ontological Issues Concerning Money. In: *15th International Workshop on Value Modelling and Business Ontologies (VMBO 2021)*. 2021. Dostupné také z: <http://ceur-ws.org/Vol-2835/paper8.pdf>.
14. GUIZZARDI, Giancarlo; AMARAL, Glenda; PORELLO, Daniele; PRINCE SALES, Tiago. A Core Ontology for Economic Exchanges. In: 2020, s. 364–374. ISBN 978-3-030-62521-4. Dostupné také z: <https://philarchive.org/archive/PORACO-6>.
15. AMARAL, Glenda; PRINCE SALES, Tiago; GUIZZARDI, Giancarlo. Modeling Payments and Linked Obligation Settlements. In: 2022. Dostupné také z: [http://www.inf.ufes.br/~gguizzardi/Modeling\\_Payments\\_EEWC2021.pdf](http://www.inf.ufes.br/~gguizzardi/Modeling_Payments_EEWC2021.pdf).
16. MARTÍNEZ FERRANDIS, Ana Ma; PASTOR LÓPEZ, Oscar; GUIZZARDI, Giancarlo. Applying the Principles of an Ontology-Based Approach to a Conceptual Schema of Human Genome. In: NG, Wilfred; STOREY, Veda C.; TRUJILLO, Juan C. (ed.). *Conceptual Modeling*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, s. 471–478. Dostupné také z: <https://link.springer.com/content/pdf/10.1007/978-3-642-41924-9.pdf>.
17. AHMAD, Aakash; BANDARA, Madhushi; FAHMIDEH, Mahdi; PROPER, Henderik; GUIZZARDI, Giancarlo; SOAR, Jeffrey. An Overview of Ontologies and Tool Support for COVID-19 Analytics. In: 2021, s. 1–8. Dostupné z DOI: 10.1109/EDOCW52865.2021.00026.
18. GUIZZARDI, Giancarlo; BENEVIDES, Alessandro; FONSECA, Claudenir; PORELLO, Daniele; ALMEIDA, João; PRINCE SALES, Tiago. UFO: Unified Foundational Ontology. *Applied Ontology*. 2022. Dostupné z DOI: 10.3233/AO-210256.
19. OMG. *Unified Modeling Language* [online]. 2017 [cit. 2022-05-02]. Dostupné z: <https://www.omg.org/spec/UML/2.5.1/PDF>.
20. PERGL, Robert. *OntoUML: Základy, rigidní a anti-rigidní sortály* [online] [cit. 2021-05-02]. Dostupné z: <https://courses.fit.cvut.cz/BI-KOM/slides/html/lectures-czech/02-ontouml-basics-sortals.htm>.
21. PERGL, Robert. *OntoUML: Non-sortály, relace, relátor* [online] [cit. 2021-05-02]. Dostupné z: <https://courses.fit.cvut.cz/BI-KOM/slides/html/lectures-czech/03-ontouml-nonsortals-relations.htm>.
22. PERGL, Robert. *OntoUML: Vztahy Celek-Část, typy agregací; Aspekty* [online] [cit. 2021-05-02]. Dostupné z: <https://courses.fit.cvut.cz/BI-KOM/slides/html/lectures-czech/04-ontouml-aggr-aspects.htm>.
23. CELBOVÁ, Ludmila. Sémantická interoperabilita. *Information*. 2003. ISSN 2078-2489. Dostupné také z: [https://aleph.nkp.cz/F/?func=direct&doc\\_number=000000555&local\\_base=KTD](https://aleph.nkp.cz/F/?func=direct&doc_number=000000555&local_base=KTD).
24. FAIR, GO. *FAIR Principles* [online] [cit. 2022-05-02]. Dostupné z: <https://www.go-fair.org/fair-principles/>.
25. FAIR, GO. *FAIRification Process* [online] [cit. 2022-05-02]. Dostupné z: <https://www.go-fair.org/fair-principles/fairification-process/>.
26. REMMARK, a.s. *Remmark* [online] [cit. 2022-04-23]. Dostupné z: <http://www.remark.cz/>.
27. ODI. *What are open standards for data?* [Online] [cit. 2022-04-23]. Dostupné z: <https://standards.theodi.org/introduction/what-are-open-standards-for-data/>.
28. ODI. *The Data Spectrum* [online] [cit. 2022-04-23]. Dostupné z: <https://theodi.org/about-the-odi/the-data-spectrum/>.
29. ALERC. *An Introduction to Open, Shared and Closed Data* [online] [cit. 2022-04-23]. Dostupné z: [https://www.alerc.org.uk/uploads/7/6/3/3/7633190/an\\_introduction\\_to\\_open\\_shared\\_and\\_closed\\_data.pdf](https://www.alerc.org.uk/uploads/7/6/3/3/7633190/an_introduction_to_open_shared_and_closed_data.pdf).

30. ODI. *What is open data?* [Online]. 2016 [cit. 2022-04-23]. Dostupné z: <https://theodi.org/article/what-is-open-data/>.
31. ODI. *Types of open standards for data* [online]. 2016 [cit. 2022-04-24]. Dostupné z: <https://standards.theodi.org/introduction/types-of-open-standards-for-data/>.
32. CVUT, FIT. *OpenPonk modeling platform* [online]. 2022 [cit. 2022-05-09]. Dostupné z: <https://openponk.org/>.
33. CCMI. *OpenPonk* [online]. 2016 [cit. 2022-05-09]. Dostupné z: <https://ccmi.fit.cvut.cz/nastroje/openponk/>.
34. PHARO. *About Pharo* [online] [cit. 2022-05-09]. Dostupné z: <https://pharo.org/about>.
35. ČTÚ. *Všeobecné oprávnění č. VO-S/2/07.2005-10*. Sokolovská 219, Praha 9, 2005. Dostupné také z: [https://www.ctu.cz/1/download/0patreni%20obecne%20povahy/VO\\_S\\_02\\_07\\_2005\\_10.pdf](https://www.ctu.cz/1/download/0patreni%20obecne%20povahy/VO_S_02_07_2005_10.pdf).
36. *Všeobecné oprávnění č. VO-S/2/07.2007-12*. Sokolovská 219, Praha 9, 2007. Dostupné také z: [https://www.ctu.cz/1/download/00P/Rok\\_2007/VO-S\\_02\\_07\\_2007\\_12.pdf](https://www.ctu.cz/1/download/00P/Rok_2007/VO-S_02_07_2007_12.pdf).
37. *Všeobecné oprávnění č. VO-S/2/09.2008-11*. Sokolovská 219, Praha 9, 2008. Dostupné také z: [https://www.ctu.cz/cs/download/oop/rok\\_2008/vo-s\\_02-09\\_2008-11.pdf](https://www.ctu.cz/cs/download/oop/rok_2008/vo-s_02-09_2008-11.pdf).



# Obsah přiloženého média

readme.txt .....	stručný popis obsahu média
src	
_ modely .....	Konceptuální modely
_ klicove-domeny .....	modely klíčových domén
_ sablony .....	modely šablon
_ text .....	zdrojová forma práce ve formátu L <sup>A</sup> T <sub>E</sub> X
bp-sirvacla-2022.pdf .....	text práce ve formátu PDF