



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



MINISTERSTVO ZDRAVOTNICTVÍ
ČESKÉ REPUBLIKY

Projekt „Strategické řízení rozvoje elektronického zdravotnictví v resortu MZ“,
registrační číslo CZ.03.4.74/0.0/0.0/15_025/0006212,
je spolufinancován Evropskou unií.

ZPRACOVÁNÍ METODIK TVORBY NÁSTROJŮ PRO IMPLEMENTACI NÁRODNÍ STRATEGIE ELEKTRONICKÉHO ZDRAVOTNICTVÍ

**WP.F - Metodika správy terminologie
elektronického zdravotnictví
P10 – Soubor metodik pro správu terminologie
elektronického zdravotnictví**

Projekt:	Strategické řízení rozvoje elektronického zdravotnictví v resortu MZ, registrační číslo CZ.03.4.74/0.0/0.0/15_025/0006212 je spolufinancován Evropskou unií		
Klíčová aktivita:	Zpracování metodik tvorby nástrojů pro implementaci Národní strategie elektronického zdravotnictví		
Datum:	12. 4. 2019	Stav:	final
Balík práce:	WP.F - Metodika správy terminologie elektronického zdravotnictví		
Název produktu:	P10 – Soubor metodik pro správu terminologie elektronického zdravotnictví		
Autor:			
Dodavatel:	Asseco Central Europe, a.s.		
Zákazník:	Ministerstvo zdravotnictví ČR		
Číslo dokumentu:		Verze:	1.0

Schválení

Jméno	Podpis	Pozice	Datum
Ing. Martin Zeman		Sponzor projektu	12. 4. 2019
MUDr. Miroslav Zvolský		Hlavní uživatel	12. 4. 2019
Ing. Hynek Kružík		Hlavní dodavatel	12. 4. 2019
Ing. Martina Hábová, Ph.D.		Manažer projektu	12. 4. 2019

Distribuční seznam

Jméno	Subjekt / organizační jednotka	Datum	Verze
Ing. Martin Zeman	Ministerstvo zdravotnictví ČR	12. 4. 2019	1.0
Ing. Jiří Borej	Ministerstvo zdravotnictví ČR	12. 4. 2019	1.0
Ing. Eliška Urbancová	Ministerstvo zdravotnictví ČR	12. 4. 2019	1.0

Obsah

1	WPF: Struktura a popis produktů	5
2	Úvod	5
2.1	Zadání pro produkt P10.....	5
2.1.1	Zahrnutá oblast	5
2.2	Struktura výstupu	5
2.3	Vazba výstupu na další výstupy balíku práce WPF	6
2.4	Přístup k řešení.....	6
1.1	Vybrané použité pojmy	6
1.2	Seznam zkratk	7
3	Metodologické principy tvorby a rozšiřování sémantického slovníku	9
3.1	Metodické principy pro vytváření glosáře.....	9
3.1.1	Orientace na koncepty	9
3.1.2	Preferovaný název pojmu.....	9
3.1.3	Definice pojmu	10
3.1.4	Vyjádření hierarchických sémantických vazeb	10
3.1.5	Související pojmy	10
3.1.6	Příklady použití pojmu.....	10
3.2	Metodické principy pro vytváření ontologického modelu	10
3.2.1	Diagram binárních vazeb	10
3.2.2	Komplexní příklady s využitím binárních vazeb.....	11
3.2.3	Ontologické typy pro pojmy a vztahy	11
3.2.4	Ontologický model	11
4	Praktická typologie slovníků využitá v metodice.....	11
4.1	Jednoduchý slovník	12
4.2	Formalizovaný slovník	12
4.3	Kompletní sémantický slovník.....	12
5	Vlastní metodika pro tvorbu/rozšiřování sémantického slovníku	12
5.1	Kontext	12
5.2	Úvod	12
5.3	Role participující na agendě	13
5.3.1	Člen týmu projektu/agendy	13
5.3.2	Mluvčí týmu projektu/agendy.....	13
5.3.3	Slovníkový specialista	13

5.3.4	Ontologický inženýr.....	13
5.3.5	Správce Sémantického modelu	13
5.4	Základní popis a struktura agendy	13
5.5	Popis jednotlivých procesů agendy	15
5.5.1	Specifikace požadavků na Slovník a rozhodnutí o typu rozšiřujícího slovníku.	15
5.5.2	Sběr pojmů	15
5.5.3	Nominace mluvčího a vytvoření konsensu terminologie týmu	17
5.5.4	Import nových pojmů do RDF úložiště	19
5.5.5	Kodifikace atributů pojmů a vytvoření (validace) sémantických vazeb	20
5.5.6	Vytvoření ontologického modelu	21
5.5.7	Publikování nové verze Sémantického slovníku	22

1 WPF: Struktura a popis produktů

Balík práce WPF obsahuje následující produkty:

- P9 – Rešerše standardů a metodologií pro správu odborných slovníků
- **P10 – Soubor metodik pro správu terminologie elektronického zdravotnictví**
- P11 – Základní sémantický slovník NCEZ
- P12 – Rozšířený sémantický slovník NCEZ

2 Úvod

2.1 Zadání pro produkt P10

Produkt je zamýšlen ke konkrétnímu praktickému využití pro vznikající Národní centrum elektronického zdravotnictví. Má být souborem metodik pro vytváření, správu a aktualizaci terminologie všech konkrétních projektů, výstupů a služeb NCEZ v podobě Sémantického slovníku NCEZ. Předpokládá, že NCEZ bude autoritou, která bude Sémantický slovník NCEZ spravovat a využívat pro své výstupy a služby (všechny projekty, výstupy a služby NCEZ budou moci být referencovány na tento Slovník) a spolu s nimi ho poskytovat veřejnosti. Metodiky budou tedy přímo použitelné v kontextu vznikajícího NCEZ a aplikovatelné ve všech jeho projektech, počínaje tímto projektem.

Metodiky musí zaručovat, aby jimi vytvářený Sémantický slovník NCEZ zajišťoval syntaktickou a sémantickou interoperabilitu přes všechna logická schémata slovníku, tedy, aby všechna jednotlivá, doménově specifická, logická schémata slovníku byla plně propojena na explicitní základní sémantický model společný pro Sémantický slovník NCEZ jako celek.

Výstup bude ve formátu dokumentu MS Office.

2.1.1 Zahrnutá oblast

Soubor metodik bude uveden i relevantními obecnými metodologickými principy pro tvorbu slovníků.

2.2 Struktura výstupu

Tento dokument obsahuje výstup P10 **Soubor metodik pro správu terminologie elektronického zdravotnictví** balíku práce **WPF Správa terminologie**. Navzdory velmi obecnému názvu má výstup mít ryze praktický význam v konkrétním kontextu vznikajícího Národního centra elektronického zdravotnictví, kdy podle zadání má být metodikou, podle které lze prakticky postupovat při vytváření a správě Sémantického slovníku NCEZ.

Dokument obsahuje v první části popis obecných metodologických principů pro tvorbu slovníku, následně praktickou typologii slovníků, která je pak použita ve vlastní metodice, která tvoří poslední a hlavní část výstupu.

2.3 Vazba výstupu na další výstupy balíku práce WPF

Výstup využívá Rešerši standardů a metodologií pro správu odborných slovníků (Prodekt P9) a je založen na zkušenostech s vytvářením Základního slovníku NCEZ (Produkt P11). Metodika byla použita pro dodání Rozšiřujícího slovníku NCEZ (Produkt P12).

2.4 Přístup k řešení

Sémantický slovník NCEZ je (resp. bude), ve smyslu správy terminologie, dvouvrstevná hierarchická struktura, v které první vrstvu tvoří vlastní Základní slovník NCEZ a druhou vrstvu pak slovníky, které bude NCEZ vytvářet v rámci jednotlivých dalších projektů, a které budou sloužit jako slovníky jednotlivých agend a služeb NCEZ (dále agendové slovníky). Agendové slovníky mají pokrývat specifickou terminologii své domény, k zachycení sémantických vazeb¹ pak využívat koncepty a vztahy definované základním slovníkem NCEZ, popřípadě Základním obecným slovníkem pojmů veřejné správy.

Tento výstup vychází z předpokladu, že Sémantický slovník NCEZ tak, jak je popsán výše, bude vznikat postupně vždy v rámci nějakého konkrétního projektu, kterým bude ustavena konkrétní agenda NCEZ resp. implementována konkrétní centrální služba spravovaná NCEZ, případně jakýkoliv jiný výstup NCEZ určený veřejnosti. Integrální součástí produktu všech takových projektů má být buď vytvoření nového agendového slovníku pokrývající specifickou terminologii materiálních výstupů projektu a/nebo rozšíření již existujících slovníků (včetně Základního slovníku NCEZ). **Proto vytvořená metodika pro vytváření a rozšiřování Sémantického slovníku NCEZ nutně předpokládá kontext konkrétního projektu.**

Správu Sémantického slovníku NCEZ jako takto popsané struktury slovníků pak chápeme jako agentu, jež **sama musí být ustavena prostřednictvím konkrétního projektu.** Konkrétní metodika pro správu Sémantického slovníku pak nutně musí vycházet ze struktury agendy ustavené tímto projektem a být de facto produktem projektu, kterým se ustaví tato agenda.

Při vytváření agentových slovníků v rámci konkrétních projektů NCEZ budou nutně vznikat podněty na aktualizaci základního slovníku ve smyslu rozšíření jeho glosáře o další obecné koncepty, zpřesnění definic stávajících konceptů, případně úpravy ontologického modelu, jež spolu s podněty od uživatelů produktů NCEZ jsou vstupy pro agendu správy Sémantického slovníku NCEZ.

Proto jsme pokládali za účelné vytvořit jako výstup tohoto projektu především konkrétní metodiku pro vytváření resp. rozšiřování sémantického slovníku NCEZ, přičemž metodika pro jeho správu v kontextu uvedeném výše je v tomto výstupu chápána jako analogie této metodiky, jež musí být tak jako tak konkretizována v projektu ustavení agendy správy Sémantického slovníku NCEZ.

Vybrané použité pojmy

Agenda:

Struktura navzájem korespondujících **procesů**, jejichž smyslem je dodat určitý komplexní výstup (hodnotu).

HL7 (Health level 7):

¹ S různou mírou přesnosti a formalizace podle typu slovníku (viz metodiku)

Mezinárodní standard pro přenos informací mezi informačními systémy zejména poskytovatelů zdravotních služeb²

Projekt:

Projekt ve smyslu metodiky projektového řízení PRINCE2

Výstup projektu:

Materiální výstup projektu (produkt projektu ve smyslu PRINCE2).

Glosář:

Vlastní sada konceptů (pojmů) sémantického slovníku, tedy názvů pojmů, definic a dalších atributů sloužících pro porozumění významu, vytvořených a spravovaných podle definovaných pravidel.

Sémantická vazba:

Vztah (relace) mezi jednotlivými koncepty Slovníku nesoucí kontextový význam daného typu.

Technologická infrastruktura Sémantického slovníku:

Konkrétní zvolený soubor IT nástrojů použitý pro vytváření, udržování, sdílení, rozšiřování a aktualizaci Sémantického modelu.

Příklady technologické infrastruktury použité v Metodice:

- Nástroj na sběr nových konceptů : sdílený dokument **Google sheet**
- Nástroj na import obsahu Google sheet do datového úložiště sémantického modelu : vytvořený importní skript
- RDF datové úložiště sémantického slovníku : **RDF4J** (frameworku VocBench)
- Prostředí pro udržování sémantického modelu : **VocBench**
- Jazyk pro modelování sémantických vazeb : **UML** respektive **onto -UML**
- Prostředí pro modelování sémantických vazeb : **Modelio** s implementací onto-UML
- Nástroj pro obousměrnou synchronizaci Vocbench a Modelio
- Prostředí pro publikaci Slovníku na webu : **Skosmos**

Seznam zkratk

NCEZ:	Národní centrum elektronického zdravotnictví
RDFS :	Resource Description Framework Scheme
SKOS :	Simple knowledge organization system

HL7: Health Level 7
RIM (HL7): Reference information model HL7
UFO: Unified Foundational Ontology

3 Metodologické principy tvorby a rozšiřování sémantického slovníku

Sémantický slovník NCeZ je vhodné budovat tak, aby mohl plně využívat Sémantický slovník veřejné správy. Pro dodržení syntaktické a sémantické kompatibility mezi oběma slovníky je nutné dodržovat zejména 3 hlavní principy [1]:

- **globální identita:**
Každý odborný pojem slovníku má přiřazený globální identifikátor, který musí umožnit snadné získání informace o významu daného pojmu. Při zavedení nového odborného pojmu je vždy vytvořen nový globální identifikátor.
- **ukotvení významu:**
Každý odborný pojem má být doplněn strojově čitelnými vlastnostmi a vztahy, které ukotvují jeho význam. V případech, kdy je význam pojmu definován ve veřejném normativním dokumentu, je pojem doplněn rovněž odkazem na tento dokument (zákon, vyhláška, jiný předpis) nebo jeho část.
- **princip určené zodpovědnosti:**
Každá ucelená a konzistentní sada pojmů, jejich vlastností a vztahů je popsána v rámci jednoho slovníku. Za správu každého slovníku, jeho aktuálnost a konzistenci je zodpovědná právě jedna agenda NCeZ.

3.1 Metodické principy pro vytváření glosáře

3.1.1 Orientace na koncepty

Základní jednotkou odborného pojmosloví by měl být pojem, který reprezentuje právě jeden konkrétní význam (koncept). Slovník tedy musí explicitně rozlišovat mezi pojmem a jeho lexikálními reprezentacemi [2], které budeme v následujícím textu nazývat jako „termín“. Jako příklad uvažujme, že termíny “nemoc”, “onemocnění”, “NM” jsou lexikálními variantami se stejným významem vyjadřujícím nedostatek zdraví jednotlivce. Tyto 3 termíny je potřeba reprezentovat jedním pojmem, tedy identifikátorem (dle principu globální identity definované výše), který reprezentuje jeho význam a slučuje tyto lexikální varianty (synonyma).

3.1.2 Preferovaný název pojmu

Pojem v rámci jednoho slovníku identifikuje jeden konkrétní význam, no může odpovídat vícerym termínům (lexikálním formám). Pro dodržení konzistence slovníku je potřeba jeden z těchto termínů vybrat jako primární a zbylé termíny jako alternativní. Primární termín budeme dále pojmenovávat *preferovaný název pojmu* a zbylé termíny *alternativní názvy pojmu*, nebo synonyma. Výběr konkrétního termínu je nutné zakládat na specifikaci požadavků slovníku a to zejména dle očekávaného použití slovníku a cílové skupiny uživatelů.

Příklad: Uvažujme slovník určený pro lékaře, do kterého chceme začlenit termíny “hypertenze” a “vysoký krevní tlak” reprezentující stejný koncept. Jelikož jsou cílovou skupinou uživatelů lékaři, je vhodné zvolit jako preferovaný název pojmu “hypertenze” a jeho alternativní název “vysoký krevní tlak”.

3.1.3 Definice pojmu

Každý pojem by měl mít právě jednu definici. Častým problémem je existence více definic jednoho pojmu, které jsou vzájemně nekonzistentní, nebo popisují pojem ze dvou rozličných úhlů pohledů. Řešením je syntéza těchto definic do jedné definice, ponechání jedné z definic dle posouzení na základě cíle a rozsahu slovníku, nebo vytvoření nového pojmu pro každou definici zvlášť.

Příklad: Uvažujme dvě definice pojmu pacient -- “Fyzická osoba, která vyžaduje lékařské ošetření” a “Fyzická osoba, které jsou poskytovány zdravotní služby”. Řešení konfliktu můžeme vyřešit např. vytvořením dvou separátních pojmů “osoba vyžadující lékařské ošetření” a “pacient”.

3.1.4 Vyjádření hierarchických sémantických vazeb

Relace obecnější – užší (skos:broader resp. skos:narover ve standardu SKOS) umožňuje zachytit kontext pojmů glosáře vytvářením neformálních hierarchií.

3.1.5 Související pojmy

Asociační relace mezi souvisejícími pojmy glosáře umožňuje pojmy organizovat do asociačních sítí bez nutnosti specifikovat jméno a význam sémantických vazeb mezi pojmy. Případně je možné využít specifických pojmenovaných asociačních vazeb, což do určité míry nahrazuje vytváření diagramů binárních vazeb.

3.1.6 Příklady použití pojmu

Příklady použití pojmu se popisují jenom pro abstraktní pojmy (např. “zdravotní služba”), nikoli pojmy konkrétní (např. “ministerstvo zdravotnictví”). Slouží zejména pro validaci pojmu, nejsou tedy nutné u jednoduchých slovníků (viz dále). Vhodný počet příkladů záleží od komplexnosti daného pojmu. Příklady jsou důležité zejména v případech, kdy v slovníku nejsou specifičtější pojmy (např. “zdravotnická dopravní služba” pro pojem “zdravotní služba”). V případech, kdy definice používá další pojmy, které nepatří do slovníku, nebo jejich význam v kontextu definice nemusí být zřejmý, je potřeba příklad popsat v celé šíře definice.

Častou chybou příkladů použití pojmu je uvedení specializace pojmu místo uvedení instance (tedy příkladu) pojmu.

Příklad: Uvažujme pojem “zdravotnická dokumentace” s definicí “Zdravotnická dokumentace je souborem informací o zdravotních službách vztahujících se k pacientovi..”. Nevhodným příkladem, kde je použita specializace místo instance pojmu, je “zpráva z ambulantního vyšetření” jelikož to je příklad pojmu “typ zdravotnické dokumentace”. Korektním příkladem je tedy “zpráva ambulantního vyšetření pacienta Franty Vomáčky ze dne 1.2.2019”.

3.2 Metodické principy pro vytváření ontologického modelu

3.2.1 Diagram binárních vazeb

Diagram binárních vazeb slouží na zpřesnění asociačních vazeb pro účely vytváření ontologického modelu. U tohoto zpřesnění je vhodné specifikovat jméno a kardinalitu vazby. Je možné ho specifikovat pomocí UML diagramu tříd, nebo pomocí tabulky, viz. následující příklad.

Příklad: Uvažujme pojmy “poskytovatel zdravotních služeb” a související pojem “typ zdravotní služby”. Zpřesnění tohoto vztahu je vazba “poskytuje” a platí, že “poskytovatel zdravotní služby” “poskytuje” 1..N “typů zdravotních služeb” jak je naznačeno v následující tabulce.

Vztah	zdrojový pojem	cílový pojem	max. kardinalita zdrojového pojmu	max. kardinalita cílového pojmu
poskytuje	poskytovatel zdravotních služeb	typ zdravotní služby	0..N	1..N

3.2.2 Komplexní příklady s využitím binárních vazeb

Pro validaci pojmosloví jsou důležité komplexní příklady s využitím binárních vazeb. Důraz je tady kladen zejména na vytvoření co největší souvislé komponenty spojující koncepty dle jednoho scénáře použití slovníku. Dalším kritériem je různorodost příkladů z hlediska použití pojmů.

K úkolu je vhodný zejména UML objektový diagram, který využívá již definovaných binárních vazeb. Alternativními nástroji pro vyjádření příkladu jsou mindmapovací nástroje nebo volný text.

3.2.3 Ontologické typy pro pojmy a vztahy

Cílem tohoto kroku je určení základních ontologických typů pro pojmy a ontologicky významné vazby mezi pojmy. Tyto typy definuje jako pojmy Základního slovníku veřejné správy (např. "Objekt", "Role", "Typ vztahu", "Typ vlastnosti") a jsou již přímou součástí základní ontologie UFO.

3.2.4 Ontologický model

Ontologický model využívá výstupy předcházejících kroků pro validaci pojmů a vazeb. Výstupem je ontologicky korektní hierarchie pojmů a formalizace zbylých vztahů mezi pojmy. Části modelu jsou organizované do diagramů využívajících jazyk OntoUML pro zachycení ontologických vlastností pojmů a vazeb.

4 Praktická typologie slovníků využitá v metodice

Vytváření kompletního sémantického slovníku je velmi náročné na čas a zdroje. Předpokládáme, že pro vytváření agentových slovníků v rámci konkrétních projektů obvykle ve skutečnosti nebude příliš velký prostor jak ve smyslu harmonogramu projektu, tak zdrojů projektu. Proto jsme navrhli následující typologii, kterou používáme pro účely metodiky, a která dovoluje již na začátku projektu přizpůsobit metodiku specifickým požadavkům projektu na míru formalizace a validace sémantických pojmů vytvářeného slovníku, a tím přesněji odhadnout požadavky na zdroje projektu alokované na vytvoření/rozšíření Sémantického slovníku NCeZ.

Typologie rozeznává tři typy slovníků z hlediska nároků na zdroje projektu:

4.1 Jednoduchý slovník

Jednoduchý slovník odpovídá standardu SKOS a zachycuje pouze neformální hierarchické a asociační vazby. Nevyžaduje účast role „Ontologický inženýr“ a účast role „Slovníkový specialista“ je jen velmi omezená.

4.2 Formalizovaný slovník

Formalizovaný slovník odpovídá standardu SKOS rozšířenému o pojmenované specifické asociační vazby, případně diagramy binárních vazeb a případně využívá (podle míry kompetence role „Slovníkový specialista“) ontologické typy. Zaručuje rozumnou míru konsistence pojmů a vztahů. Nevyžaduje účast role „Ontologický inženýr“.

4.3 Kompletní sémantický slovník

Kompletní sémantický slovník obsahuje kompletní ontologický model. Nutně vyžaduje účast role „Ontologický inženýr“.

5 Vlastní metodika pro tvorbu/rozšiřování sémantického slovníku

5.1 Kontext

Metodika pro vytváření/rozšiřování Sémantického slovníku NCEZ je koncipována pro použití výhradně uvnitř konkrétních projektů NCEZ, které budou dodávat buď konkrétní materiální výstupy nebo implementovat konkrétní agendy případně služby NCEZ. Metodika tedy předpokládá nutně existenci projektového týmu, který dodává výstupy daného projektu, kde jednotliví členové spolu v rámci projektu komunikují a v rámci projektu sdílejí, popřípadě upřesňují specifické koncepty (pojmy), relevantní pro dodání výstupů projektu, pro předání výstupů jejich uživatelům a pro další komunikaci s nimi po ukončení projektu.

5.2 Úvod

Metodika popisuje vytváření sémantického slovníku jako **agendu** ve smyslu definice uvedené výše. Jde o standardní procesní popis postupu, jak dodat zejména agendový sémantický slovník, popřípadě rozšířit již existující agendový slovník, popřípadě rozšířit Základní slovník NCEZ jako produkt konkrétního projektu NCEZ. Metodika předpokládá existenci **Základního slovníku NCEZ**.

Metodika je založena na definovaných **rolích**, je strukturována na jednotlivé **procesy**, které dohromady vytvářejí **agendu**.

5.3 Role participující na agendě

5.3.1 Člen týmu projektu/agendy

Pracovník, který se podílí na dodání výstupů nebo na řízení daného projektu NCEZ resp. je součástí týmu zodpovědného za konkrétní agendu NCEZ. Má tzv. „doménovou znalost“, tj. rozumí materii projektu/agendy, zná autoritativní zdroje doménové znalosti a přirozeně ovládá terminologii domény, v které projekt/agenda dodává výstupy. Není expertem na formalizování terminologie a ontologické modelování. V této roli se mohou zúčastnit tvorby/rozšiřování Sémantického slovníku všechny osoby patřící do týmu projektu, může jich být tedy celá řada.

5.3.2 Mluvčí týmu projektu/agendy

Člen týmu projektu, který je týmem vybrán, aby zprostředkoval doménovou znalost specialistům při vytváření Sémantického slovníku. Měl by to být (pokud je dostupný) ten člen týmu, který nejlépe rozumí materii projektu.

5.3.3 Slovníkový specialista

Pracovník se specializovanou znalostí datové reprezentace konceptů slovníku a sémantických vazeb. Jedná se především o RDFS datový model³ pro tzv. **ontologie** (RDF slovníky), a na něm postavený SKOS (Simple Knowledge Organization System) model⁴ pro organizaci informací o konceptech a jejich sémantických vazbách, v konkrétní zvolené implementaci⁵ a je schopen zajistit import nových konceptů pracovního glosáře databáze RDF. Slovníkový specialista dále musí mít znalost o abstraktních obecných konceptech již obsažených v Základním slovníku NCEZ a je schopen definovat binární sémantické vazby v UML v konkrétní zvolené aplikaci (Modelio). Slovníkový specialista získá svojí kompetenci školením (rozsah školení cca 8-16 hodin). NCEZ by měla disponovat alespoň jedním pracovníkem s touto kompetencí⁶. Slovníkový datový specialista postupně přebírá část kompetencí od ontologického inženýra (viz dále) v onto-UML (UFO).

5.3.4 Ontologický inženýr

Externí konsultant, expert disponující hlubokou znalostí teorie strojové reprezentace znalostí, sémantických webů a vytváření a správy ontologií, který ovládá Unified Foundational Ontology (UFO) a jazyk onto-UML. Je zodpovědný za vytvoření a aktualizace ontologického modelu Sémantického slovníku NCEZ.

5.3.5 Správce Sémantického modelu

Osoba zodpovědná za správu datového úložiště a webového portálu sémantického modelu a za technické operace spojené s rozšiřováním sémantického modelu podle této metodiky.

5.4 Základní popis a struktura agendy

Agenda začíná specifikací požadavků na Slovník, určením typu Slovníku a specifikací proudu projektu s požadavkem na rozšíření Slovníku. Následuje sběr konceptů Slovníku. Procesy sběru pojmů probíhají paralelně a

³ Pro info viz https://en.wikipedia.org/wiki/RDF_Schema

⁴ Pro info viz https://en.wikipedia.org/wiki/Simple_Knowledge_Organization_System

⁵ VocBench a RDF4J

jsou závislé na postupu prací na projektu – na dodávání draftů výstupů projektu. Výstupem procesů sběru pojmů je jeden sdílený dokument (např. Google sheet) naplněný novými koncepty a jejich definicemi.

Následně tým uspořádá setkání všech, kteří nové koncepty vytvářeli, na kterém je v diskusi konsensuálně přijat definitivní soubor konceptů (pojmů), jejich definice a případně příklady použití a základní neformální sémantické vazby. Tým zároveň zvolí jednoho z pracovníků, kteří nové koncepty vytvářeli, jako svého mluvčího pro další práci se slovníkovým specialistou.

Slovníkový specialista vytvořenou strukturu pojmů, definic a vazeb importuje do datového úložiště Sémantického modelu a dále již pracuje společně s mluvčím týmu ve VocBench a Modelio. Společně vytvoří případně nové terminologické skupiny (skos:collection), definitivně kodifikují sémantické vazby nových pojmů navzájem a identifikují případné vazby na již existující pojmy Sémantického slovníku. Následně se odehraje série schůzek slovníkového specialisty a mluvčího týmu s externím konsultantem – ontologickým inženýrem, na které konsultant validuje navržené sémantických vazeb a vytvoří ontologický model kompatibilní s UFO.

Poté je vytvořena a publikována nová verze Sémantického slovníku.

Agenda vytváření/rozšíření sémantického slovníku o relevantní pojmy projektu se skládá z těchto procesů:

- 1) Specifikace požadavků na Slovník projektu a určení produktů projektu rozšiřujících Slovník projektu.
- 2) Sběr pojmů
- 3) Nominace mluvčího a vytvoření konsensu terminologie týmu
- 4) Import nových pojmů do úložiště Slovníku projektu
- 5) Kodifikace atributů pojmů a neformálních sémantických vazeb
- 6) Vytvoření ontologického modelu
- 7) Publikování nové verze Sémantického modelu

5.5 Popis jednotlivých procesů agendy

5.5.1 Specifikace požadavků na Slovník a rozhodnutí o typu rozšiřujícího slovníku.

5.5.1.1 Popis procesu:

Proces je součástí aktivit etapy nastavení projektu, konkrétně popisu produktu projektu. Produktem projektu, který je zde specifikován, je Slovník projektu. Slovník je jako produkt specifikován jako poslední produkt projektu, tedy až poté, co jsou popsány všechny produkty projektu, zpracovaný produktový rozpad včetně specifikace jednotlivých meziproduktů a product flow diagram. Procesu se účastní Projektový manažer, Hlavní uživatel a Slovníkový specialista. Je nutné, aby se Slovníkový specialista seznámil s výše vyjmenovanými manažerskými produkty Dokumentace nastavení projektu. Poté jsou na jediném setkání vyjmenovaných rolí specifikovány požadavky na Slovník a určen jeho typ. Na stejném setkání jsou identifikovány produkty (meziprodukty) projektu vhodné (nutné) pro rozšíření terminologie. Pro výstup je rozhodující autorizace hlavním uživatelem. Výjimečně lze setkání opakovat, pokud se na prvním setkání nepodaří dosáhnout autorizace výstupů Hlavním uživatelem.

5.5.1.2 Hodnoty atributů procesu:

Majitel procesu: Projektový manažer

Začátek procesu: Poté, co je dokončený popis produktů projektu a product flow diagram (s výjimkou Slovníku).

Vstupy: Popis produktů projektu, Product flow diagram

Aktivita procesu:

1. Nastudování vstupů Slovníkovým specialistou
2. Setkání Projektového manažera, Hlavního uživatele a Slovníkového specialisty
 - a. Určení typu Slovníku
 - b. Identifikace produktů pro rozšíření Slovníku

Konec procesu: Výstupy autorizován hlavním uživatelem projektu

Výstupy: Určený typ Slovníku

Aktualizována dokumentace nastavení projektu

(Popis produktu, Product flow diagram, Registr kvality)

5.5.2 Sběr pojmů

Jde tolik paralelně probíhajících procesů, kolik členů týmu se podílí na vytváření terminologie projektu, resp. kolik je v rámci projektu dodáváno produktů, které byly specifikovány pro rozšíření terminologie v etapě nastavení. Všichni pracují současně s jedním on line sdíleným dokumentem (google sheet) s danou strukturou polí.

5.5.2.1 Popis procesu:

Proces začíná dokončením (dopsáním) dokumentu draftu výstupu, před jeho odevzdáním. Zúčastní se ho osoba zodpovědná za dodání draftu výstupu, případně další pracovník, který se na dodání draftu podílel a je pověřen sběrem pojmů. Dokument je čten od začátku do konce a jsou postupně vybírány vhodné pojmy a zanášeny do sdíleného dokumentu pro sběr pojmů (google sheetu), avšak jen v případě, že se již nevyskytují v Sémantickém slovníku NCEZ ani v Google sheetu. Vybraný pojem je označen v draftu **tučně kursivou**. Do google sheetu jsou

zaneseny jen názvy pojmů a zaznamenán autor – tedy ten, kdo pojem do sešitu zanesl, a pracovní označení dokumentu draftu, z kterého byl pojem získán. Po zpracování celého dokumentu draftu jsou nové pojmy opatřeny definicemi. Pokud definici následně vytváří jiný pracovník než ten, kdo pojem do sešitu zanesl, stává se on autorem. Součástí zpracování definice je i vložení odkazu na zdroj definice. Proces končí zpracováním všech nových pojmů v google sheetu. Přesný postup, tedy, kdo se bude podílet na zanášení pojmů do google sheetu a kdo se bude podílet na vytváření definic a další podrobnosti si dohodnou mezi sebou ti, kdo dokument draftu zpracovávají.

Poznámka: Nejde o to zanést do google sheetu co nejvíce pojmů, ale vybrat jen ty, jejichž význam není přirozeně „všeobecně známý“, tedy pojmy speciální pro danou doménu, a dále ty pojmy, které jsou sice známé, ale u kterých je pocíťováno riziko různých výkladů. Počet nových pojmů by neměl převýšit tři pojmy na jednu stránku textu dokumentu draftu výstupu, **spíše méně**, záleží ovšem na konkrétním kontextu.

5.5.2.2 Hodnoty atributů procesu:

Majitel procesu: Pracovník zodpovědný za dodání draftu výstupu případně pracovník, který se na dodání draftu výstupu podílel a je pověřen sběrem pojmů.

Začátek procesu: Dokončení draftu výstupu, před odevzdáním

Vstupy: Dokument obsahující draft výstupu
Sdílený dokument Google sheet pro sběr konceptů projektu⁷

Aktivita:

1. Zanesení pojmů do Googlesheetu

Cyklus:

- a. Výběr vhodného pojmu – konceptu
- b. Ověření, že pojem již není součástí Slovníku ani Google sheetu
- c. Zanesení pojmu do Google sheetu
- d. Označení autorství pojmu jménem autora
- e. Označení dokumentu draftu dohodnutou značkou, nejlépe kódem výstupu
- f. Označení všech výskytů pojmu v draftu **tučně kursivou**

Opakování cyklu, dokud nejsou zpracovány všechny vhodné koncepty draftu výstupu.

Poznámka: Tento cyklus provádí nad jedním dokumentem draftu výstupu právě jen jedna osoba

2. Zpracování definic

Cyklus:

- a. Otevření řádku s pojmem
- b. Zpracování definice pojmu
- c. Zpracování odkazu na zdroj definice, pokud existuje
- d. Editace autorství, pokud definici zpracovává někdo jiný než ten, kdo pojem zanesl do Google sheetu

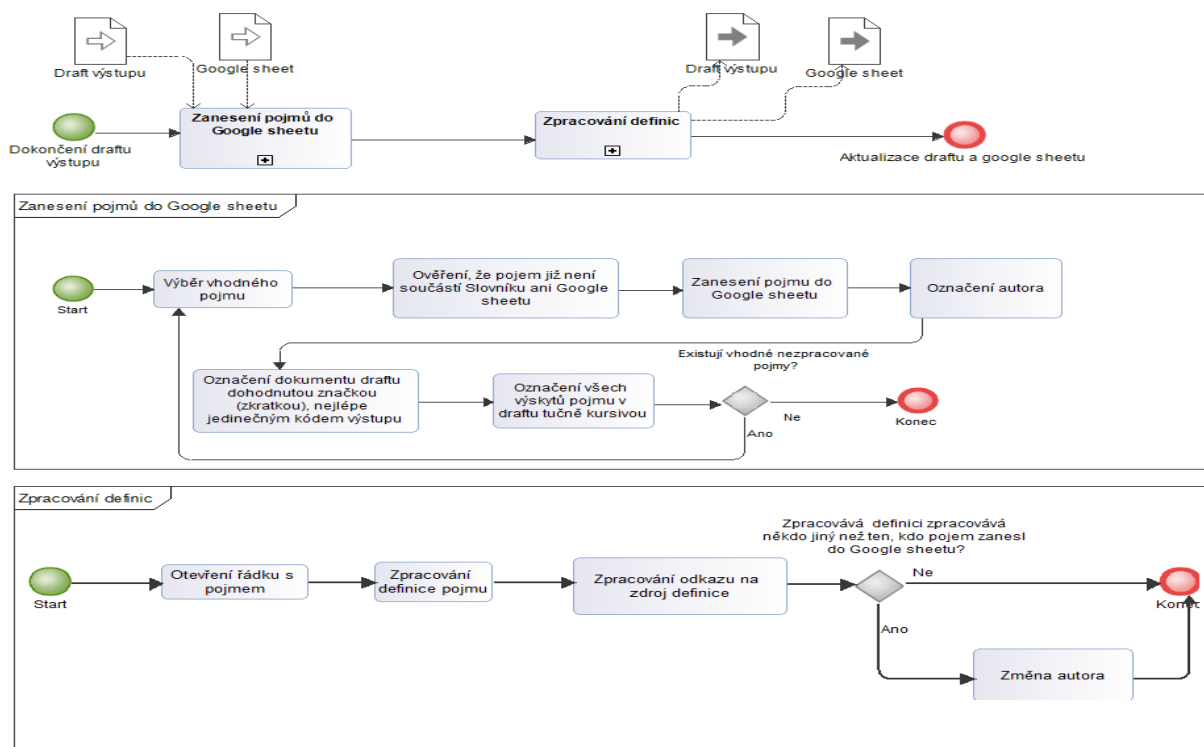
Opakování cyklu, dokud nejsou zpracovány všechny vhodné koncepty draftu výstupu

Poznámka : Zpracování definic může provádět více pracovníků současně

⁷ Tento dokument je třeba založit v rámci nastavení projektu, zakládá ho projektová kancelář

- Konec:** všechny vhodné koncepty z dokumentu draftu výstupu jsou zpracovány
- Výstup:** Aktualizovaný google sheet rozšířený o nové pojmy z daného draftu výstupu.
Aktualizovaný dokument draftu výstupu o označené nové pojmy⁸

5.5.2.3 Procesní diagram



5.5.3 Nominace mluvčího a vytvoření konsensu terminologie týmu

5.5.3.1 Popis procesu

Tato část agendy probíhá jako setkání všech členů týmu, kteří se podíleli na sběru pojmů. Cílem procesu je:

- Určení mluvčího
- Doplnění základních atributů pojmů
- Vytvoření konsensu

Na začátku tým nominuje mluvčího. Mluvčí moderuje schůzku, hlídá čas, v případě konfliktu má poslední slovo a dále za tým pracuje v následujících procesech na dokončení agendy.

Po nominaci mluvčího tým bere jeden nasbíraný pojem za druhým a pokud možno konsensuálně:

- Přijme definici pojmu
- Doplní případně:

⁸ Dokument draftu výstupu by měl být předán k dalšímu zpracování (připomínkám, oponentuře) až poté, co je takto zpracován. Proces může použít i ten, kdo draft připomínkuje, pokud uzná za vhodné, a stejným způsobem do Goofle sheetu zanést další pojmy.

- Synonyma
- Příklady
- Poznámky
- Nehierarchické vazby (souvislosti) na sesbírané pojmy podle typů⁹

Doplnění zmíněných atributů není povinné. Atributy se vyplňují jen v případě, že v týmu spontánně padne návrh na vyplnění toho kterého atributu a tým tento návrh „snadno“ přijme. Kontroverzní hodnoty atributů (na kterých není dostatečný konsensus) nemají být zaznamenány.

V průběhu zpracování pojmu může dojít ke změně autora, pokud autorský příspěvek nějakého člena týmu při zpracování pojmu převáží příspěvek autora, který navrhl pojem a definici, ale zpravidla zůstává původní autorství. O autorství rozhoduje mluvčí týmu.

Zpracování jednoho pojmu by nemělo **v průměru** trvat déle než 5 minut. Vzhledem k nákladnosti tohoto procesu je kriticky důležité hlídat čas a nesnažit se o dokonalost.

Je třeba pamatovat na 5-10 minut přestávku po zpracování zhruba devíti pojmů a plánovat dobu trvání schůzky striktně podle počtu zpracovávaných pojmů s rezervou + 10% (+ čas na přestávku).

Plánovaný celkový čas jedné schůzky nemá přesáhnout 120 minut, plánovaný čas nesmí být překročen.

Pokud je vzhledem k počtu pojmů organizovat více schůzek, mluvčí určený na první schůzce se nemění (tým má pouze jednoho mluvčího).

Schůzky mohou být organizovány jako virtuální (např. na skypu).

5.5.3.2 Hodnoty atributů procesu:

Majitel procesu : Mluvčí týmu projektu

Začátek procesu : První schůzka týmu k vytvoření konsensu terminologie

Vstupy: Sdílený dokument Google sheet pro sběr konceptů projektu naplněný novými koncepty

Aktivity:

1. 1. schůzka
 - a. Nominace mluvčího
 - b. Cyklus nad každým konceptem
 - i. Konsensus definice
 - ii. Případné doplnění dalších atributů:
 1. Synonyma
 2. Příklady
 3. Poznámky
 - iii. Určení autora

Cyklus pokračuje, dokud:

- a) nejsou zpracovány všechny nové pojmy (koncepty)

nebo

- b) není čas na přestávku (po devíti zpracovaných konceptech)

⁹

Viz strukturu polí a legendu Google sheet

nebo

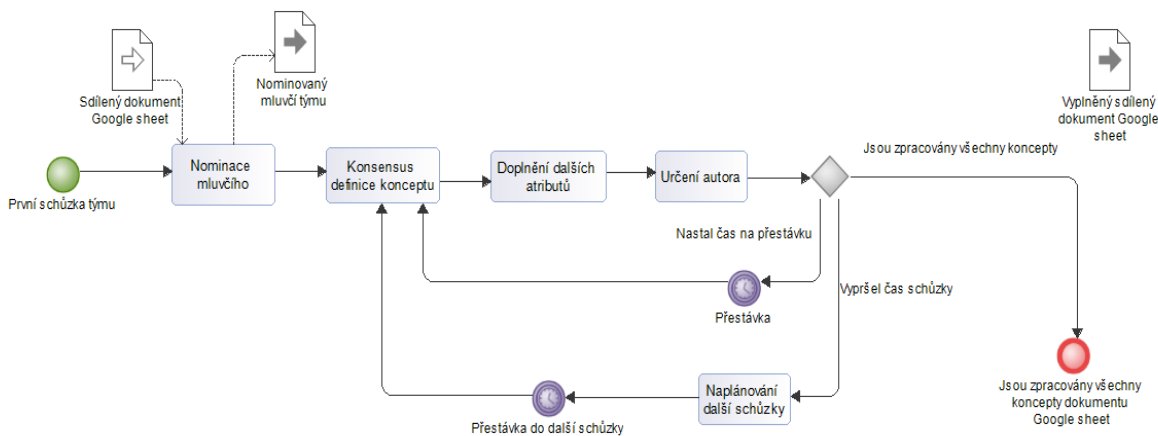
- c) nevypršel čas schůzky
 2. Druhá a další schůzka¹⁰
 - a. Cyklus nad každým pojmem jako výše

Konec: všechny koncepty dokumentu Google sheet jsou zpracovány

Výstupy: Nominovaný mluvčí týmu

Vyplněný sdílený dokument Google sheet pro sběr konceptů projektu

5.5.3.3 Procesní diagram



5.5.4 Import nových pojmů do RDF úložiště

5.5.4.1 Popis procesu

Po ukončení předchozího procesu správce sémantického slovníku projektu vytvoří novou pracovní databázi sémantického modelu v RDF4J, kam naimportuje obsah stávajícího sémantického slovníku NCEZ a do ní provede pomocí příslušného nástroje import obsahu sdíleného dokumentu Google sheet. Dále agenda pracuje s touto pracovní databází. **Pokud je vytvářen pouze jednoduchý slovník, agenda vynechává všechny následující procesy a pokračuje publikací nové verze Sémantického slovníku (proces 5.5.7).**

5.5.4.2 Atributy procesu

Majitel procesu : Správce sémantického modelu

Začátek procesu : Konec předchozího procesu

Vstup: Sdílený dokument Google sheet se zpracovanými koncepty

Aktivita: import dokumentu Google sheet do RDF4J

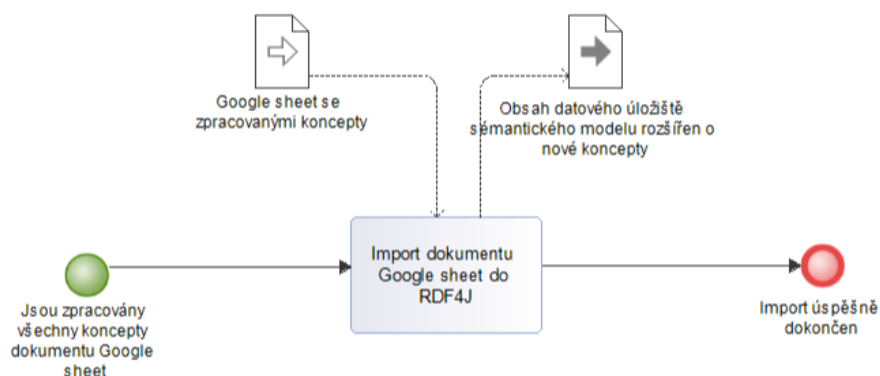
Konec: import úspěšně dokončen

¹⁰

jen v případě, že všechny pojmy nebyly zpracovány v průběhu první schůzky

Výstup: obsah datového úložiště sémantického modelu rozšířen o nové koncepty z dokumentu Google sheet.

5.5.4.3 Procesní diagram



5.5.5 Kodifikace atributů pojmů a vytvoření (validace) sémantických vazeb

5.5.5.1 Popis procesu

Mluvící týmu a slovníkový specialista na společné schůzce společně projdou všechny nové koncepty, zařadí je do terminologických skupin, přiřadí je, pokud lze, ke kategoriím RIM HL7 a identifikují relevantní sémantické vazby mezi novými a starými pojmy Slovníku.

Po zpracování všech nových konceptů synchronizuje slovníkový specialista obsah RDF4J a nástroj na modelování sémantických vazeb.

Následně slovníkový specialista vizualizuje postupně každou terminologickou skupinu a kategorii HL7, v které se vyskytuje alespoň jeden nový pojem, v modelovacím UML nástroji a společně s mluvcím týmu případně doplní či modifikují sémantické vazby. Pokud to dovolí kompetence slovníkového specialista a komplikovanost vazeb, může slovníkový specialista kodifikovat na této schůzce i hierarchické sémantické vazby na koncepty UFO případně přiřadit konceptům.

Proces může probíhat ve více schůzkách. Při zpracování konceptů platí pro průběh schůzky, její trvání a počet zpracovaných konceptů obdobná pravidla jako v procesu 2.4.2. s tím, že na zpracování jednoho konceptu je třeba počítat zhruba dvojnásobný čas (10 minut). Slovníkový specialista je uveden u všech zpracovaných nových konceptů jako Posuzovatel. **Pokud není vytvářen kompletní sémantický slovník, agenda vynechává následující proces a pokračuje publikací nové verze Sémantického slovníku (proces 5.5.7).**

5.5.5.2 Hodnoty atributů procesu

Majitel procesu : Slovníkový specialista

Začátek procesu : Konec předchozího procesu

Vstup: RDF4J rozšířený o nové koncepty

Aktivita:

1. Rozšíření struktury terminologických oblastí ve Vocabench
2. Cyklus nad každým novým konceptem:
 - a. Zařazení do terminologické skupiny

- b. Přiřazení ke kategorii HL7
- c. Identifikace a přiřazení případných dalších hierarchických vazeb na koncepty slovníku
- d. Identifikace a přiřazení nehierarchických vazeb na koncepty slovníku podle typů.

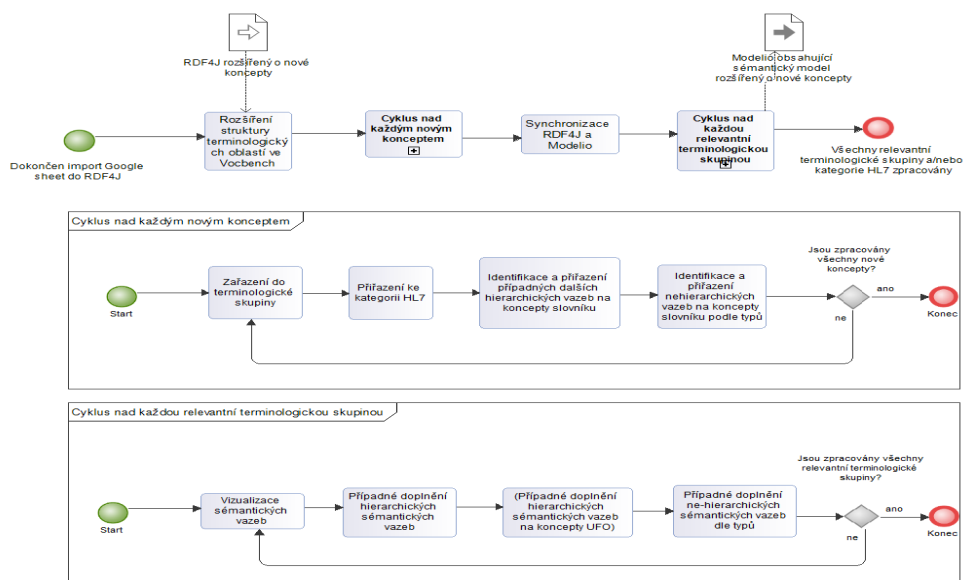
Cyklus pokračuje, dokud nejsou takto zpracovány všechny nové koncepty.

3. Synchronizace RDF4J a Modelio (provede správce sémantického modelu)
4. Cyklus nad každou relevantní terminologickou skupinou a/nebo kategorií HL7:
 - a. Vizualizace sémantických vazeb
 - b. Případné doplnění hierarchických sémantických vazeb
 - c. Případné doplnění UFO typů k jednotlivým konceptům resp. vazbám
 - d. Případné doplnění ne-hierarchických sémantických vazeb dle typů
 Cyklus pokračuje, dokud nejsou takto zpracovány všechny relevantní terminologické skupiny a/nebo kategorie HL7.

Konec: Všechny relevantní terminologické skupiny a/nebo kategorie HL7 zpracovány import úspěšně dokončen

Výstup: Modelio obsahující sémantický model rozšířený o nové koncepty a jejich sémantické vazby

5.5.5.3 Procesní diagram



5.5.6 Vytvoření ontologického modelu

5.5.6.1 Popis procesu

Pokud v předchozím procesu nebyly pro nové obecné koncepty doplněny hierarchické sémantické vazby na terminologickou skupinu konceptů UFO a pokud tak rozhodne slovníkový specialista, jsou tyto vazby doplněny na společné schůzce slovníkového specialisty s ontologickým inženýrem, na kterou může být přizván i mluvčí týmu. Na schůzce (schůzkách) prezentuje slovníkový specialista nové obecné koncepty sémantického modelu a společně hledají a doplňují sémantické vazby na koncepty (objekty) UFO. Ontologický inženýr zároveň provádí ontologické testování celého sémantického modelu k odhalení případných nekonsistencí. Ty jsou odstraněny změnou příslušných sémantických vazeb na základě konsensu zúčastněných, případně může být i opraveny detaily glosáře.

5.5.7 Publikování nové verze Sémantického slovníku

5.5.7.1 Popis procesu

Po dokončení předchozích procesů správce sémantického modelu archivuje aktuální dokumentaci sémantických vazeb sémantického modelu a vytvoří novou dokumentaci sémantických vazeb exportem diagramů sémantických vazeb do příslušných výstupních souborů. Poté synchronizuje Modelio a pracovní databázi sémantického slovníku, s kterou pracuje tato agenda. Následně archivuje aktuální produkční databázi sémantického modelu, kterou nahradí pracovní databází sémantického modelu, s kterou pracuje tato agenda. Tu pak publikuje na webový portál Sémantického modelu jako novou platnou verzi.

5.5.7.2 Atributy procesu

Majitel procesu : Správce sémantického modelu

Začátek procesu : Konec předchozího procesu

Vstupy: Modelio obsahující definitivní sémantické vazby modelu

Pracovní databáze sémantického modelu v RDF4J

Aktuální produkční databáze sémantického modelu v RDF4J

Aktivita:

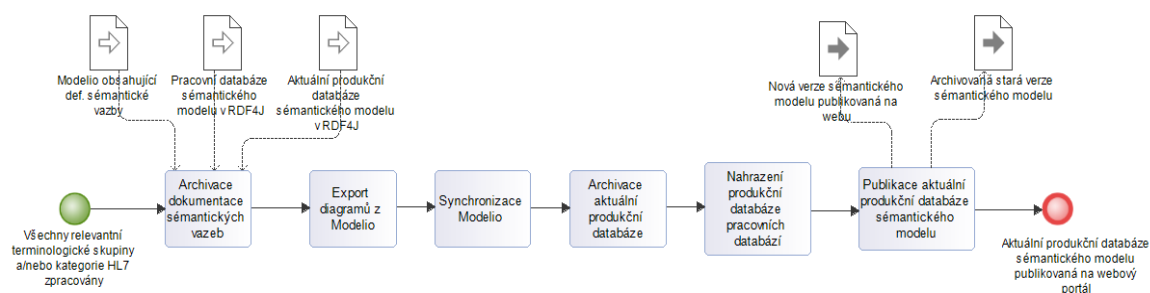
1. Archivace dokumentace sémantických vazeb
2. Export diagramů sémantických vazeb z Modelio a vytvoření nové dokumentace sémantických vazeb
3. Synchronizace Modelio a pracovní databáze sémantického modelu
4. Archivace aktuální produkční databáze sémantického modelu
5. Nahrazení produkční databáze pracovní databází
6. Publikace aktuální produkční databáze sémantického modelu na webový portál sémantického modelu

Konec: Aktuální produkční databáze sémantického modelu publikovaná na webový portál

Výstupy: Nová verze sémantického modelu publikovaná na webu a dokumentovaná v nové sadě diagramů sémantických vazeb

Archivovaná stará verze sémantického modelu

5.5.7.3 Procesní diagram



Web strategie: <http://www.nsez.cz>

Toto dílo podléhá licenci Creative Commons CC BY 4.0. Dílo je možné libovolně šířit a upravovat za předpokladu uvedení citace tohoto díla. Pro zobrazení podrobných licenčních podmínek navštivte <http://creativecommons.org/licenses/by/4.0/>. Licence se nevztahuje na použití loga Ministerstva zdravotnictví České republiky mimo reprodukci tohoto díla. Veškerá práva k logu jsou vyhrazena.

Citace dle ČSN ISO 690:2011:

MINISTERSTVO ZDRAVOTNICTVÍ ČESKÉ REPUBLIKY. *Soubor metodik pro správu terminologie elektronického zdravotnictví*. Verze 1.0. Praha, 2019. Licencováno pod CC BY 4.0, licenční podmínky dostupné z: <http://creativecommons.org/licenses/by/4.0/>.

