

Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
Katedra informatiky a výpočetní techniky

## Diplomová práce

# BodyInNumbers a prostředky sémantického webu

Místo této strany bude  
zadání práce.

# Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 16. května 2019

Vítězslav Vacek

## **Abstract**

### **BodyInNumbers and semantic web resources**

The main goal of this master thesis is to automate or otherwise simplify the conversion of collected data and metadata from Body in Numbers system to the popular RDF format. This format is being often used in the field of medical informatics and is also used to share collected medical data among researchers inside the same area of expertise. To successfully convert the collected data into any of the possible RDF formats, it will be necessary to define an ontology describing the terminology used inside the BiN system and during the data collection process. It will also be necessary to select a suitable RDF store for the job and then deploy it inside the BiN infrastructure. This deployed solution will then be filled with real, converted data from the BiN system and it will be tested again during its live service.

## **Abstrakt**

Tato práce si klade za cíl automatizovat či jinak zjednodušit konverzi nasbíraných dat a metadat ze systému Body in Numbers do populárního formátu RDF. Tento formát se často využívá na poli medicínské informatiky a zároveň se používá i pro sdílení nasbíraných medicínských dat mezi výzkumnými pracovníky ve stejném oboru. Pro úspěšnou konverzi do některého z možných RDF formátů bude nutné nadefinovat ontologii popisující používanou terminologii v systému BiN a i během procesu sbírání dat. Také bude potřeba vybrat vhodné úložiště a následně ho nasadit do provozu v rámci infrastruktury BiN. Toto realizované řešení bude následně naplněno daty z BiN a znovu otestováno již v provozu.

# Obsah

<b>1</b>	<b>Úvod</b>	<b>8</b>
<b>2</b>	<b>State of the Art</b>	<b>9</b>
2.1	Body in Numbers . . . . .	9
2.1.1	Systémy podobné BiN . . . . .	11
2.1.2	Datový cyklus BiN . . . . .	12
2.1.3	Používaná terminologie v rámci BiN . . . . .	13
2.1.4	Příklad - založení projektového dne . . . . .	14
2.1.5	Architektura systému BiN . . . . .	14
2.1.6	Uživatelské role v rámci BiN . . . . .	17
2.1.7	Sběr metadat dat v systému BiN . . . . .	18
2.1.8	Sběr medicínských dat v systému BiN . . . . .	18
2.1.9	Stanoviště BCI . . . . .	19
2.1.10	Průběh měření v systému BiN . . . . .	21
2.2	Sémantický web . . . . .	23
2.2.1	Linked Data . . . . .	24
2.3	RDF . . . . .	25
2.3.1	RDF trojice . . . . .	25
2.3.2	Formáty RDF (serializace RDF) . . . . .	25
<b>3</b>	<b>OWL ontologie</b>	<b>30</b>
3.1	Podjazyky OWL a OWL 2 . . . . .	30
3.1.1	OWL 2 EL . . . . .	32
3.1.2	OWL 2 QL . . . . .	32
3.1.3	OWL 2 RL . . . . .	32
3.2	Ontologické portály . . . . .	33
3.3	Ontologické dobré mravy . . . . .	33
3.3.1	Postup při vytváření ontologie popisující systém Body in Numbers (BIN) . . . . .	36
3.3.2	Používané anotace uvnitř BiN ontologie . . . . .	37
<b>4</b>	<b>RDF datová úložiště</b>	<b>40</b>
4.1	Realizace konverzního algoritmu z CSV na RDF . . . . .	40
4.1.1	Testování dostupných řešení . . . . .	41
4.1.2	CSV2RDF modul . . . . .	42
4.1.3	Úprava CSV2RDF modulu . . . . .	43

4.1.4	Podporované přepínače . . . . .	43
4.1.5	Příklad přeložení vstupního CSV souboru . . . . .	45
4.2	Populární RDF úložiště . . . . .	45
4.2.1	Prvotní výběr zájmových úložišť - pomocí článku . . . . .	49
4.3	Výběr tří RDF úložišť pro podrobnější testování . . . . .	52
<b>5</b>	<b>Testování vhodných RDF úložišť</b>	<b>53</b>
5.1	Parametry testovacího stroje . . . . .	53
5.2	Vybrané RDF formáty . . . . .	53
5.3	Použité SPARQL dotazy . . . . .	54
5.3.1	SPARQL dotaz 1 . . . . .	55
5.3.2	SPARQL dotaz 2 . . . . .	55
5.3.3	SPARQL dotaz 3 . . . . .	56
5.3.4	SPARQL dotaz 4 . . . . .	57
5.3.5	SPARQL dotaz 5 . . . . .	58
5.3.6	SPARQL dotaz 6 . . . . .	59
5.3.7	SPARQL dotaz 7 . . . . .	61
5.4	Testování rychlosti nahrávání vstupních RDF souborů . . . . .	62
5.5	Testování rychlosti dotazování . . . . .	63
5.6	Výběr vhodného RDF úložiště . . . . .	65
<b>6</b>	<b>Nasazení a naplnění vybraného RDF úložiště</b>	<b>66</b>
6.1	Testování nasazeného řešení . . . . .	66
6.1.1	Test nahrávání různých RDF formátů . . . . .	67
6.1.2	Test dotazování se na načtená data v různých seriali- zacích . . . . .	67
6.2	Zhodnocení nasazení a otestování RDF úložiště . . . . .	71
<b>7</b>	<b>Zhodnocení dosažených výsledků</b>	<b>72</b>
<b>8</b>	<b>Závěr</b>	<b>75</b>
<b>A</b>	<b>Seznam zkratk</b>	<b>76</b>
<b>B</b>	<b>Seznam literatury</b>	<b>77</b>
<b>C</b>	<b>Seznam obrázků</b>	<b>80</b>
<b>D</b>	<b>Seznam tabulek</b>	<b>83</b>
<b>E</b>	<b>Terminologie BiN</b>	<b>84</b>



# 1 Úvod

V dnešní době informačních technologií je společnost zavalena velkým množstvím dat, které je problém zpracovat a interpretovat. Proto vznikají specializované systémy na jejich zpracování, ukládání a vizualizaci a to jak s rozhraním pro člověka, tak pro stroj. Medicínská data, zejména v oblasti monitoringu neurálních funkcí, jsou rozsáhlá a nepřehledná.

Prvním krokem k jejich předzpracování, je rozpoznání jednotlivých částí problému a posléze kategorizace takovým způsobem, aby byla data přenositelná a sdílitelná mezi různými pracovními skupinami, mnohdy rozptýlenými po celém světě. Je potřeba vytvořit tyto ontologie schematizující vztahy mezi termíny, aby byly lépe sdílitelné. Proto bylo úkolem zpracovat systém, který by tato data dokázal rozumně předzpracovat a přiblížit člověku.

Projekt Body in Numbers si klade za cíl vytvořit unikátní anotovanou kolekci heterogenních medicínských dat, která budou sloužit k dalším analýzám. Body in Numbers sbírá kromě metadat ve formě dotazníku i medicínská data (hmotnost, výška, krevní tlak, ...) a EEG data z BCI experimentů. Takto sesbíraná data jsou poté anonymizována a publikována ve formě datových článků, které mohou sloužit právě ke komplexnější analýze nejen pro výzkumné pracovníky Neuroinformatické laboratoře na Fakultě aplikovaných věd Západočeské univerzity v Plzni, ale i pro širší vědeckou obec.

Právě proto vznikla v poslední době potřeba zjednodušit proces předzpracování takto sesbíraných dat, aby je bylo možné sdílet mezi širší vědeckou obcí. K umožnění tohoto kroku je potřeba převést sesbíraná data do některého z všeobecně používaných RDF formátů a vytvořit zaštitující ontologii, která případně nadefinuje dodatečné vlastnosti sesbíraných dat.



## 2 State of the Art

Kapitola State of the Art se zabývá nadefinováním všech potřebných terminologií a předpokladů, se kterými je nutné se seznámit. V této kapitole se také řeší problematika systému Body in Numbers a způsob sběru dat uvnitř tohoto systému. Také se přiblíží myšlenka sémantického webu včetně některých jeho funkčních částí, které jsou již realizovány a v praxi používány.

### 2.1 Body in Numbers

Projekt Body in Numbers (dále jen „BiN“) si klade za cíl vytvořit unikátní anotovanou kolekci heterogenních medicínských dat, která budou sloužit k dalším analýzám (např. souvislost životního stylu a chronických onemocnění). Tento projekt již vytvořil funkční systém, který podporuje sběr dat ve formě dvou různých mobilních aplikací a jedné webové, která umožňuje zobrazit jednoduché statistiky [26]. Systém BiN je unikátní a již autorizovaný software vytvořený kolektivem KIV sloužící pro vědeckou práci nejen z oblasti neuroinformatiky.

Heterogenita dat spočívá jak ve sběru metadat (např. pomocí dotazníků), tak ve sběru samotných medicínských dat (např. síla, hmotnost, spirometrie), ale zároveň obohacenou unikátní částí dat o mozkové aktivitě jedince.

Dalším úkolem, i když méně výrazným, projektu BiN je šíření obecného povědomí o zdravém životním stylu a jeho potencionálních účincích na psychiku jedince.

Samotný systém BiN je rozdělen dle funkčního hlediska do dvou celků a to:

- **Mobilní aplikace pro sběr dat** – Slouží primárně ke sběru dat a metadat uvnitř procesu měření. Zároveň je mobilní aplikace schopná zobrazit všechna již sesbíraná data určitého měřeného subjektu[2].
- **Mobilní fitness aplikace** – Dokáže kontrolovat a doporučit úpravy jídelníčku (proporcionální snížení dávek) a také doporučuje i cviky pro zlepšení kondice.
- **Webová aplikace** – Hlavním cílem je zobrazovat základní přehledové statistiky různých změřených veličin. Může se jednat například o průměrné reakční doby horních končetin v rámci jednoho projektového

dne, nebo celkové přehledy reakčních dob. Webová aplikace také dokáže sledovat aktuální stav účastníků v rámci probíhajícího měření a v případě, že účastník již všechna stanoviště absolvoval, změřit i jeho celkový čas, věnovaný experimentu.

Projekt BiN se zaměřuje primárně na věkové kategorie od 11 do 60 let. Mimo toto věkové rozmezí jsou naměřené výsledky zkresleny určitými chybami, vyplývajícími buď z příliš nízkého a nebo pokročilejšího věku. V případě předškolních dětí může být například problém přečíst čísla, nebo nedosáhnou na stůl s tlačítky. V případě lidí s motorickými problémy zejména pokročilejšího věku, může být problém zareagovat rychle pomocí nohou a podobně.

Dále se sledované věkové rozmezí dělí na zájmové kategorie, které jsou zobrazeny na následujícím obrázku 2.1, kdy ke každé z uvedených věkových kategorií se během datového cyklu musí přistupovat individuálně. Například jinak se bude postupovat se souhlasem a vyplněním dotazníku pro malé dítě<sup>1</sup> a odlišně pro dospělého.

## Použité technologie v projektu BiN

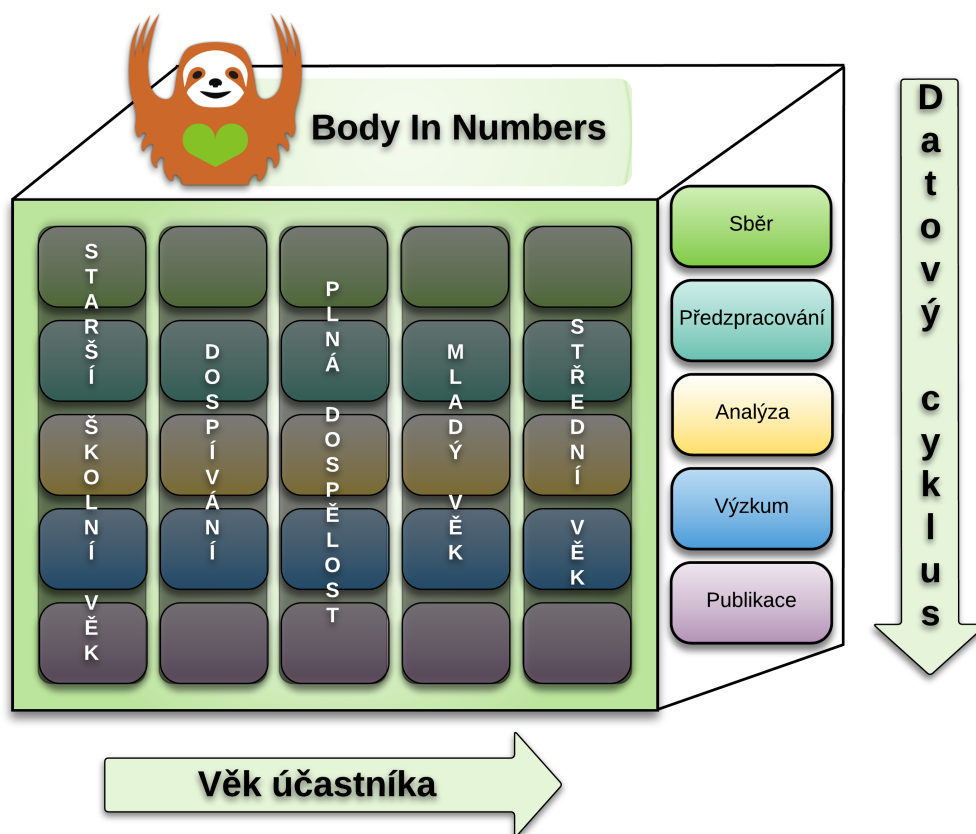
Následující seznam použitých technologií obsahuje technologie potřebné jak pro webovou aplikaci tak i obě mobilní aplikace (fitness, sběr dat). Během vývoje a provozování systému BiN byly využity následující programovací jazyky, frameworky a technologie:

- HTML
- CSS
- Javascript
- Python
- Docker
- Flask
- SQLAlchemy
- PostgreSQL
- Ionic
- React Native

Vzhledem k tomu, že značná část aplikací BiN využívá programovací jazyk Python, tak by bylo výhodné najít takové nástroje, které podporují vytváření a nebo využívání služeb sémantického webu za použití programovacího jazyku Python.

---

<sup>1</sup>Účast v projektu musí odsouhlasit zákonný zástupce.



**Obrázek 2.1:** Pro každou věkovou kategorii může probíhat sběr, předzpracování (případně i výzkum) a publikace dat trochu jinak. Věková rozmezí pro každou z uvedených kategorií jsou následující: starší školní věk je od 11 do 14 let, dospívání je od 15 do 18 let, plná dospělost je od 19 do 30 let, mladý věk je od 31 do 45 let a střední věk je od 46 do 60 let.

## Přehled naměřených dat v projektu BiN

Za dobu trvání projektu BiN bylo již změřeno přes 650 účastníků, kde zhruba 100 měření obsahuje i naměřené EEG. Část nasbíraných dat byla již publikována v časopise Elsevier [4]. Tato již publikovaná data byla původně pořízena v rámci Dnů vědy a techniky 2016 pořádaných v Plzni.

### 2.1.1 Systémy podobné BiN

Již existují aplikace, které mají podobnou myšlenku jako projekt BiN.

Například systém Vitabot<sup>2</sup>, se zaměřuje na funkčnost výživového poradce. Kde se podle zadaných informací upraví doporučený jídelníček účast-

<sup>2</sup>Dostupný na <https://www.vitabot.com/web/>. [cit. 4.2.2019]

níka, aby obsahovala vyvážený podíl všech vitamínů a dalších nutričních prvků. Tento systém vyváženosti stravy je založen na proporcionálním snižování oblíbených potravin uživatele. Tento systém také monitoruje fyzickou aktivitu uživatele pomocí kombinace chytrých náramků (např. FitBit), které mohou posílat informace do specializované mobilní aplikace. Tento systém je určen pro individuální použití jedním uživatelem, kdežto systém BiN je určen pro hromadné ukládání biomedicínských dat.

Dalším příkladem podobného systému tentokrát pracuje s kognitivními hrami, které zlepšují paměť, pozornost, rychlost, řešení problémů a další. Tento systém se nazývá Lumosity<sup>3</sup>.

Každý z těchto systémů se specializuje na určitou specifickou problematiku, ale žádný z nich nepracuje přímo s naměřenými daty o mozkové aktivitě (např. EEG).

### 2.1.2 Datový cyklus BiN

Fáze, v jakých fázích se mohou nalézat data pořízená v rámci projektových dnů, lze vidět na obrázku 2.2.

Fáze sběr se zabývá pořízením dat od účastníka projektu BiN, která jsou následně předpracována. Předzpracování se skládá z očištění dat<sup>4</sup> a případné úpravy pro další krok datového cyklu tedy o převod do požadovaného formátu analýzy.

Fáze analýzy se v současné chvíli zabývá určením statistické významnosti mezi naměřenými daty účastníků a dotazníkovými otázkami. Z tohoto kroku jsou vytvořeny přehledové tabulky a grafy, které jsou následně využívány pro nalezení další výzkumné činnosti.

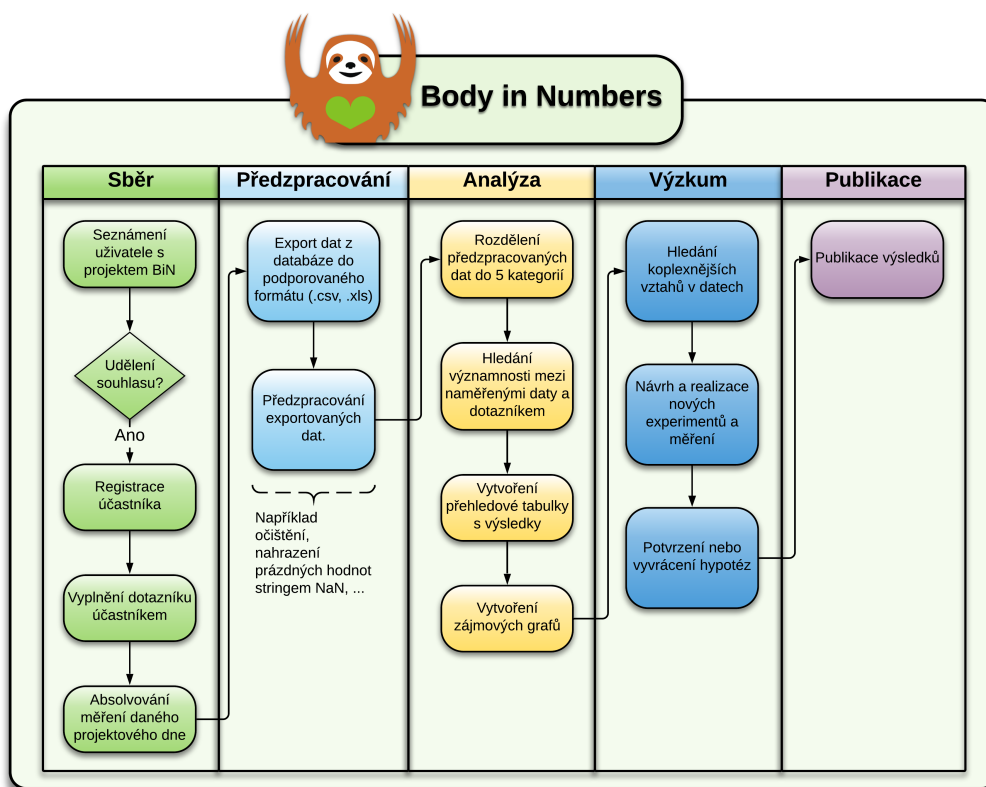
Během fáze výzkumu dochází k nastínění nových hypotéz, které se následně pomocí pilotního experimentu snaží vědečtí pracovníci buď vyvrátit a nebo potvrdit.

V poslední fázi datového cyklu dochází k publikaci anonymizovaných primárně naměřených dat včetně výsledků dosažených prvotní analýzou a pilotním experimentem. Takto publikovaná data mohou pomoci širší vědecké obci při zodpovídání jiných otázek a hypotéz.

---

<sup>3</sup>Dostupný na <https://www.lumosity.com/en/>. [cit 4.2.2019]

<sup>4</sup>Odstranění prázdných nebo nesmyslných hodnot (např. hmotnost 500 kg), přecíslování některých atributů (např. odpovědi na otázky z dotazníku se převedou na čísla), a další.



Obrázek 2.2: Datový cyklus naměřených dat v projektu BiN. Jednotlivé fáze datového cyklu byly již nastíněny na předchozím obrázku 2.1.

### 2.1.3 Používaná terminologie v rámci BiN

V BiN je používána vlastní specifická terminologie [26]. Jedná se o následující pojmy:

- **Pomůcky** - Přístroje a nebo nástroje, které jsou používány během experimentů. Například tlakoměr pro měření tlaku a tepu, spirometr pro měření kapacity plic nebo metr pro měření výšky.
- **Experiment** - Jedno vyšetření účastníka, provedené podle příslušného scénáře. Například měření krevního cukru vyžaduje píchnutí do prstu, nabrání krve na měřící proužek a následné vyhodnocení výsledků glukometrem. Výstupem experimentu jsou změřené veličiny (krevní cukr = 6.4 mmol/l).
- **Stanoviště** - Místně určené měření, které může obsahovat jeden nebo více experimentů.

- **Měření** - Definice všech stanovišť a experimentů s přidělenými pomůckami. Jedná se o upřesnění toho, co a jak se bude v daný projektový den měřit.
- **Plánování** - Založení nového měření, které obsahuje nové definice toho co, jak a pomocí čeho se bude měřit.

### 2.1.4 Příklad - založení projektového dne

Pokud bychom chtěli v systému BiN založit nové měření, abychom mohli provádět sběr dat v rámci projektového dne, bude nejprve potřeba založit některé instance z výše definované terminologie.

Jako první krok je nutné vytvořit všechny nové **pomůcky**, které chceme během projektového dne použít. Pokud se v experimentu využijí pomůcky, které se již používaly v rámci jiných experimentů, tak je systém BiN umožňuje přiřadit k nově definovanému experimentu. Může se například jednat o zesilovač EEG, EEG čepici, stimulátor reakčních dob a dvě stolní PC (stimulační a měřicí). Definují se pouze ty pomůcky, které jsou nezbytně nutné pro měření, takže se například nedefinují pomůcky typu stoly, nábytek a místnosti. U nových pomůcek je potřeba vyplnit název zařízení, sériové číslo, jednotky měření a na závěr je možné přidat komentář (např. o kolik váha proměřuje, apod.).

Když se takto nedefinují **pomůcky**, je potřeba vytvořit **experiment**. V experimentu je nutno určit, jaká veličina se bude měřit pomocí jakého přístroje. Například hmotnost se bude měřit vahou Omron BF511 v kilogramech.

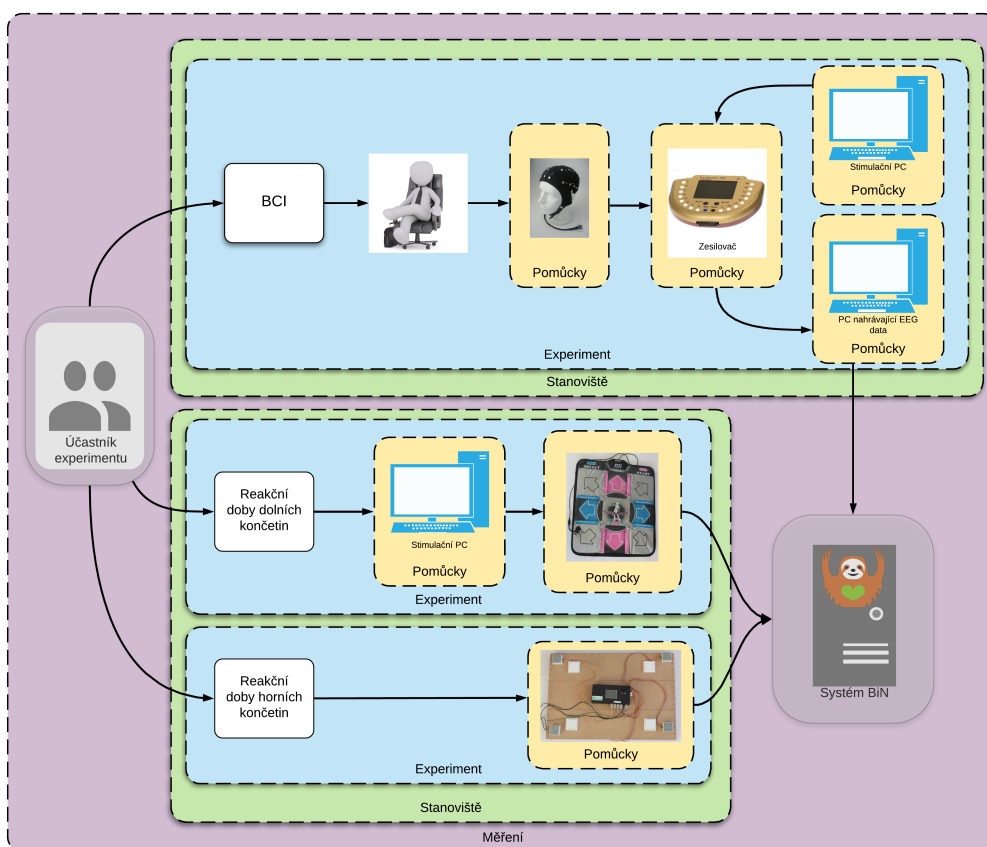
Nastavení **stanoviště** upřesní jaké experimenty se budou realizovat. V rámci jednoho stanoviště se může vykonávat více experimentů. Například v BiN existuje stanoviště Srdce a krev, které měří krevní tlak, tep účastníka v klidu a případně i krevní cukr. Pro založení nového stanoviště je nutné nadefinovat název, přiřazení experimentů a může mít i vygenerovaný QR kód určený pro usnadnění sběru pomocí mobilní aplikace.

V poslední části se musí vytvořit **měření**, které obsahuje všechna stanoviště, která se budou v rámci jednoho projektového dne realizovat.

Jak takto vytvořený projektový den může vypadat v použité terminologii je vidět na následujícím obrázku 2.3.

### 2.1.5 Architektura systému BiN

Serverová část systému BiN je dělena na celkem čtyři Dockerové kontejnery. Jak je rozdělena architektura BiN z pohledu Docker kontejner lze vidět na obrázku 2.4. Jedná se o následující oddělené funkcionality:

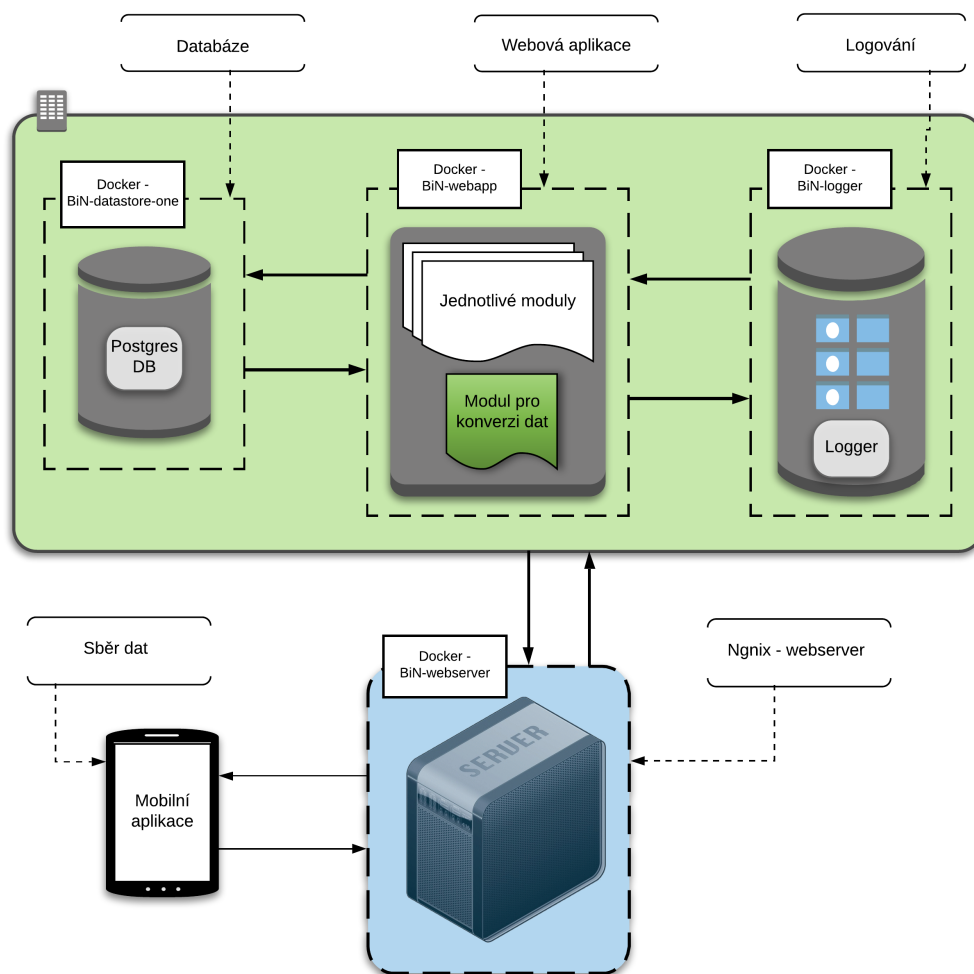


**Obrázek 2.3:** Definice projektového dne v rámci terminologie systému BiN.

- **Webserver** – Obsahuje samotnou aplikaci Ngnix, které kontroluje vstupy do webového serveru a přeposílá je na patřičné aplikace (webová aplikace, databáze, logger).
- **Webapp** – Docker kontejner webové aplikace BiN.
- **Datastore-one** – Databáze Postgres, která je využívána webovou aplikací BiN.
- **Logger** – Logovací server umístěný v rámci systému BiN. V současné době je však otázkou zda by neporušoval GDPR, a tak se zatím tento modul ještě neimplementoval.

### Formát dat pro komunikaci mezi mobilní a webovou aplikací

Mobilní aplikace převádí sesbíraná heterogenní data do formátu JSON. Takto sesbíraná data jsou odesílána do webové aplikace pomocí REST API, kde jsou dále zpracovávána a následně uložena do Postgres databáze.

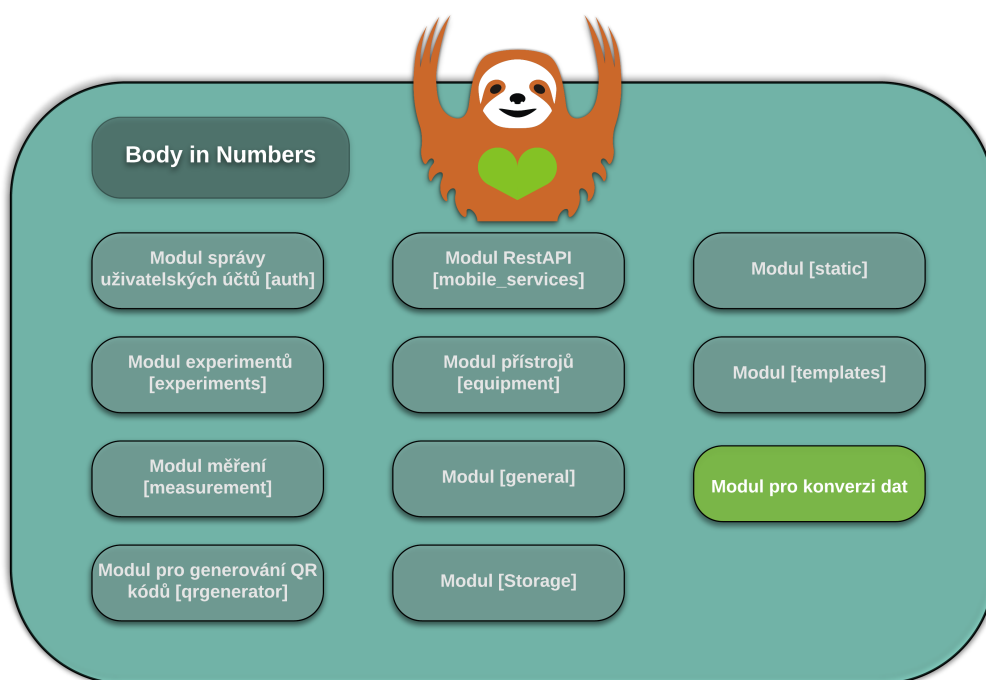


**Obrázek 2.4:** Rozdělení architektury systému BiN dle Docker kontejnerů. Bude vytvořen nový modul do webové aplikace, který bude řešit konverzi dat. V této architektuře je také obsažen sporný modul Logger.

### Dělení webové aplikace BiN

Samotná webová aplikace se dále dělí na moduly, kde každý modul řeší určitou část funkčnosti. Jedná se celkem o deset funkčních modulů. V rámci této práce bude vytvořen nový modul zajišťující převod naměřených dat do formátu RDF. Plánovaný modul je však natolik rozsáhlý, že množství práce, potřebné pro vytvoření tohoto modulu přesahuje svou velikostí rozsah této diplomové práce. Proto byla odzkoušena pouze konzolová aplikace, které ověřila možnosti řešení. Schématické zakreslení těchto modulů lze vidět na následujícím obrázku 2.5.





**Obrázek 2.5:** Uspořádání modulů v rámci webová aplikace BiN včetně přidaného modulu pro konverzi naměřených dat do RDF.

### 2.1.6 Uživatelské role v rámci BiN

V rámci systému BiN jsou implementovány základní uživatelské role, které mají za cíl rozdělit a omezit přístup k určitým částem systému. V současné chvíli je celkem vytvořeno pouze pět uživatelských rolí. Jedná se o:

- **Měřený subjekt** – Účastník měření, který si může zobrazit svoje naměřená data, která jsou uchovávána v systému BiN.
- **Experimentátor** – Osoba provádějící určitý experiment. Jejím úkolem je seznámit měřený subjekt s tím co se bude v rámci experimentu provádět a zadává naměřená data do aplikace BiN.
- **Správce serveru** – Osoba, která se stará o provoz, dostupnost a funkčnost aplikace pouze na serverové straně nasazeného systému BiN.
- **Správce aplikace** – Osoba, která může přistupovat přes webovou aplikaci k naměřením datům ze všech projektových dnů. Spravuje také přístup jednotlivých rolí k webové aplikaci (autorizace experimentátorů, apod.).

- **Datový analytik** – Vůbec nepřistupuje k systému BiN, ale pracuje pouze s anonymizovanými daty, která jsou poskytnutá od správce aplikace.

### 2.1.7 Sběr metadat dat v systému BiN

Předtím, než se účastník projektu BiN může zúčastnit samotného měření, je potřeba náležitě vyplnit dva dokumenty.

V prvním z nich je účastník seznámen s průběhem celého experimentu a souhlasí se samotným měřením.

V druhém dokumentu se provádí již sběr metadat před samotným měřením. V rámci tohoto dokumentu musí účastník odpovědět na 19 uzavřených otázek, kde právě jedna možnost je správná. Při použití mobilní aplikace pro sběr dat je možné vyplnit dotazník přímo v ní.

Naměřená data v projektu BiN jsou ukládána odděleně od osobních informací a zároveň jsou veškeré záznamy anonymizovány, aby je nebylo možné použít k přímé nebo nepřímé identifikaci jedince v případě, že by došlo k úniku informací.

### 2.1.8 Sběr medicínských dat v systému BiN

V rámci systému BiN jsou v roce 2019 měřeny následující informace:

- |                                 |                                  |
|---------------------------------|----------------------------------|
| • Reakční doby horních končetin | • Výška                          |
| • Reakční doby dolních končetin | • BMI                            |
| • Barvocit                      | • Zastoupení svalstva v těle [%] |
| • Spirometrie                   | • Zastoupení vody v těle [%]     |
| • Krevní tlak                   | • Zastoupení tuku v těle [%]     |
| • Tep                           | • Síla                           |
| • Glukóza v krvi                | • Rovnováha                      |
| • Hmotnost                      | • BCI                            |

Bližší informace lze nalézt v diplomové práci Pavla Šnejdara [26].

## Opakovatelnost měření

Výše popsaný postup měření se dodržuje při jednorázovém provedení měření, kdy se již neočekává, že by účastník absolvoval experiment znovu. Ale systém BiN umožňuje vytvoření identifikační karty účastníka, která obsahuje jeho unikátní identifikátor v podobě QR kódu. Tato karta umožňuje spárovat opakující se měření s jediným měřeným subjektem a zároveň udržet anonymizovanou podobu naměřených dat.

Další unikátní QR kód se nachází na dokumentu používaném pro sběr metadat. Tento QR kód umožňuje po načtení<sup>5</sup> zobrazit uložená data daného účastníka naměřená v rámci projektového dne.

### 2.1.9 Stanoviště BCI

Data v systému BiN obsahují i naměřená EEG data z různých experimentů<sup>6</sup>, kde sběr EEG probíhá vždy neinvazivní metodou pomocí různých EEG čepic a nebo elektrod. Takto získaná hrubá EEG data se ukládají mimo samotný systém BiN.

#### Příprava měřeného subjektu

Správné nasazení EEG čepice není úplně triviální záležitostí. Subjekt musí být nejdříve řádně očištěn (pomocí abrazivní pasty) na styčných plochách zemnicí i referenční elektrody<sup>7</sup>. Samotný proces čištění bývá pro účastníka často nepříjemný.

Když se účastníkovi nasadí a upevní standardní EEG čepice se systémem 10-20, následuje potření elektrod, potřebných pro daný experiment, vodivým elektrogelem. Poté se propojí zesilovač potřebnou kabeláží a připojí stimulační a nahrávající PC k experimentu. Po tomto kroku je spuštěn samotný stimulační protokol jehož náplň se liší v rámci prováděných experimentů.

Takováto BCI měření probíhají nejen za laboratorních podmínek, ale i nelaboratorních prostorách mimo neuroinformatickou laboratoř. BCI se měřilo i v rámci Dnů vědy a techniky 2018 pořádaných v Plzni, kde se samotné měření BCI provádělo uvnitř stanu na náměstí Republiky.

---

<sup>5</sup>Například pomocí chytrého telefonu s aplikací na čtení QR kódů.

<sup>6</sup>Např. OQ experiment, Hádání čísel, robotické koule, Save the princess, Smart Train a nebo asistenční systém BASIL.

<sup>7</sup>Jedná se například o ušní lalůček pro zemnicí elektrodu a kořen nosu pro referenční elektrodu.

## Asistenční systém BASIL

BCI experiment, který byl použit v rámci Dnů vědy a techniky 2018, vznikl v rámci projektu BASIL, jehož cílem bylo vytvořit asistenční systém pro motoricky postižené účastníky. Hlavním úkolem bylo rozpoznat, jakou z nabízených devíti činností<sup>8</sup> potřebuje provést účastník, který nemůže mluvit a hýbat se. Tvar a obsah stimulační matice použité v projektu BASIL je prezentován na obrázku 2.6.



**Obrázek 2.6:** Stimulační matice použitá v projektu BASIL. Vždy se aktivuje pouze jeden sloupec a nebo řádek této matice. Na obrázku byl aktivován první sloupec matice, který obsahuje následující činnosti v pořadí od shora dolů: otevřít dveře, rozsvítit světlo a jít na toaletu.

Pro zjednodušení popisu má stimulační protokol následující strukturu:

- Účastníkovi je nasazena EEG čepice a spuštěn stimulační protokol s maticí (3x3), která obsahuje celkem devět možných činností, které může chtít účastník vykonávat.
- Úkolem účastníka je vybrat si jednu z nabízených možností a v duchu si je počítat.
- Experimentátor vyhledává v naměřených EEG datech pomocí zprůměrovaných vln P300, na který řádek a sloupec činností účastník myslí.

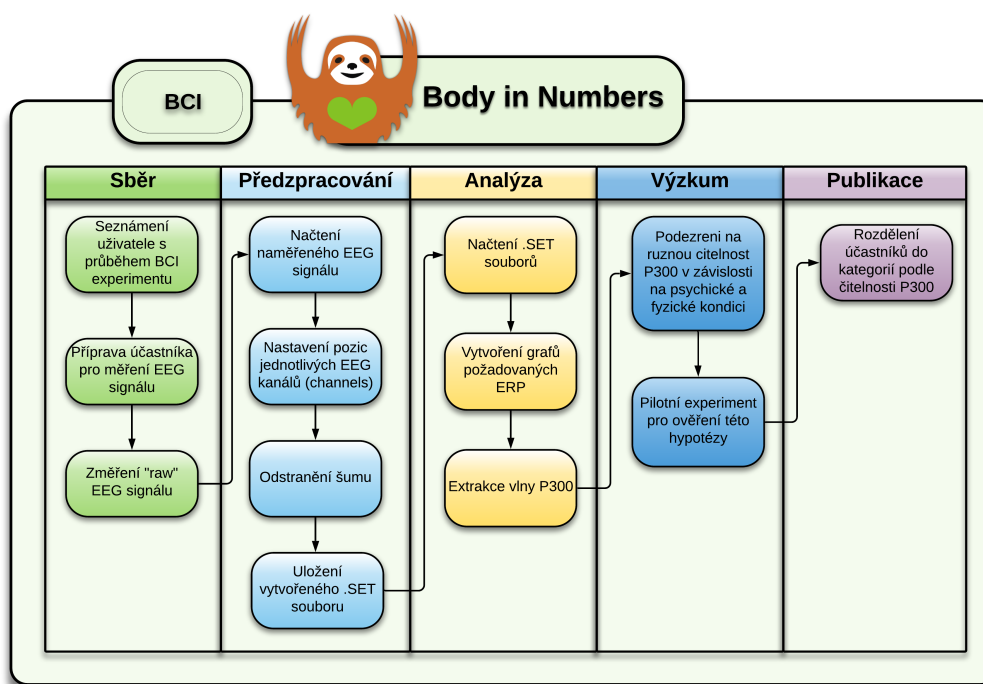
<sup>8</sup>Nabízené činnosti byly: otevřít dveře, najíst se, volat o pomoc, rozsvítit světlo, zvednout telefon, spustit rádio, jít na toaletu, spustit televizi, otevřít okno.

Pomocí získaného řádku a sloupce pak lze odhadnout, o jakou činnost se jedná.

Samotný odhad řádku/sloupce, na který účastník myslel, byl prováděn jak experimentátorem, tak automatickým klasifikátorem.

## Datový cyklus BCI

Datový cyklus stanoviště BCI lze vidět na následujícím obrázku 2.7. Datový cyklus je rozdělen do kategorií sběr, předzpracování, analýza, výzkum a publikace, podobně jako bylo zmíněno na předešlých obrázcích 2.1 a 2.2.



**Obrázek 2.7:** Příklad datového cyklu z pohledu stanoviště BCI. Experiment se zaměřoval primárně na rozdělení účastníků do kategorií v závislosti na citelnosti vlny P300.

### 2.1.10 Průběh měření v systému BiN

Na začátku experimentu musí nejdříve účastník podepsat dokument - Souhlas s účastí v projektu (viz kapitola 2.1.7). Pokud účastník souhlasil s účastí, musí v dalším kroku vyplnit dotazník obsahující metadata.

V tuto chvíli si může účastník vybrat, jaká stanoviště chce absolvovat (viz kapitola 2.1.8).

Doporučené pořadí vypadá následovně:

### 1. Krevní tlak a tep

- Změření krevního tlaku a tepu předtím než se začne účastník namáhat (např. spirometrie, reakční doby).

### 2. Glukóza v krvi

- Zjištění krevního cukru před námahou.

### 3. Spirometrie před zátěží

- Hodnoty VC, FEV1 a PEF<sup>9</sup>.

### 4. Spirometrie po zátěži

- Hodnoty VC, FEV1 a PEF potom co chodil účastník jednu minutu do schodů.

### 5. Reakční doby horních a dolních končetin

- Tyto testy zejména sledují rychlost odpovědi účastníka na zrakové podněty.

### 6. Měření síly

- Měření síly rukou (zvláště pravá a levá) pomocí siloměru v libovolné poloze a v přepažení (v kg).

### 7. Měření rovnováhy

- Účastník se postaví na dvě váhy a sleduje se distribuce hmotnosti na každou nohu (zvláště pravá a levá).

### 8. Barvocit

- Barvocit je měřen pomocí osmi pseudo-izochromatickými tabulkami. Účastník měl za úkol odpovědět jaké číslo se na dané tabulce nachází.

### 9. Měření tělesných proporcí

---

<sup>9</sup>Vitální kapacita plic (VC), časová vitální kapacita (FEV1) a maximální výdechový proud vzduchu (PEF).

- Změření hmotnosti, výšky, BMI, zastoupení svalstva v těle [%], zastoupení vody v těle [%] a zastoupení tuku v těle [%] pomocí specializované váhy.

## 10. Mozková aktivita

- Měří se pomocí BCI stanoviště, které se sestává ze snímače elektrické aktivity ze skalpu účastníka v podobě EEG čepice, zesilovačů, vyhodnocovacího počítače a generátoru stimulů.

## 2.2 Sémantický web

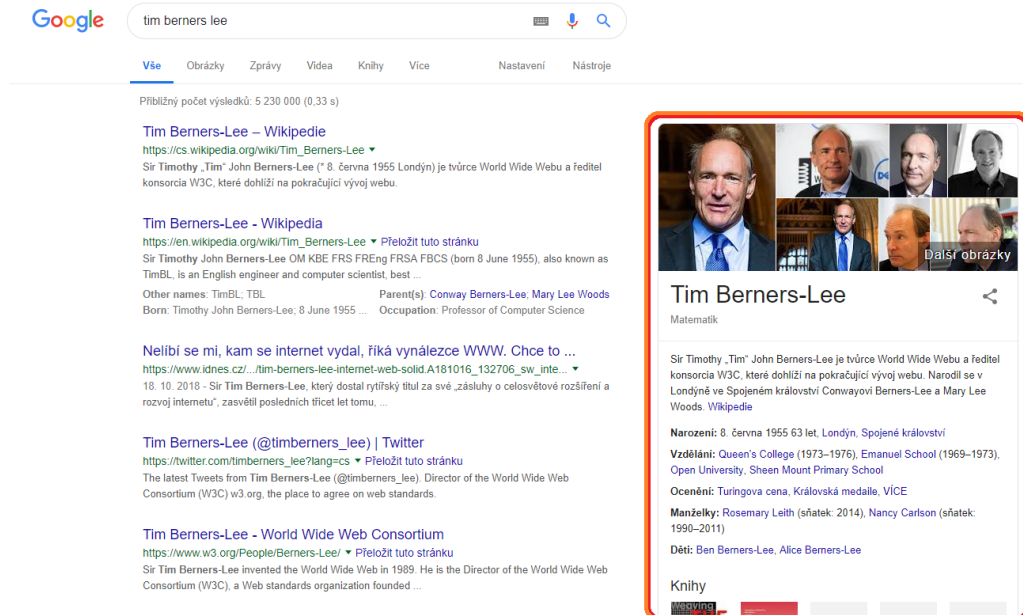
Hlavní myšlenkou sémantického webu je zajistit čitelnost obsahu na internetu nejen lidmi, ale i stroji. Internetové stránky by byly smysluplné pro různá softwarová řešení a zároveň tato data obsahovala sémantiku, která by umožnila programům přistupovat k webům podobným způsobem jako lidé a umožnila by i výměnu a nebo sdílení dat napříč sémantickým webem, aniž by jednotlivé programy byly explicitně navrženy, aby mezi sebou dokázaly komunikovat [11].

Tato vize sémantického webu inspirovala velké množství lidí, a tak jako objevení nové technologie AJAX vedlo k zavedení pojmu **Web 2.0**, tak Tim Berners-Lee označoval sémantický web jako **Web 3.0**. Bohužel velká spousta slibů sémantického webu stále ještě nebyla naplněna.

Původní představa, že by se po sémantickém webu pohybovali inteligentní agenti a byli schopni sami o sobě vykonávat určité funkce a rozhodnutí stále zůstává v pravém slova smyslu nenaplněna [10]. Některé podobné technologie již existují (jako například Siri od firmy Apple), ale oproti původní myšlence sémantického webu umožňuje odpovědět na velice úzce specializovanou kategorii otázek. A to ještě pouze díky tomu, že přímo inženýři Apple manuálně připojili k nemalému množství web-services [12].

Šance, že se původní verze sémantického webu, který navrhnul Tim Berners-Lee, objeví není příliš pravděpodobná. Ale i přesto se vytvořilo a používá mnoho nových technologií a ideí, které původně vznikly během implementace myšlenky sémantického webu jako takového. Například Google používá sémantické technologie (nyní primárně JSON/LD) pro generování konceptuálních shrnutí vpravo od výsledků vyhledávání (viz obrázek 2.8). Webová stránka a služba <https://schema.org/> udržuje seznam sémantických slovníků, která mohou webovými vývojáři využít, aby mohli být jejich data snadno publikována širší veřejnosti. Další moderní aplikací sémantických technologií, a zároveň asi i nejvíce populární a úspěšnou mimo Google,

je OpenGraph protokol od firmy Facebook. OpenGraph protokol definuje schéma, které mohou weboví vývojáři využít pro určení, jak se mají webové stránky zobrazovat, když jsou sdíleny pomocí sociálních sítí [6, 12].



**Obrázek 2.8:** Google využívá sémantické technologie například pro vytvoření konceptuálních shrnutí, která se zobrazují vpravo od výsledků vyhledávání.

## 2.2.1 Linked Data

Linked data (česky „propojená data“) je způsob pro publikování strukturovaných dat, v prostředí webu, která jsou navzájem propojená pomocí odkazů (linků), a zároveň jsou interpretovatelná stroji. Linked data umožňují vyhledávat propojené a jednoznačně definované informace, které mohou být i rozděleny mezi několik různých webových stránek, a tak celkově pomáhají rozšiřovat znalosti pro automatizované agenty<sup>10</sup>. Linked data využívají například technologie jako HTTP (webové stránky), IRI (jednoznačné identifikátory) a nebo RDF.

<sup>10</sup>Jako je například pro vyhledávače jsou nejdříve indexovány webové stránky (pomocí webcrawlerů), apod.



## 2.3 RDF

Jedním se základních standardů, které měly napomoci k rozšířenějšímu používání sémantického webu měl být datový formát RDF, neboli Resource description framework. Jedná se o datový formát, který W3C<sup>11</sup> vytvořila, aby reprezentoval libovolný obecný druh znalostí. RDF by vlastně měl sloužit jako gramatika, pomocí které se vyjádří sémantický web [12].

### 2.3.1 RDF trojice

RDF popisuje obecné znalosti pomocí trojic skládající se z subjektu, predikátu a objektu. Například příkladem RDF trojice může být Subjekt=„Experiment“ (odpověď na otázku KDO), predikát=„používá“ (jakou činnost nebo vlastnost má subjekt) a objekt=„přístroj“ (jakou hodnotu nabývá vlastnost subjektu). Tento názorný příklad lze vidět zakreslený na následujícím obrázku 2.9, kde je i uveden výsledný RDF graf.

### 2.3.2 Formáty RDF (serializace RDF)

Samotná syntaxe RDF může být vyjádřena pomocí velkého množství formátů. Vždy se jedná o stejnou myšlenku trojice subjekt-predikát-objekt, ale pouze s rozdílným způsobem uložení a syntaxí. Mezi některé příklady lze uvést:

- RDF/XML
- Turtle
- N-Triples
- N3
- RDF/JSON
- JSON/LD
- a další...

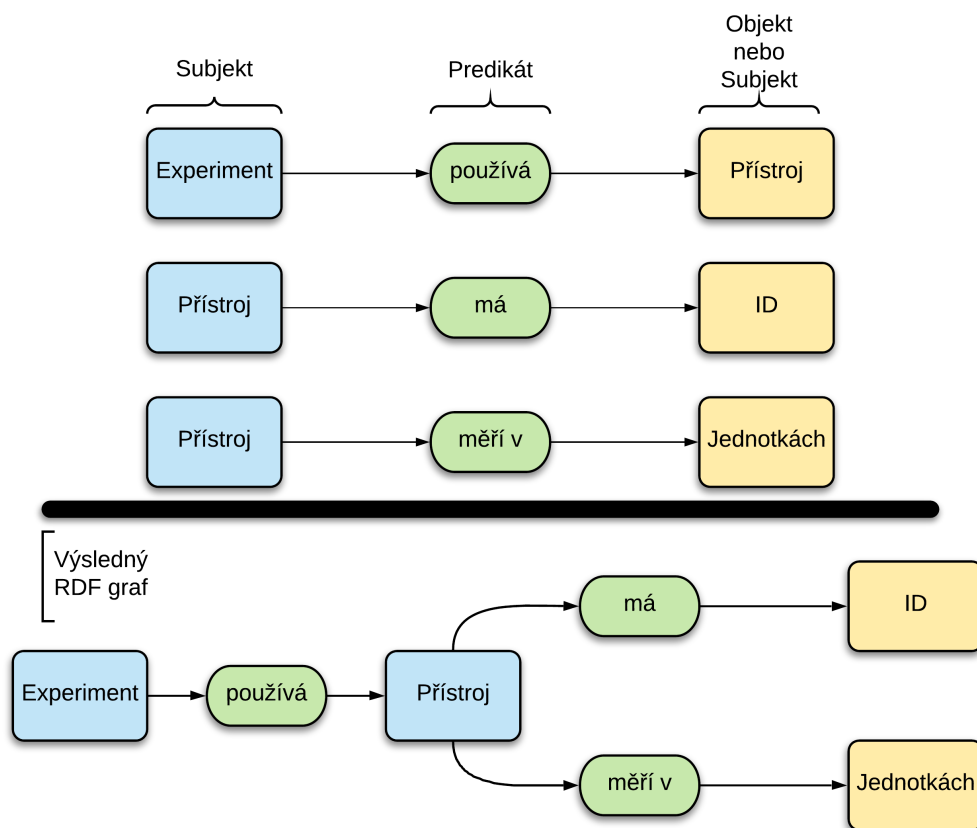
#### RDF/XML

Aby bylo možné zakódovat RDF graf pomocí XML je nutné, aby všechny predikáty a objekty byly reprezentovány pomocí platných XML značek a to:

- Název atributu
- Hodnoty atributu
- Název elementu
- Obsah elementu

---

<sup>11</sup>Mezinárodní standardizační organizace na poli nejen informačních technologií a internetu.



**Obrázek 2.9:** Jednoduchý příklad obsahující tři příklady RDF trojic včetně výsledného RDF grafu.

RDF/XML používá **QNames**, které jsou definované v XML, pro reprezentování IRI<sup>12</sup>. Všechna definovaná **Qnames** obsahují **namespace name**, které je IRI, zkrácené **local name** a případně i krátký prefix. IRI reprezentované pomocí **Qname** je určeno přidáním **local name** části **Qname** za **namespace name**. Tento postup je použit pro zkracování IRI všech predikátů. IRI identifikující subjekty a objekty mohou být také uloženy jako hodnota XML atributu. RDF literaly, které mohou být pouze objekty, se stanou datovým typem XML **element text content** a nebo hodnota XML **attribute** [19].

Graf může být považován za kolekci cest ve formě subjekt, predikát, objekt/subjekt, predikát, a tak dále... , které pokryjí celý graf. Ve formátu RDF/XML jsou tyto cesty převedeny do sekvencí elementů uvnitř elementů, které alternují mezi různými elementy a predikáty. Příklad formátu RDF/XML lze vidět na obrázku 2.10 [19].

<sup>12</sup>Jednotný identifikátor zdroje. Je to textový řetězec s definovanou strukturou, který slouží k přesné identifikaci zdroje informací.

## EXAMPLE 2

### Node Elements with IRIs added

```
<rdf:Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar">
  <ex:editor>
    <rdf:Description>
      <ex:homePage>
        <rdf:Description rdf:about="http://purl.org/net/dajobe/">
          </rdf:Description>
        </ex:homePage>
      </rdf:Description>
    </ex:editor>
  </rdf:Description>
```

Obrázek 2.10: Příklad formátu RDF/XML. Zdroj: Example 2 [19].

## Turtle

Textový formát pro RDF nazvaný **Turtle** umožňuje zapisování RDF grafů pomocí kompaktní a přirozené textové podoby, který umožňuje i používat zkratky a datové typy. **Turtle** také umožňuje kompatibilitu pro formát **N-triples** [18]. Tento datový formát také podporuje:

- **Relativní IRI** - @base
- **Prefixi a prefixová jména** - @prefix
- **Seznamy predikátů** - oddělených od sebe středníkem ';'
- **Seznamy objektů** - oddělených od sebe čárkou ','
- **Token 'a'** - který nahrazuje IRI  
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
- **Literály**

## JSON/LD

**JSON/LD** (JavaScript Object Notation for Linked Data) je jednoduchá syntaxe pro serializaci **Linked Data** pomocí **JSON** formátu. **JSON/LD** umožňuje s minimálním množstvím změn, aby již existující **JSON** dokumenty byly interpretovány jako **Linked Data**. Formát **JSON/LD** je primárně zamýšlen, aby bylo možné používat **Linked Data** v prostředí webových aplikací, vytváření zaměnitelných webových služeb a ukládání **Linked Data** do **JSON** databází. Vzhledem k tomu, že **JSON/LD** je plně kompatibilní s formátem **JSON**, tak je možné znovu využít velké množství **JSON** parserů a nebo **JSON**

### EXAMPLE 1

```
@base <http://example.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rel: <http://www.perceive.net/schemas/relationship/> .

<#green-goblin>
  rel:enemyOf <#spiderman> ;
  a foaf:Person ;      # in the context of the Marvel universe
  foaf:name "Green Goblin" .

<#spiderman>
  rel:enemyOf <#green-goblin> ;
  a foaf:Person ;
  foaf:name "Spiderman", "Человек-паук"@ru .
```

**Obrázek 2.11:** Příklad RDF formátu Turtle. Každá z RDF trojic je ukončena tečkou. Tato vzorová data popisují vztah mezi Green Goblinem a Spidermenem. Zdroj: Example 1 [18].

knihoven, které jsou již nyní dostupné [17]. Navíc ke všem vlastnostem, které umožňuje samotný JSON, přidává JSON/LD následující vlastnosti:

- Mechanismus jednoznačné identifikace JSON objektů pomocí IRI
- Způsob, jak rozlišit klíče, které jsou sdíleny napříč různými JSON dokumenty, které se namapují na IRI pomocí *contextů*
- Mechanismus, který umožňuje předat hodnotu JSON objektu tak, že se odkáže na JSON objekt na jiné webové stránce
- Vlastnost anotovat *stringy* jejich jazykem
- Způsob jak asociovat datové typy s hodnotami jako jsou například datum a čas
- Způsob jak vyjádřit jeden nebo více orientovaných grafů, jako jsou například sociální sítě a nebo samostatný dokument

JSON/LD byl navrhnout, aby byl přímo používán jako JSON, bez jakékoliv předešlé znalosti RDF. A v případě potřeby je možné ho použít i s dalšími *Linked Data* technologiemi jako jsou například SPARQL. Jaký je rozdíl mezi JSON a JSON/LD možné posoudit na následujících obrázcích 2.12 a 2.13.

#### EXAMPLE 1: Sample JSON document

```
{
  "name": "Manu Sporny",
  "homepage": "http://manu.sporny.org/",
  "image": "http://manu.sporny.org/images/manu.png"
}
```

**Obrázek 2.12:** Příklad JSON dokumentu obsahujícím jméno, homepage a obrázek. Zdroj: Example 1 [17].

#### EXAMPLE 2: Sample JSON-LD document using full IRIs instead of terms

```
{
  "http://schema.org/name": "Manu Sporny",
  "http://schema.org/url": { "@id": "http://manu.sporny.org/" },
  "http://schema.org/image": { "@id": "http://manu.sporny.org/images/manu.png" }
}
```

**Obrázek 2.13:** Příklad upraveného JSON/LD dokumentu. Změny oproti JSON dokumentu jsou zvýrazněny. Zdroj: Example 2 [17].

## 3 OWL ontologie

OWL ontologie jsou formalizované slovníky používaných termínů, které používá komunita uživatelů a souvisí se specifickou doménou, ve které se tato komunita pohybuje. Specifikují definice používaných termínů pomocí zachycení vztahů a závislostí mezi jednotlivými termíny obsažených v ontologii. Hlavním důvodem používání ontologií je učinit obsah webu mnohem přístupnějším a čitelnějším pro stroje [20].

Ontologie se popisují pomocí tříd, vlastností a jejich vztahů uvnitř specifické aplikační domény<sup>1</sup>. Poslední specifikace jazyka OWL proběhla v roce 2012, kdy byl schválen standard pod názvem OWL 2, který rozšířil OWL a odstranil některá funkční omezení či nedostatky. Těmito přidanými funkcemi byly například přidání klíčů, řetězení vlastností, bohatší výběr datových typů (oproti OWL 1), vylepšené možnosti notací a další [25].

Aby v budoucnu nedocházelo k opakovanému vynalézání kola, byly vytvořeny specializované portály pro ukládání a prohledávání ontologií. Tyto portály umožňují prohledat různě navržené ontologie a hledat již nadefinované termíny<sup>2</sup>, které lze poté vložit do nově vytvářených ontologií. Takto se zajistí stejný význam daného termínu včetně souvislostí, které byly již v minulosti nadefinovány.

### 3.1 Podjazyky OWL a OWL 2

OWL 1 i OWL 2 jazyky lze rozdělit do tří kategorií podjazyků. Pro OWL 1 se jedná o *OWL-Lite*, *OWL-DL* a *OWL-Full*. Hlavním rozdílem mezi jednotlivými podjazyky OWL 1 je rozdíl v jejich vyjadřovacích možnostech (expresivitě) pro použití ve specifických ontologických komunitách. Vyjadřovací možnosti jednotlivých OWL 1 podjazyků rostou v tomto pořadí od nejmenší po nejvyšší úroveň expresivity [16].

Vzhledem k tomu, že v roce 2012 byl publikován nový vylepšený standard OWL 2, byly standardizovány nové typy OWL 2 jazyků a i jejich profily. OWL 2 profily jsou v tomto ohledu podobné OWL 1 podjazykům.

V OWL 2 existují dva alternativní způsoby, jak přiřadit význam k definovanému výrazu a to buď pomocí přímé sémantiky (*direct semantics*) a nebo sémantice na bázi RDF (*RDF-based semantics*). Obě tyto sémantiky

---

<sup>1</sup>Např. medicína, biologie, astronomie, obrana a další.

<sup>2</sup>Termíny mohou být například, jméno, příjmení, osobní počítač, počasí, . . . Může se jednat i o vlastnosti jako například *isConnected*, *hasPart*, *hasParticipant*, . . .

jsou si velice podobné, ale existují dva hlavní rozdíly. Pod přímou sémantikou anotace nemají žádný formální význam a pod RDF sémantikou vznikají přidružená odvození, která vyplývají z RDF pohledu na svět. Konceptně můžeme uvažovat o rozdílech mezi OWL 2 DL (přímá sémantika) a OWL 2 Full (sémantika na bázi RDF) [14].

- **OWL 2 DL** – Na první pohled se může OWL 2 DL zdát jako syntakticky omezená varianta OWL 2 Full, kde jsou jednotlivá omezení navržena, aby usnadnila život implementátorům. Vzhledem k tomu, že je jazyk OWL 2 Full ve skutečnosti uveden ve výpočetní teorii jako nerozhodnutelný problém, tak to umožňuje v OWL 2 DL napsání vyhodnocovače (resonerů), který by v zásadě byl schopen vrátit odpovědi na otázku ve formě ano či ne. Díky tomuto designu jazyka již existuje několik kvalitních vyhodnocovačů, které pokrývají celý jazyk OWL 2 DL za použití přímé sémantiky. V současné chvíli neexistují OWL 2 Full vyhodnocovače [14].
- **OWL 2 Full** – Na OWL 2 Full je možné se dívat jako na nejvíce přímočaré rozšíření RDFS. Sémantika založená na RDF v rámci OWL 2 Full následuje sémantiku RDFS a její obecnou syntaktickou filosofii (tj. vše v RDF jsou trojice a RDF jazyk je plně reflexivní<sup>3</sup>) [14].

Kromě jazyků OWL 2 DL a OWL 2 Full, tak OWL 2 specifikuje další tři profily. OWL 2 je sám o sobě velmi expresivní jazyk (jak výpočetně, tak uživatelsky), a tak může být problém s ním pracovat a nebo ho i implementovat. Tyto dodatečné profily usnadňují přístup k podmnožinám jazyka OWL 2, které jsou dostatečné pro celou řadu aplikací. Stejně tak jako OWL 2 DL jsou i na profily kladeny požadavky na výpočetní složitost. A tyto profily je mnohem snadnější implementovat a jsou zároveň robustní a škálovatelné vzhledem k stávající úrovni technologie, ale i tak existuje mnoho podmnožin OWL 2, které mají dobré výpočetní vlastnosti sami o sobě [14].

Tyto vybrané OWL 2 profily byly identifikovány, že již obsahují rozsáhlé komunity uživatelů, a aby bylo možné garantovat škálovatelnost, tak stávající profily sdílejí určitá omezení týkající se jejich expresivity. Obecně se zákazu používání operací negace a disjunkce, protože tyto konstrukce komplikují jejich vyhodnocování a pro potřeby modelování byly jen velmi zřídka potřebné [14].

---

<sup>3</sup>Ve výpočetní technice znamená reflexe schopnost počítačového programu prozkoumat a modifikovat svou strukturu a chování za běhu.

### 3.1.1 OWL 2 EL

Profil OWL 2 EL je zejména užitečný v aplikacích, které využívají ontologie s velmi velkým množstvím definovaných vlastností a/nebo tříd. Tento profil zachycuje expresivitu používanou mnohými ontologiemi a zároveň spadá do podmnožiny OWL 2, kde je možné vyhodnotit základní rozhodovací problémy v polynomiálním čase s ohledem na velikost ontologie. Pro tento profil jsou k dispozici vyhodnocovací algoritmy, u kterých bylo prokázáno, že jsou implementovány vysoce škálovatelným způsobem. Akronym EL odráží fakt, že tento profil je založen na základech EL deskriptivní logiky<sup>4</sup>(EL description logic), tj. logika, která poskytuje pouze existenciální kvantifikaci (Existential Quantification) [15].

### 3.1.2 OWL 2 QL

Profil OWL 2 QL se zaměřuje na aplikace, které používají velmi velké množství instancovaných dat a kde dotazování se na ukládaná data je nejčastějším vyhodnocovacím úkonem. V tomto profilu mohou být implementovány dotazy v konjunktivním tvaru za použití konvenčních relačních databázových systémů. Podobně jako v profilu OWL 2 EL mohou být použity polynomiální algoritmy pro řešení problémů typu: ověřování konzistence ontologie a nebo vyhodnocování problému subsumce tříd. Míra expresivity tohoto profilu je nutně velmi omezená, ačkoliv obsahuje většinu hlavních funkcí konceptuálního modelování jako například diagramy UML tříd a nebo ER diagramy. Akronym QL v tomto případě odráží fakt, že dotazování může být v tomto profilu implementováno pomocí přepsání dotazů do standardního relačního dotazovacího jazyka (Query Language) [15].

### 3.1.3 OWL 2 RL

OWL 2 RL se zaměřuje na aplikace, které vyžadují škálovatelné vyhodnocování aniž by došlo k příliš velkému omezení expresivity jazyka. Tento profil je navržen, aby vyhovoval aplikacím, které mohou obětovat plnou expresivitu jazyka OWL 2 za efektivitu jeho zpracování, a také pro RDF(S) aplikace, které potřebují trochu přidané expresivity. Vyhodnocovače profilu OWL 2 RL mohou být implementovány pomocí vyhodnocovacích nástrojů založených na pravidlech. Například problémy typu: ověřování konzistence určité ontologie, ověřování splnitelnost výrazů, které využívají třídy, kontrola instancí

---

<sup>4</sup>Více informací o EL deskriptivní logice lze nalézt na <https://lat.inf.tu-dresden.de/research/papers/2005/BaaderBrandtLutz-IJCAI-05.pdf> a pro EL++ na [http://webont.org/owlled/2008dc/papers/owlled2008dc\\_paper\\_3.pdf](http://webont.org/owlled/2008dc/papers/owlled2008dc_paper_3.pdf)



tříd a hledání odpovědi na typ otázek v konjunktivním tvaru, mohou být vyřešeny v polynomiálním čase, samozřejmě s ohledem na velikost ontologie. Akronym RL odráží fakt, že vyhodnocovač tohoto profilu je možné implementovat pomocí standardních pravidlového jazyka (Rule Language) [15].

## 3.2 Ontologické portály

Jedná se webové portály se zaměřením na publikování a prohledávání biomedicínských ontologií a terminologií. Tyto portály pak umožňují komunitní spolupráci na vyhodnocování a vývoj ontologického obsahu v oblasti biomedicíny tím, že umožňují nezávisle na použitých ontologických datových formátech .OWL a .OBO<sup>5</sup> vyhledávání již nadefinovaných termínů, včetně výtahu souvisejících metadat daných termínů.

- BioPortal[22]
- OntoFox[23]
- Ontology Lookup Service[7]
- OntoBee[24]
- LOV[13]

Ontologický portál	Počet uložených ontologií
BioPortal	761
LOV	664
OLS	226
OntoFox	223
OntoBee	191

**Tabulka 3.1:** Přehled celkem uložených ontologií v jednotlivých ontologických portálech. Některé ontologie jsou uloženy i ve více portálech najednou. Příklad takovéto ontologie může být NCIT (National Cancer Institute Thesaurus), kde tato ontologie je uložena ve všech uvedených portálech. Uvedené počty jsou vztaženy k datu 2019-05-05 (RRRR-MM-DD).

## 3.3 Ontologické dobré mravy

V současné chvíli existuje poměrně velké množství doporučení a názorů na to co všechno by měla užitečná ontologie obsahovat. Bohužel neexistuje jed-

<sup>5</sup>Biomedical ontologies, v současné chvíli se považuje za součást .OWL datového formátu.

noznačný názor na to co všechno by bylo vhodné v ontologii u definovaných termínů zmínit.

Další problém s touto různorodostí spočívá v nazývání totožných termínů pomocí synonym. Může se jednat o funkčně totožné termíny (např. termín, který se používá pro označení autora definice), se označuje v každé ontologii jiným způsobem.

A tak se některé organizace snaží zavést do tvorby ontologií určitý pořádek a nebo standard. Mezi tyto organizace lze zařadit například **OBO Foundry** [8], **Open Semantic Framework** [9] a **W3C** [21]. Mezi hlavní body své standardizace řadí mimo jiné i pravidla pro importování již existujících termínů z jiných ontologií, názvosloví nových termínů, a nebo nutnost uvádět textové definice<sup>6</sup> u každého termínu v ontologii.

Mimo jiné bylo prozkoumáno i několik již publikovaných ontologií, aby se ověřila platnost těchto doporučení a případně i odhalily jiné postupy při řešení ontologií. Výťah některých důležitých bodů je uveden v následujících bodech.

- **Určení rozsahu nové ontologie** – Definice, kde, kdo a jak bude ontologie využívat a snažit se během návrhu novou ontologii zaměřit na uspokojení těchto potřeb.
- **Učit se z již existujících ontologií** – V případě, že s vytvářením ontologií nemá tvůrce dostatečnou zkušenost, je vhodné projít některé již existující a publikované ontologie ve stejném oboru. V tomto případě je důležité se zaměřit na způsob, jakým jsou termíny definovány, jaké anotace jsou využívány případně i na formální řád a štábní kulturu celé ontologie.
- **Zvolení vhodných anotací (annotation properties)** – I v tomto případě je vhodné maximalizovat re-use. Vhodné jsou například anotace z ontologie IAO (Information Artifact Ontology)<sup>7</sup>. Případně pokud žádné z existujících definic anotací nevyhovují, tak lze vytvořit vlastní.
- **Název termínu** – Nejlépe uvést název v plain english, nepoužívat CamelCase ani podtržítkovou\_ notaci. Pokud existují synonyma, která se mohou použít ve stejném významu jako původní termín, tak je

---

<sup>6</sup>Nejčastěji v angličtině, ale je možné přidat i několik textových definic pro různé jazyky, aby se maximalizovalo znovupoužití nedefinovaného termínu.

<sup>7</sup>Vhodné anotace jsou například term editor, textual definition, has curation status, alternative term a další.

vyjmenovat (i včetně zkratk) <sup>8</sup>. Pokud možno, tak využít `rdfs:label` jako primární název termínu.

- **Stanovení identifikátor termínu (IRI)** – Pokud se znovu využívají (re-use) již naimportované definice z jiné ontologie ve stejném významu, tak se zachová původní IRI, jinak se vytvoří vlastní.
- **Vlastní IRI termínu** – Unikátní identity v prostředí internetu nejlépe lomítková notace. Například ID termínu může vypadat následovně `http://purl.bioontology.org/ontology/akronym ontologie/identifikátor termínu`. Existují organizace, které dokážou PURL (Persistent URL) vygenerovat a přiřadit, které zajistí onu unikátnost. Samotný identifikátor termínu by se měl skládat z akronymu ontologie a číselného identifikátoru (Např. MESH\_00000123).  
Výsledné IRI nadefinovaného termínu může vypadat následovně:  
`http://purl.bioontology.org/ontology/MESH/MESH_00000123`  
a nebo  
`http://purl.obolibrary.org/obo/OBI_0000185`.
- **Textová definice termínu** – Textová definice termínu stanovuje popis k čemu se termín využívá nebo slovní vymezení. Nejčastěji psané v angličtině.
- **Importování externích termínů** – V případě, že se jedná o stejné využití termínů, tak musí být jak IRI, tak atributy podobné. Výjimku mohou tvořit uvedené pokyny pro importování termínu v anotaci `rdfs:comment` u zdrojové ontologie, ve kterém se občas uvádějí.
- **Otevření ontologie veřejné komunitě** – Kolaborativní spolupráce ontologické komunity může vést k zvýšení kvality samotné publikované ontologie.
- **Uvedení licence k dané ontologii** – Pro většinu otevřených ontologií se jedná o licenci Creative Commons v aktuální verzi. Pro placené ontologie licence specifikuje, kdy ji lze využít, v jakých projektech a nebo jak získat k ontologii přístup.
- **Použitá serializace výsledného .OWL souboru** – Nejčastěji se používají OWL (OWL 1 formát), RDF/XML (OWL 2 formát) a nebo OBO (OBO Foundry formát).

---

<sup>8</sup>Například termín celsius (jednotka tepla), tak synonyma mohou být C, °C, degree celsius

- **Budovat ontologii postupně (inkrementálně)** – Vzhledem k tomu, že rozpočet a čas jsou omezené a konečné veličiny, tak mohou být zkonstruované ontologie vytvořené v menším rozsahu a nebo nadefinované termíny nejsou dostatečně objasněny vůči původnímu plánu. Proto postupné a malé přírůstky mohou přinést mnohem lepší pragmatické výsledky.

### 3.3.1 Postup při vytváření ontologie popisující systém Body in Numbers (BIN)

Postup, který byl při vytváření ontologie BIN dodržován, vycházel zejména z doporučených postupů, které jsou uvedeny v předchozí kapitole [8, 9, 21].

V úvodní fázi návrhu nové ontologie byl vytvořen slovník, ve kterém je potřeba sledovat všechny termíny v prostředí systému BiN. Zejména se jednalo o termíny, které souvisí s procesem sběru dat a nebo jsou následně používána během jejich zpracovávání. Ke každému z těchto termínů, byl přiřazen prvotní anglický název a krátká slovní definice termínu v češtině.

Dalším krokem bylo rozdělit tyto termíny do kategorií tříd a anotací, tj. které termíny budou sloužit k popisu zbylých termínů (např. *autor definice*), a na třídy které budou popisovány pomocí anotací (např. *siloměr*, který bude mít specifikovaného *autora definice*).

Po této fázi bylo nutné maximalizovat re-use z již existujících ontologických zdrojů, aby nedocházelo k opakované definici významově stejných termínů. Pro tento účel byly používány právě ontologické portály (viz kapitola 3.2), ve kterých bylo možné podobné třídy a nebo anotace termínů vyhledávat. V následujícím kroku se postupovalo odlišně podle toho, zda termín již existoval ve stejném nebo podobném významu uvnitř již vy publikované ontologie, či zda samotná definice termínu byla dostupná v jiné formě<sup>9</sup>. Pokud se termín nepodařilo nalézt byla vytvořena nová definice termínu.

To jak vypadal výsledný dokument, s nadefinovanou terminologií lze nalézt v příloze E.

Výsledná ontologie byla vyvíjena v programu Protege verze 5.5. Během vytváření byla také ověřována konzistence (consistency) a souvislost (coherency) mezi nadefinovanými termíny pomocí vyhodnocovačů HermiT (1.4.3.456), Pellet a FaCT++ (1.6.5). Výsledné ontologie byla také ověřena a úspěšně publikována na webové stránce BioPortal.

<sup>9</sup>Například na webových stránkách, stránkách online slovníků (oxfordský, merriam webster, ...).

## Postup při přidávání již nadefinovaného termínu

Pokud se podařilo nalézt termín uvnitř publikované ontologie, který byl významově podobný české definici termínu, byly veškeré anotační vlastnosti tohoto termínu převzaty. V případě, že by bylo potřeba daný termín v určitém smyslu upravit, tak by podle dobrých ontologických mravů neměl tento upravený termín používat původní IRI převzatého termínu.

V tomto případě byly převzaty původní atributy termínů, které byly pouze popsány pomocí jiných anotačních vlastností se stejným významem. Například, když původní termín obsahoval definici termínu uvnitř anotace s názvem `description`<sup>10</sup>, tak byla tato definice přepsána do anotační vlastnosti `textual definition`<sup>11</sup>. Zároveň byl zachován zdroj, ze kterého byl termín převzat<sup>12</sup>.

## Postup při přidávání nového termínu

V případě, že termín nebyl v požadovaném významu nadefinován v některé z již publikovaných ontologií, bylo potřeba vytvořit novou definici termínu. Ještě předtím, než by se však přistoupilo k tvorbě oné vlastní definice, byly prohledávány i jiné možnosti, které by mohly požadovanou definici obsahovat. V tomto případě jde zejména o výkladové slovníky, které obsahují definice termínů, nebo jiných zdrojů, obsahující nadefinované pojmy.

Pokud se podařilo nalézt nějaký externí zdroj definice, byly tyto informace zahrnuty do ontologie. Jednalo se zejména o uvedení odkazu na zdroj definice (pomocí anotace `definition source`) ve formě URL.

Když se i přes všechno úsilí nepodařilo nalézt vhodnou definici termínu, byla tato definice vytvořena pomocí vlastních slov a zdroj definice v tomto případě byl uveden jako osoba (`PERSON:`).

### 3.3.2 Používané anotace uvnitř BiN ontologie

Kde to bylo možné, byly vyplněny všechny následující informace a uvedeny v následujících bodech. V případě, že byly převzaty termíny z již publikovaných ontologií, byly převzaty pouze ty důležité informace, pro které existovala anotace ve stejném významu uvnitř ontologie BiN.

<sup>10</sup>Která je definována uvnitř ontologie Dublin Core. Zdroj této anotace lze nalézt na <http://purl.org/dc/elements/1.1/description>.

<sup>11</sup>Anotace uvnitř ontologie IAO. Zdroj této anotace lze nalézt na [http://purl.obolibrary.org/obo/IAO\\_0000115](http://purl.obolibrary.org/obo/IAO_0000115)

<sup>12</sup>Zdroj definice termín byl uveden v anotaci `definition source` a také byl poskytnut původní odkaz na nezměněnou definici termínu ve formě IRI na specifickou `.owl` ontologii.

V ontologii BiN se většinou vyskytují následující anotační vlastnosti, které byly převzaty z již publikovaných ontologií:

- **Alternative term** – Označení názvů termínů, které lze použít jako synonymum k definovanému termínu. Může se jednat i o zkratky či slova v podobném významu.
  - IRI – [http://purl.obolibrary.org/obo/IAO\\_0000118](http://purl.obolibrary.org/obo/IAO_0000118)
- **Creator** – Jméno tvůrce ontologie, může se jednat i o výzkumnou skupinu.
  - IRI – <http://purl.org/dc/elements/1.1/creator>
- **Comment** – Obecný komentář, který byl využíván pro uvedení dodatečných informací o termínu (např. použití nebo způsob využití). Tyto informace nejsou nezbytně nutné k pochopení definice termínu, ale pomáhají objasnit jeho používání.
  - IRI – <http://www.w3.org/2000/01/rdf-schema#comment>
- **Contributor** – Osoba, instituce či výzkumná skupina, která je zodpovědná za přidávání obsahu v rámci ontologie. Zejména se jedná o spolupracovníky autora ontologie (viz creator).
  - IRI – <http://purl.org/dc/elements/1.1/contributor>
- **Definition editor** – Jméno autora, který je zodpovědný za přidání daného termínu do ontologie. Pokud se jedná o převzatý termín, jsou uvedeni původní autoři termínu z převzaté ontologie.
  - IRI – [http://purl.obolibrary.org/obo/IAO\\_0000117](http://purl.obolibrary.org/obo/IAO_0000117)
- **Definition source** – Obsahuje zdroj, odkud pochází textová definice termínu. Může se jednat o jinou ontologii, odkaz na webovou stránku s definicí a nebo o jméno autora definice.
  - IRI – [http://purl.obolibrary.org/obo/IAO\\_0000119](http://purl.obolibrary.org/obo/IAO_0000119)
- **Editor preferred term** – Preferovaný název termínu upřednostňovaný jeho autorem (viz definition editor).
  - IRI – [http://purl.obolibrary.org/obo/IAO\\_0000111](http://purl.obolibrary.org/obo/IAO_0000111)
- **Example of usage** – Příklad věty a nebo činnosti, při které lze termín využít.

- IRI – [http://purl.obolibrary.org/obo/IAO\\_0000112](http://purl.obolibrary.org/obo/IAO_0000112)
- **Label** – Výchozí nebo běžný název termínu.
  - IRI – <http://www.w3.org/2000/01/rdf-schema#label>
- **Has curation status** – Určení stavu, ve kterém se definice daného termínu nachází. Tyto úrovně vycházejí ze stavů termínů, které jsou definované uvnitř ontologie IAO. Příkladem těchto stavů může být: Metadata complete (metadata termínu jsou nadefinovaná, ale termín se může přemístit na jiné místo v ontologii) nebo Ready for release (termín již obsahuje všechna metadata a nebude se uvnitř ontologie přemísťovat). Tyto stavy jsou popsány pomocí IRI symbolizujících tyto stavy (např. pro ready for release – [http://purl.obolibrary.org/obo/IAO\\_0000122](http://purl.obolibrary.org/obo/IAO_0000122)).
  - IRI – [http://purl.obolibrary.org/obo/IAO\\_0000114](http://purl.obolibrary.org/obo/IAO_0000114)
- **Imported from** – Zde se uvádí zdroj převzatého termínu, ve formě odkazu na původní .owl soubor obsahující původní ontologii, ze které se termín převzal. Například se mohlo jednat o zdroj termínu fahrenheit což obsahuje ontologie <http://purl.obolibrary.org/obo/uo.owl>.
  - IRI – [http://purl.obolibrary.org/obo/IAO\\_0000412](http://purl.obolibrary.org/obo/IAO_0000412)
- **Relation** – Tato anotace obsahuje odkaz na jiný termín uvnitř ontologie BiN, se kterým nově nadefinovaný termín souvisí.
  - IRI – <http://purl.org/dc/elements/1.1/relation>
- **Rights** – Anotace ontologie, která obsahuje licenci a nebo práva, za kterých lze ontologii nebo její obsah využít. V případě BiN ontologie se jedná o odkaz na licenci Creative Commons ve verzi 4.0 (<https://creativecommons.org/licenses/by/4.0/>).
  - IRI – <http://purl.org/dc/elements/1.1/rights>
- **Title** – Oficiální název ontologie. Může se jednat i o její oficiální zkratku.
  - IRI – <http://purl.org/dc/elements/1.1/title>
- **VersionInfo** – Obsahuje dodatečné informace o dané verzi ontologie. Může se jednat o číslo verze či datum publikace dané verze.
  - IRI – <http://www.w3.org/2002/07/owl#versionInfo>

## 4 RDF datová úložiště

RDF lze ukládat mnoha různými způsoby, ale je nutné nejdříve posoudit k čemu se bude dané úložiště převážně používat. Základním typem ukládání jsou `triple stores` (úložiště trojic) nebo `graph database` (grafové databáze). `Triple store` ukládají přímo RDF trojice do své databáze a používají se primárně k rozhodování. Na druhé straně grafové databáze vytvářejí přímo grafy závislostí a používají se primárně pro dotazování a nebo specifické grafové operace<sup>1</sup> [5].

### 4.1 Realizace konverzního algoritmu z CSV na RDF

Zadáním pro následující část práce bylo vytvořit konverzní skript nebo program, který by dokázal převést vstupní data ze systému BiN (JSON nebo CSV), do některého z výstupních formátů RDF dat.

Důvodem pro tuto konverzi bylo převést data ze systému BiN do takového publikovatelného formátu, který se často používá na poli biomedicíny. Běžnými standardy jsou v tomto případě RDF formáty `RDF/XML`, `JSON-LD`, `Turtle` a `N-Triples`. Od původního návrhu, že by se jednalo o přímou konverzi z formátu JSON do RDF, bylo upuštěno z důvodu, že data, která se vyexportují ze systému BiN je často nutné ještě předzpracovat předtím, než dojde k jejich samotné publikaci. Může se v tomto případě jednat jak o čištění, předzpracování, tak i o případnou anonymizaci a nebo pseudonymizaci dat. Úprava formátu JSON je v tomto případě poměrně nepohodlná, a tak byl zvolena druhá možnost, že samotná data budou nejdříve exportována do formátu CSV, který lze například otevřít uvnitř tabulkových procesorů<sup>2</sup> pro poměrně snadnou editaci.

A právě v rámci tohoto řešení byl vytvořen konverzní program, který využívá formát CSV pro konverzi do některého z formátů RDF.

---

<sup>1</sup>Jako například hledání nejkratší cesty.

<sup>2</sup>Jedním ze známějších tabulkových procesorů je například kancelářský balík od společnosti Microsoft Office Excel, OpenOffice Calc a nebo LibreOffice Calc.



### 4.1.1 Testování dostupných řešení

Vstupní formát CSV byl zvolen z toho důvodu, že samotný systém BiN již tento export umožňuje a zároveň toto již implementované řešení funguje správně. A proto byly pro tento problém vyhledávány již vytvořené a odzkoušené nástroje, které byly vytvořeny v programovacím jazyce Python, protože by umožňovaly snazší začlenění výsledného řešení do webové aplikace BiN. Některými z těchto nástrojů byly například následující knihovny a programy:

- **CSVWlib** – Tento nástroj obsahuje třídu CSVWConverter, která umožňuje převedení datového formátu CSV on the Web (CSVW) buď do formátu JSON a nebo RDF s využitím stejnojmenného W3C standardu (dostupného na <https://w3c.github.io/csvw/>). Jednalo se o běžný CSV soubor, který ale byl dostupný na veřejné nebo privátní URL adrese. Pro tento případ se nejednalo o vhodné řešení, protože by byla nutnost vždy daný soubor umístit na privátní a nebo veřejnou URL adresu, než bude možné tento soubor zkonvertovat do některé ze serializací RDF.
- **TabLinker v0.2** – Je experimentální software, který dokáže konvertovat anotovaný excelovský formát `.xls` a `.xlsx` do RDF v serializaci N3. RDF formát `Turtle` je také částečně podporován. Nevýhodou tohoto nástroje je jeho složité nastavování, jediným podporovaným vstupním formátem je specifický excelovský datový formát, a také fakt, že byl vytvořen pro zastaralou verzi Pythonu 2.7, když k aktuálnímu dni vytváření této diplomové práce existuje verze 3.7.2).
- **RDFLib** – Specializovaná Python knihovna určená pro práci se sémantickým webem a RDF daty, která podporuje mnoho serializací i specifických parserů. Tato knihovna také umožňuje chovat se k RDF datům jako ke grafovým problémům, řešit perzistentní i paměťové ukládání dat a obsahuje i implementaci SPARQL dotazovacího rozhraní. Nevýhoda v tomto případě by byla potřeba navrhnout a vytvořit vlastní implementaci v Pythonu s využitím RDFLib pro konverzi z datového formátu CSV na RDF.
- **CSV2RDF** – Tento nástroj byl vytvořený skupinou Semantic Computing Research Group (SeCo)<sup>3</sup> a umožňuje konverzi vstupního CSV

---

<sup>3</sup>Nástroj CSV2RDF je dostupný na následující adrese <https://github.com/SemanticComputing/CSV2RDF>. Hlavním autorem je Mikko Koho a nástroj spadá pod MIT licenci.

souboru do některého z formátů RDF. Tento program také umožňuje definovat některé užitečné parametry jako jsou například typ kódování vstupního CSV souboru<sup>4</sup>, typ výstupního souboru<sup>5</sup>, definovat oddělovací znak pro CSV, definovat znak pro označení NaN hodnot, znaků uvozující komentáře a dalších vlastností. Samotný konverzní program je napsaný v Pythonu pro verzi 3.0. Bohužel tento nástroj by bylo potřeba ještě přizpůsobit, aby splňoval všechny požadované vlastnosti pro výslednou konverzi dat, ale i tak byl právě kvůli mnoha již podporovaným funkcionalitám vybrán.

### 4.1.2 CSV2RDF modul

Původní program CSV2RDF používal pro konverzi z CSV na RDF následující postup. První řádek CSV souboru obsahoval popis všech atributů, které budou následně převedeny na anotační vlastnosti OWL a budou využity pro přiřazení naměřených hodnot k pseudonáhodně vygenerovaným identifikátorům jednotlivých řádků. Tento nástroj je vytvořen v programovacím jazyce Python pro verzi 3.0.

Jeden z problémů byl obsažen ve způsobu převodu názvu atributu na anotační vlastnost. Během tohoto postupu byla využívána Python knihovna Slugify, která převádí vstupní znaky do UTF-8 kódování, což vedlo k problémům, kdybychom chtěli mít v attributech IRI požadovaného termínu z určité ontologie (např. IRI pro sílu levé ruky – [http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FBIN%2FBIN\\_00000054#classPermalinkModal](http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FBIN%2FBIN_00000054#classPermalinkModal)). Fyzicky totiž v tomto případě docházelo k nahrazení lomítek a dalších zvláštních znaků za pomlčky, což samo o sobě poškodilo vlastnosti IRI a stal se tak odkaz neplatným.

Další z problémů základního řešení spočíval v používání českých znaků v názvů atributů a nebo v označení již naměřených hodnot. Obzvláště pak byl problém s písmeny "ž" a "ř", které se z neurčitého důvodu pro kódování UTF-8 nepodařilo původnímu programu úspěšně přeložit. Samotné kódování UTF-8 nepodporovalo české znaky vůbec, kódování Latin2 ve většině případů fungovalo až na písmena ž a ř, která se špatně zkonvertovala. Funkčním řešením pro češtinu bylo v tomto případě kódování Windows-1250. Kódování vstupního souboru je tak nejspíše závislé na nástroji, ve kterém se

---

<sup>4</sup>Tato vlastnost podporuje psaní nejen v angličtině, ale i v češtině, ruštině a případně i v dalších jazycích využívající zvláštní znakovou sadou.

<sup>5</sup>Možnost exportovat až do osmi typů serializací RDF, RDF/XML, N3, N-Triples, Turtle, Pretty-XML, TriG, TriX, N-Quads.

vstupní CSV soubor vytvořil a nebo byl upravován. V tomto konkrétním případě byly zásadní změny v testovacím souboru dodělávány uvnitř tabulového editoru Microsoft Excel.

Tento nástroj využívá následující verze Python knihoven:

- **Pandas** – 0.17.0
- **RDFLib** – 4.2.1
- **Python-Slugify** – 1.2.1

### 4.1.3 Úprava CSV2RDF modulu

CSV2RDF modul byl mírně upraven, aby podporoval určité nové vlastnosti. Bohužel s těmito úpravami došlo i k několika nárokům na tvar či úpravu vstupního CSV souboru do validního formátu. V tomto případě názvy atributů **NESMÍ** obsahovat mezery. Pokud je nutnost použít víceslovný textový název atributu, tak se doporučuje použít podtržítkovou notaci (např. `sil_a_leve_ruky`). Také není vhodné, aby se název atributu skládal pouze z mezer a nebo jiných bílých znaků. Také je vhodné, když prázdné hodnoty naměřených dat budou obsahovat stejný řetězec označující NaN hodnoty.

První změnou byla úprava tvorby anotačních vlastností. Po této úpravě je možné vkládat do názvů jednotlivých vlastností IRI a nebo URL pro již nedefinované termíny uvnitř ontologií <sup>6</sup>.

Další změnou byla úprava některých zastaralých volání funkcí za aktuálně podporované varianty.

Také byly prozkoumány různé možnosti pro export do výstupních formátů RDF, podporované varianty zakódování vstupního CSV souboru pro různé znakové sady a zmapování některých již podporovaných přepínačů nástroje CSV2RDF.

### 4.1.4 Podporované přepínače

Upravená varianta CSV2RDF je navržena, aby byla spouštěna jako konzolová aplikace. Typické spuštění pythonovské konzolové aplikace<sup>7</sup> probíhá v

---

<sup>6</sup>Například místo textu, který se použije jako název atributu, `sil_a_leve_ruky`, lze uvést následující IRI `http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Furl.bioontology.org%2Fontology%2FBIN%2FBIN_00000054#classPermalinkModal`.

<sup>7</sup>Konzolová aplikace neobsahuje žádné grafické rozhraní a často bývá spouštěna právě ze systémové konzole (Windows) a nebo terminálu (Linux). Konzole se v prostředí Windows vyvolá stisknutím tlačítek [`windows button`] + R a napsáním příkazu `cmd` do tohoto nového okna, nebo spuštěním aplikace příkazového řádku (command line).

následujícím tvaru:

```
python [1] [2] [3]
[1] {navez zdrojoveho souboru}.py
[2] {navez vstupniho souboru}.csv
[3] {dalsi prepinače} {parametry prepinače}
```

Příklad takového spuštění může vypadat následovně:

```
python csv2rdf.py testovaci_data.csv --sep ";"
--encoding "windows-1250" --format "turtle"
```

Podporované přepínače nástroje CSV2RDF:

- **--outdata** – Název výstupního souboru, který bude obsahovat už samotná zkonvertovaná data, včetně přípony. Přípona se zadává nezávisle na serializaci RDF (může vzniknout turtle soubor s příponou .n3). Například "output\_data.ttl".
- **--outschema** – Název výstupního souboru, který bude obsahovat definice anotačních vlastností (první řádek ve vstupním CSV souboru) včetně přípony. V tomto případě by měl být stejný jako u přepínače **--outdata**. Například "output\_schema.ttl".
- **--tclass** – Definice třídy, která bude obsahovat anotační vlastnosti. V tomto případě se jedná o IRI na ontologii, ve které jsou všechny používané termíny nadefinovány (např. <http://purl.bioontology.org/ontology/BIN>).
- **--tnamespace** – Jmenný prostor pro zkonvertované naměřené hodnoty. Zde se uvádí textový řetězec, který bude identifikovat daný řádek z CSV souboru (např. text `user` bude poté číslován jako `user_1`, `user_2`, ...).
- **--schemanamespace** – Jmenný prostor, ve kterém jsou nadefinované URI anotačních vlastností. Takže se může například jednat o následující URI <http://purl.bioontology.org/ontology/BIN> a názvy atributů (anotačních vlastností) ze vstupního souboru budou dodávány na konec onoho URL. Takže by se pak mohlo jednat o [http://purl.bioontology.org/ontology/BIN/left\\_arm\\_strength](http://purl.bioontology.org/ontology/BIN/left_arm_strength), či <http://purl.bioontology.org/ontology/BIN/id>.
- **--format** – Výstupní formát RDF. V současné chvíli jsou podporovány následující formáty RDF/XML ("xml"), Notation3 ("n3"), Turtle ("turtle"), Pretty XML ("pretty-xml"), N-Triples ("nt") a TriG ("trig").

- **--quotechar** – Tento přepínač určuje znak, který se používá ve vstupním CSV jako komentář. Takže všechny znaky, které budou mezi začátkem a koncem tohoto znaku budou ignorovány.
- **--sep** – Tento přepínač specifikuje znak, který se používá ve vstupním CSV jako oddělovač. Běžně se jedná o středník (";") a nebo čárku (",").
- **--encoding** – Zde se určuje jaký typ kódování používá vstupní CSV soubor. Specifikace kódování může například umožnit používání českých znaků v IRI a nebo v názvech atributů.
- **--na\_values** – Textové řetězce, které symbolizují hodnotu Not a Number. Může se jednat například o řetězce NaN nebo NA.

#### 4.1.5 Příklad přeložení vstupního CSV souboru

Pro následující názorný příklad konverze mějme následující vstupní CSV soubor, který je uveden na obrázku 4.1.

Na následujícím obrázku 4.2 lze vidět použitou kombinaci textových atributů a upravených IRI stejného vstupního CSV souboru, ale otevřenou uvnitř tabulkového editoru. Modře zvýrazněné hodnoty jsou běžné textové atributy, které neobsahují žádný odkaz do ontologií. Žlutá barva zvýrazňuje atributy, které používají IRI do ontologií. Zelená pole poukazují na NaN hodnoty. NaN hodnoty v tomto případě znamenají, že daný řádek neobsahuje danou atributovou hodnotu pro korespondující sloupec.

Pro následující konverzi z CSV do RDF byl použit tento spouštěcí příkaz:

```
python csv2rdf.py vacekvi_data.csv
--sep ";" --encoding "windows-1250" --format "turtle"
--tclass "http://purl.bioontology.org/ontology/BIN"
```

Výsledkem tohoto spuštění je následující soubor v RDF serializaci Turtle. Výstupní soubor lze vidět na následujícím obrázku 4.3.

Podobný výsledek byl dosažen i na obrázku 4.4, kde byl pozměněn jeden přepínač ze spouštěcího příkazu. Jednalo se o výměnu přepínače `--format "turtle"` za `--format "xml"` a tak je výstup v RDF serializaci RDF/XML.

## 4.2 Populární RDF úložiště

Existuje velké množství RDF úložišť různého typu, které není možné během relativně rozumné doby všechna ověřit. Tak byl vyhledán fundovaný názor, či

```

vacekvi_data.csv – Poznámkový blok
Soubor Úpravy Formát Zobrazení Nápověda
id;group;sex;age;bioportal.bioontology.org/ontologies/BIN/?
p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FBIN
%2FBIN_00000080#classPermalinkModal;bioportal.bioontology.org/ontologie
s/BIN/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org
%2Fontology%2FBIN
%2FBIN_00000081#classPermalinkModal;bioportal.bioontology.org/ontologie
s/BIN/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org
%2Fontology%2FBIN
%2FBIN_00000082#classPermalinkModal;bioportal.bioontology.org/ontologie
s/BIN/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org
%2Fontology%2FBIN
%2FBIN_00000084#classPermalinkModal;bioportal.bioontology.org/ontologie
s/BIN/?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fxml%2Fowl
%2FEVS%2FThesaurus.owl
%23C53322#classPermalinkModal;bioportal.bioontology.org/ontologies/BIN/
?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology
%2FBIN
%2FBIN_00000085#classPermalinkModal;bioportal.bioontology.org/ontologie
s/BIN/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org
%2Fontology%2FBIN%2FBIN_00000086#classPermalinkModal
0;0;0;0;0;0;0;0;NaN;0;0
1;1;1;13;493;0;0;636;NaN;505;999
2;1;1;13;610;0;0;627;NaN;539;735
3;1;1;12;489;0;0;695;NaN;532;1327
4;1;0;17;464;0;0;441;NaN;374;637
5;1;1;17;423;0;0;576;392;454;2000
6;1;0;16;NaN;NaN;NaN;NaN;NaN;NaN;NaN
7;1;0;19;NaN;NaN;NaN;NaN;NaN;NaN;NaN
8;1;1;19;NaN;NaN;NaN;NaN;NaN;NaN;NaN
9;1;1;13;NaN;NaN;NaN;NaN;NaN;NaN;NaN
10;1;0;15;NaN;NaN;NaN;NaN;NaN;NaN;NaN
11;1;0;10;NaN;NaN;NaN;NaN;NaN;NaN;NaN
12;1;0;16;NaN;NaN;NaN;NaN;NaN;NaN;NaN
13;1;1;24;NaN;NaN;NaN;572;371;414;2000
14;1;1;29;NaN;NaN;NaN;NaN;NaN;NaN;NaN
15;1;0;30;NaN;NaN;NaN;NaN;NaN;NaN;NaN

```

**Obrázek 4.1:** Vstupního CSV soubor obsahoval následující informace. Základní nastavení skriptu doplňuje před všechny atributy (první řádek CSV souboru) textový řetězec "http://". Tato funkci samozřejmě lze upravit buď ve vstupním souboru a nebo lze tento efekt potlačit pomocí přepínače --schemanamespace. Proto v tomto příkladě neobsahují IRI převzaté z ontologií tento textový řetězec.

	A	B	C	D	E	F	G	H	I	J	K
1	id	group	sex	age	bioportal.	bioportal.	bioportal.	bioportal.	bioportal.	bioportal.	bioportal.
2	0	0	0	0	0	0	0	0	NaN	0	0
3	1	1	1	13	493	0	0	636	NaN	505	999
4	2	1	1	13	610	0	0	627	NaN	539	735
5	3	1	1	12	489	0	0	695	NaN	532	1327
6	4	1	0	17	464	0	0	441	NaN	374	637
7	5	1	1	17	423	0	0	576	392	454	2000
8	6	1	0	16	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	7	1	0	19	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	8	1	1	19	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	9	1	1	13	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	10	1	0	15	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	11	1	0	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	12	1	0	16	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	13	1	1	24	NaN	NaN	NaN	572	371	414	2000

**Obrázek 4.2:** Na tomto obrázku je uveden obsah stejného vstupního CSV souboru jako v předchozím případě pouze v prostředí tabulkového editoru. Modře zvýrazněné hodnoty obsahují běžné textové atributy. Žluté zvýraznění obsahuje atributy, které jsou uvedeny formou IRI (platný odkaz do ontologií) a zelená pole obsahují textový řetězec korespondující s prázdnou hodnotou. V tomto případě se jedná o textový řetězec NaN.

```

0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
11
1 @prefix ns1: <http://> .
2 @prefix ns2: <http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpubl.bioontology.org%2Fontology%2FBIN%2FBIN_00000086#> .
3 @prefix ns3: <http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fowl%2Fowl%2Fthesaurus.owl%2F353322#> .
4 @prefix ns4: <http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpubl.bioontology.org%2Fontology%2FBIN%2FBIN_00000080#> .
5 @prefix ns5: <http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpubl.bioontology.org%2Fontology%2FBIN%2FBIN_00000085#> .
6 @prefix ns6: <http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpubl.bioontology.org%2Fontology%2FBIN%2FBIN_00000084#> .
7 @prefix ns7: <http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpubl.bioontology.org%2Fontology%2FBIN%2FBIN_00000081#> .
8 @prefix ns8: <http://bioportal.bioontology.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpubl.bioontology.org%2Fontology%2FBIN%2FBIN_00000082#> .
9 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
10 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
11 @prefix xml: <http://www.w3.org/XML/1998/namespace> .
12 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
13
14 ns1:r_0 a <http://publ.bioontology.org/ontology/BIN> .
15
16 ns1:r_1 a <http://publ.bioontology.org/ontology/BIN> ;
17 ns1:age 1.3e+01 ;
18 ns4:classPermalinkNodal 4.93e+02 ;
19 ns6:classPermalinkNodal 6.36e+02 ;
20 ns5:classPermalinkNodal 5.05e+02 ;
21 ns2:classPermalinkNodal 9.99e+02 ;
22 ns1:group "1" ;
23 ns1:id "1" ;
24 ns1:sex 1e+00 .
25
26 ns1:r_10 a <http://publ.bioontology.org/ontology/BIN> ;
27 ns1:age 1.5e+01 ;
28 ns1:group "1" ;
29 ns1:id "10" .
30
31 ns1:r_100 a <http://publ.bioontology.org/ontology/BIN> ;
32 ns1:group "2" ;
33 ns1:id "100" .
34
35 ns1:r_101 a <http://publ.bioontology.org/ontology/BIN> ;
36 ns1:group "2" ;
37 ns1:id "101" .

```

**Obrázek 4.3:** Zde je uveden výstup z konverzního programu CSV2RDF, kde výsledný RDF soubor je zapsán v notaci Turtle. Tento výsledek byl dosažen pomocí dříve uvedeného spouštěcího příkazu.

```

0          10          20          30          40          50          60          70          80          90          100         110         120         130         140         150
1 <?xml version="1.0" encoding="UTF-8"?>
2 <rdf:RDF
3   xmlns:s1="http://
4   xmlns:s2="http://bioportal.bioontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Fpurl.bioontology.org%2Fontology%2FBIIN%2FBIIN_00000081#"
5   xmlns:s3="http://bioportal.bioontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Fpurl.bioontology.org%2Fontology%2FBIIN%2FBIIN_00000086#"
6   xmlns:s4="http://bioportal.bioontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Fpurl.bioontology.org%2Fontology%2FBIIN%2FBIIN_00000082#"
7   xmlns:s5="http://bioportal.bioontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Fpurl.bioontology.org%2Fontology%2FBIIN%2FBIIN_00000084#"
8   xmlns:s6="http://bioportal.bioontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Fpurl.bioontology.org%2Fontology%2FBIIN%2FBIIN_00000087#"
9   xmlns:s7="http://bioportal.bioontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Fpurl.bioontology.org%2Fontology%2FBIIN%2FBIIN_00000080#"
10  xmlns:s8="http://bioportal.bioontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Fpurl.bioontology.org%2Fontology%2FBIIN%2FBIIN_00000085#"
11  xmlns:rdfs="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
12 >
13   <rdf:Description rdf:about="http://r_26">
14     <ns1:sex rdf:datatype="http://www.w3.org/2001/XMLSchema#double">1.0</ns1:sex>
15     <ns1:id>26</ns1:id>
16     <ns7:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">601.0</ns7:classPermalinkNodal>
17     <ns4:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">80.0</ns4:classPermalinkNodal>
18     <ns3:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">799.0</ns3:classPermalinkNodal>
19     <rdfs:type rdf:resource="http://purl.bioontology.org/ontology/BIIN"/>
20     <ns8:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">478.0</ns8:classPermalinkNodal>
21     <ns5:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">626.0</ns5:classPermalinkNodal>
22     <ns1:age rdf:datatype="http://www.w3.org/2001/XMLSchema#double">32.0</ns1:age>
23     <ns1:group>1</ns1:group>
24   </rdf:Description>
25   <rdf:Description rdf:about="http://r_134">
26     <ns1:id>134</ns1:id>
27     <ns4:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">237.0</ns4:classPermalinkNodal>
28     <ns2:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">11.0</ns2:classPermalinkNodal>
29     <ns3:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">1116.0</ns3:classPermalinkNodal>
30     <ns1:sex rdf:datatype="http://www.w3.org/2001/XMLSchema#double">1.0</ns1:sex>
31     <ns5:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">867.0</ns5:classPermalinkNodal>
32     <ns7:classPermalinkNodal rdf:datatype="http://www.w3.org/2001/XMLSchema#double">598.0</ns7:classPermalinkNodal>
33     <ns1:group>2</ns1:group>
34     <ns1:age rdf:datatype="http://www.w3.org/2001/XMLSchema#double">70.0</ns1:age>

```

**Obrázek 4.4:** Zde je uveden výstup z konverzního programu CSV2RDF, kde výsledný RDF soubor je zapsán v notaci RDF/XML. Byl použit podobný spouštěcí příkaz jako v předchozím případě na obrázku 4.3, pouze byl vyměněn přepínač `--format "turtle"` za `--format "xml"`.

průzkum dostupných řešení, který poměruje, testuje a porovnává populárně používaná úložiště v mnoha kategoriích. Tento publikovaný článek [1] byl vypracován na velmi dobré úrovni, a zároveň byl tento průzkum aktuální, dostatečně podrobný a působil věrohodným dojmem.

Autor této práce otestoval celkem osm populárních RDF úložišť. Jednalo se o Blazegraph, Stardog, Virtuoso, GraphDB, AnzoGraph, AllegroGraph, MarkLogic a Apache Rya. Tato jmenovaná úložiště byla následně testována v několika kategoriích. Jednalo se o test rychlost načítání dat (ve formátech RDF/XML a N-Triples), rychlost zodpovídání dotazů (v osmi různě náročných SPARQL dotazech<sup>8</sup>) a subjektivním hodnocení. Subjektivní hodnocení obsahovalo následující kategorie:

- Jednoduchost nahrávání dat
- Jednoduchost dotazování se na data
- Jednoduchost prozkoumávání dat
- Jednoduchost monitorování úložiště (monitoring systému)
- Jednoduchost správy dat

<sup>8</sup>Autor všechny dotazy a vzorová data veřejně publikoval, odkazy na stáhnutí jsou uvedeny v textu článku.



### 4.2.1 Prvotní výběr zájmových úložišť - pomocí článku

Po prvotním testu, který se zabýval rychlostí načítání dat, byla úložiště *MarkLogic* a *Apache Rya* z následujících dvou kategorií testů (test dotazů a subjektivní hodnocení) vyřazena, protože těmto úložištím trvalo příliš dlouho načíst objemnější data.

V následujícím testu zodpovídání SPARQL dotazů byla různá úložiště úspěšná v jiných situacích. Ve většině dotazů však dominovala (byla jedna z nejrychlejších) dle mého názoru *Virtuoso*, *GraphDB* a *AnzoGraph*.

V následujících subjektivních testech se jevila jako uživatelsky přívětivější následující úložiště: *Stardog*, *GraphDB*. *Virtuoso* bylo v této kategorii, označeno autorem průzkumu, jako nejméně uživatelsky přívětivé.

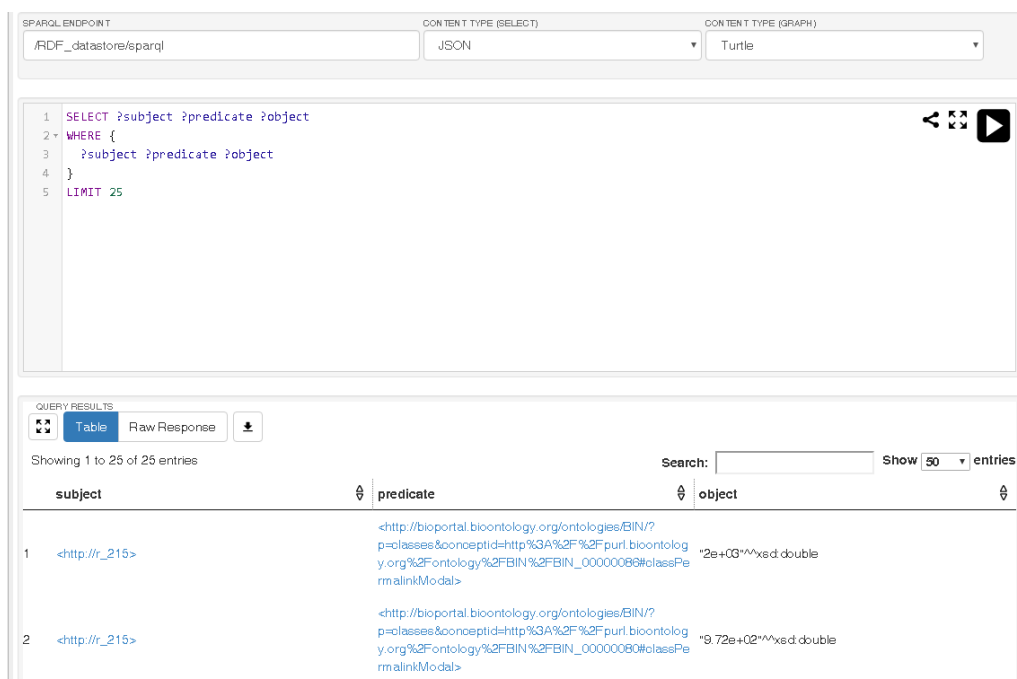
Na základě těchto výsledků a doporučení vedoucího práce bylo zvoleno pro další zkoumání úložiště *Virtuoso* a dle mého osobního názoru bylo zvoleno úložiště *GraphDB*. Třetím testovaným systémem bylo úložiště *Apache Jena Fuseki* zvolené pro svoji jednoduchost.

#### Apache Jena Fuseki

Jena Fuseki server je databázová aplikace pro ukládání, práci a dotazování nad RDF daty. Výhodou tohoto serveru je jednoduchost jeho zprovoznění a spuštění, protože samotná databázová aplikace vyžaduje pouze zprovozněnou Javu (multiplatformní programovací jazyk). Díky této vlastnosti je možné spustit Jena server buď jako službu operačního systému, jako Java aplikaci a nebo jako samostatný server. Jedná se o čistě Java aplikaci, která ve Fuseki variantě poskytuje HTTP rozhraní pro komunikaci se SPARQL serverem.

Během vývoje konverzní aplikace byl tento server použit pro ověření správné transformace dat do některého z vybraných RDF formátů. Jak vypadalo SPARQL dotazovací rozhraní uvnitř Jena lze vidět na následujícím obrázku 4.5.

Fuseki varianta Jeny obsahuje opravdu základní funkčnosti, jako jsou například zakládání nového datasetu, nahrávání dat do datasetu a uložení výsledků SPARQL dotazů do souboru CSV. Nepodporuje žádné možnosti spravování nebo přístupových účtů. Existuje varianta Jeny, která právě tyto funkčnosti přidává (Jena TBD). Jena je vydávána pod open source licenci Apache License, Version 2.0 (dostupná na <http://www.apache.org/licenses/LICENSE-2.0>).



**Obrázek 4.5:** Příklad SPARQL dotazu v prostředí Jena Fuseki Server, kde byla vypsána uložená testovací data (původně načtena ve formátu Turtle = celkem 3355 trojic).

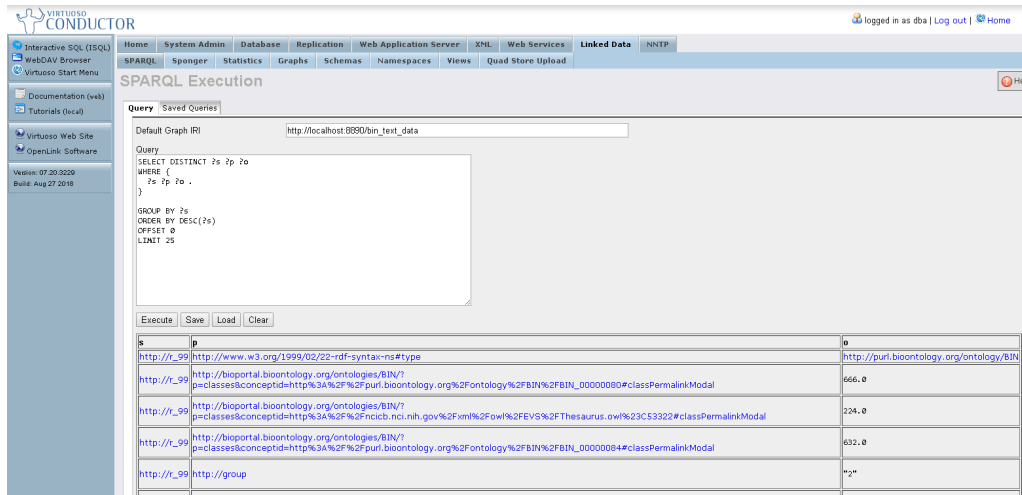
## OpenLink Virtuoso

V tomto případě byla testována free varianta Virtuosa (Virtuoso Open Source Edition). Jde o databázový systém napsaný v jazyce C se všemi výhodami i nevýhodami tohoto jazyka. Tento server je sám o sobě poměrně robustní a spolehlivý bohužel však za cenu uživatelské přívětivosti. Ve výsledcích SPARQL dotazovacích testů provedených v tomto článku [1] si Virtuoso vedl velmi dobře.

Tento server je zejména vhodný pro pokročilé uživatele, kteří rozsah jeho vlastností, podporovaných jazyků a funkcí využijí. Virtuoso zvládá velmi výkoně načítání i dotazování se na data. Správa samotného serveru také poskytuje poměrně vhodné množství funkcí jako je například monitoring zdrojů, spuštěných operací, správa uživatelských účtů až po specifikaci práv (read, write) na jednotlivé datasety.

Bohužel díky robustnosti serveru Virtuoso je potřeba pokročilejší znalost samotného serveru a případně i problematiky týkající se samotné správy serveru. Například dotazování se na určitý RDF datastore probíhá pomocí běžného SPARQL Query interface, bez jakékoliv interaktivní kontroly, takže případně sestavený složitý dotaz je nutné uložit uvnitř serveru, jinak bude

potřeba každý dotaz znovu definovat od začátku (včetně prefixů). Jak vypadají načtená vstupní data v prostředí Virtuoso Conductor (webová aplikace pro přístup k serveru) lze vidět na následujícím obrázku 4.6.



**Obrázek 4.6:** Příklad SPARQL dotazu v prostředí Virtuoso Conductor, kde byla vypsána uložená testovací data (původně načtena ve formátu Turtle = celkem 3355 trojic).

Virtuoso je vydáván jak v open source omezené variantě (aktuálně verze 7.2.5), tak při komerční licenci (8.2). Virtuoso je tak vhodný pro nasazení živé aplikace (díky své robustnosti), než pro vývoj nebo testování nějaké RDF aplikace.

## Ontotext GraphDB

GraphDB je RDF databáze napsaná v Javě, která dokáže pracovat s poměrně velkou škálou RDF formátů. Tento server je vydáván jak ve variantě s omezenou free edicí, tak i placenou licenci. GraphDB je poměrně jednoduchý na zprovoznění a obsahuje větší část nástrojů pro správu serveru, uživatelských účtů, práv a monitoringu zdrojů.

Také obsahuje poměrně rozsáhlé a uživatelsky přívětivé grafické rozhraní, které se poměrně snadno ovládá. Bohužel však rychlost dotazování a nahrávání dat je ve většině případů horší než v případě Virtuosa (viz zdroj [1]). Jak vypadá dotazování v prostředí GraphDB lze vidět na následujícím obrázku 4.7.

Pokud bychom porovnali tento server s Virtuosem, vyměňuje se zde výkonnost úložiště Virtuoso za jednoduchost nastavení a správy serveru.

The screenshot shows the GraphDB interface. On the left is a navigation menu with options: Import, Explore, SPARQL (highlighted), Monitor, Setup, and Help. The main area contains a SPARQL query editor with the following query:

```

2 select ?s ?p ?o
3 where {
4   ?s ?p ?o .
5 }
6 ORDER BY DESC(?s)
7 Limit 25

```

Below the query editor are tabs for 'Table', 'Raw Response', 'Pivot Table', and 'Google Chart'. A 'Download as' button is on the right. The results section shows 'Showing results from 1 to 25 of 25. Query took 0.1s, minutes ago.' and a table with the following data:

	s	p	o
1	ns1r_99 ns1bioportal.biontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Ffurl.biontology.org%2Fontology%2FBIIN%2FBIIN_00000086#classPermalinkModal		"1.041e+03""xsd:double
2	ns1r_99 ns1stage		"1.8e+01""xsd:double
3	ns1r_99 ns1bioportal.biontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Fncicb.nci.nih.gov%2Fxml%2Fowl%2FEVS%2FThesaurus.owl%2Z3C5322#classPermalinkMo		"2.24e+02""xsd:double
4	ns1r_99 ns1bioportal.biontology.org/ontologies/BIIN/?p=classes&conceptid=http%3A%2Ffurl.biontology.org%2Fontology%2FBIIN%2FBIIN_00000085#classPermalinkModal		"4.82e+02""xsd:double

**Obrázek 4.7:** Příklad SPARQL dotazu v prostředí GraphDB, kde byla vypsána uložená testovací data (původně načtena ve formátu Turtle = celkem 3355 trojic).

### 4.3 Výběr tří RDF úložišť pro podrobnější testování

Všechna tato tři úložiště byla otestována a jsou funkční. Zvolení vhodného úložiště však závisí na způsobu jeho použití a také na znalostech budoucího správce tohoto serveru.

Z těchto prozkoumaných řešení je možno konstatovat, že pokud se jedná o vývoj a nebo provozování jednoduchého RDF úložiště, bylo by vhodné použít Jena Fuseki respektive GraphDB.

V případě, že by se jednalo o provozování náročnějšího RDF úložiště, kde by bylo potřeba využít plný výkonnostní potenciál vybraného úložiště, tak Virtuoso je tou vhodnější volbou z těchto uvedených.

# 5 Testování vhodných RDF úložišť

Vzhledem k tomu, že byly v předchozí kapitole vybrány celkem tři zajímaví kandidáti na funkční RDF úložiště, bylo nutné sestavit sadu testů, které by posloužily k jejich podrobnějšímu otestování. V tomto případě byly sledovány následující metriky: rychlost načítání dvou různých RDF formátů (Turtle a N-Triples), rychlost zodpovídání celkem sedmi SPARQL dotazů (s postupně zvyšující se složitostí) a jakým subjektivním dojem působila na autora této práce během testování.

## 5.1 Parametry testovacího stroje

Stolní PC, na kterém probíhalo testování mělo následující parametry:

- **Operační systém** – Windows 8.1 Pro 64-bit (6.3, Build 9600) (9600.winblue\_ltsb\_escrow.190313-1806)
- **Procesor** – Intel(R) Xeon(R) CPU E5-1650 v2 @ 3.50GHz (12 CPUs), 3.5GHz
- **Grafická karta** – NVIDIA GeForce GTX 780 Ti
- **Paměť** – 16 GB

Na tomto stroji bude následně vybrané úložiště také zprovozněno.

## 5.2 Vybrané RDF formáty

V tomto případě byl u všech vybraných úložišť sledován čas, který byl potřeba k načtení dvou velmi různých RDF formátů. Turtle (viz kapitola 2.3.2) je z těchto dvou vybraných formátů ten, který je lépe čitelný pro člověka a formát N-Triple<sup>1</sup>.

Vzhledem k tomu, že samotných převedených dat ze systému BiN zatím není dostatečně velké množství (zhruba několik tisíců trojic), aby je bylo

---

<sup>1</sup>N-Triple je RDF formát, který se skládá pouze z trojic a nepoužívá jakéhokoliv zkracovací (např. aliasy jako PREFIX u Turtle formátu), tak je tento formát poměrně objemný a tím pádem i náročnější na parsování.

možné využít samostatně k podrobnějšímu otestování RDF úložišť, byla využita ještě další veřejně dostupná RDF data, která k tomuto testování posloužila. Jednalo se o data obsahující 120 let historie olympijských her<sup>2</sup>. Přehled vybraných testovacích souborů i s rozdělením podle vybraných formátů a velikostí lze vidět v následujících tabulkách 5.1 a 5.2.

<b>Turtle (.ttl)</b>	<b>Označení</b>	<b>Počet trojic</b>	<b>Velikost souboru</b>
bin_dvt_2017.ttl	small	7 326	298 kB
olympics.ttl	medium	1 781 625	61 773 kB
120-years-olympics.ttl	large	7 809 410	624 791 kB

**Tabulka 5.1:** Zde jsou uvedeny jednotlivé velikosti vstupních souborů formátu Turtle, kde první a nejmenší soubor obsahuje převedená data ze systému BiN. Jedná se data změřená ze Dnů vědy a Techniky 2017 v Plzni. Zbývající data byla převzata z veřejných RDF dat obsahující až 120 let historie olympijských her, která byla případně převedena do Turtle formátu.

Největší soubor má velikost v tomto případě zhruba 0,6 GB. Turtle ale umožňuje jednoduchou a dobře lidsky čitelnou serializaci RDF. Nyní se však podívejme na stejná data, ale pouze v jiném formátu N-Triples. Velikosti tohoto formátu jsou uvedeny v následující tabulce 5.2.

<b>N-Triples (.nt)</b>	<b>Označení</b>	<b>Počet trojic</b>	<b>Velikost souboru</b>
bin_dvt_2017.nt	small	7 326	1 683 kB
olympics.nt	medium	1 781 625	329 782 kB
120-years-olympics.nt	large	7 809 410	1 182 538 kB

**Tabulka 5.2:** Zde je možné vidět, že data, která obsahují stejnou sémantickou hodnotu a pouze jsou zapsána v jiné RDF serializaci (oproti tabulce 5.1) mají mnohem větší objem. Hlavní rozdíl formátu N-Triples oproti formátu Turtle je v tomto případě velikost vstupního souboru, která je zhruba 2x až 4x větší oproti Turtle formátu.

### 5.3 Použité SPARQL dotazy

SPARQL dotazy byly převzaty [1] ze stejné stránky jako byla dodatečná RDF data<sup>3</sup> a byla částečně upravena podle potřeby (např. pro Virtuoso).

<sup>2</sup>Tato data jsou veřejně dostupná na této adrese <https://github.com/wallscope/olympics-rdf>.

<sup>3</sup>Tato data jsou veřejně dostupná na této adrese <https://github.com/wallscope/olympics-rdf> včetně dotazů.

Tyto dotazy byly vytvořeny pro největší datový set (soubor `large-120-years-olympics`) a jsou v nich obsaženy různě složité SPARQL konstrukce (agregační funkce, regulární výrazy, dotazy na externí zdroje, ...). V následující části bude stručně popsáno všech sedm použitých SPARQL dotazů.

### 5.3.1 SPARQL dotaz 1

#### Slovní popis dotazu

Výpis jmen každého atleta, který získal zlatou medaili.

#### Komentář k funkcionalitě

Každé z uvedených jmen bude obsaženo pouze jednou (DISTINCT).

#### Výsledný počet řádků odpovědi

10 414

#### SPARQL kód

```
PREFIX walls: <http://wallscope.co.uk/ontology/olympics/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
SELECT DISTINCT ?name
```

```
WHERE {
  ?instance walls:athlete ?athlete ;
  walls:medal <http://wallscope.co.uk/resource/olympics/medal/Gold> .
  ?athlete rdfs:label ?name .
}
```

### 5.3.2 SPARQL dotaz 2

#### Slovní popis dotazu

Výpis jmen každého atleta, který získal alespoň jednu medaili včetně celkového počtu medailí získaného daným atletem. Na závěr seřazených podle počtu získaných medailí.

## Komentář k funkcionalitě

Použitá agregační funkce `COUNT` pro spočítání celkového počtu medailí atleta. Také se používá seskupení výsledků (`GROUP BY`) podle jména včetně sestupného seřazení (`ORDER BY DESC`) podle počtu medailí.

## Výsledný počet řádků odpovědi

28 203

## SPARQL kód

```
PREFIX walls: <http://wallscope.co.uk/ontology/olympics/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?name (COUNT(?name) As ?noOfMedals)
WHERE {
    ?instance walls:athlete ?athlete ;
              walls:medal ?medal .
    ?athlete rdfs:label ?name .
}
GROUP BY ?name
ORDER BY DESC(?noOfMedals)
```

### 5.3.3 SPARQL dotaz 3

#### Slovní popis dotazu

Vypiš všechny státy a uveď u každého státu průměrnou výšku a hmotnost pro jejich atlety. Výsledek seřaď sestupně podle vypočtené průměrné výšky.

## Komentář k funkcionalitě

Použití agregační funkce průměru (`AVG`) pro spočítání průměrné výšky a hmotnosti atleta z daného státu. Seskupení (`GROUP BY`) výsledků podle kódu státu a sestupné seřazení (`ORDER BY DESC`) podle vypočítané průměrné výšky atleta.

## Výsledný počet řádků odpovědi

227



## SPARQL kód

```
PREFIX walls: <http://wallscope.co.uk/ontology/olympics/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?countryCode
        (AVG(?height) As ?avgHeight)
        (AVG(?weight) As ?avgWeight)
WHERE {
    ?noc dbo:ground ?team ;
        rdfs:label ?countryCode .

    ?athlete rdf:type foaf:Person ;
             dbo:team ?team ;
             dbo:height ?height ;
             dbo:weight ?weight .
}
GROUP BY ?countryCode
ORDER BY DESC(?avgHeight)
```

### 5.3.4 SPARQL dotaz 4

#### Slovní popis dotazu

Výpis všech let, kdy se konaly olympijské hry, včetně věku nejstaršího účastníka soutěže v Judu v daném roce.

#### Komentář k funkcionalitě

Použití agregační funkce z hledání maxima (MAX) pro určení nejstaršího účastníka soutěže v Judu. Omezení výpisu pouze na sport Judo a seskupení výsledků podle roku konání.

#### Výsledný počet řádků odpovědi

13

## SPARQL kód

```
PREFIX walls: <http://wallscope.co.uk/ontology/olympics/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dbp: <http://dbpedia.org/property/>
```

```
SELECT ?year (MAX(?age) As ?maxAge)
WHERE {
    ?instance walls:games ?games ;
              walls:event ?event ;
              walls:athlete ?athlete .

    ?event rdfs:subClassOf <http://wallscope.co.uk/resource/olympics/sport/Judo> .

    ?games dbp:year ?year .

    ?athlete foaf:age ?age .
}
```

**GROUP BY** ?year

### 5.3.5 SPARQL dotaz 5

#### Slovní popis dotazu

Vypiš jméno každého atleta jehož jméno obsahuje řetězec "louis" včetně názvu města kde soutěžil a také uveď o jakou olympijskou sezónu se jednalo.

#### Komentář k funkcionalitě

Tento výraz využívá regulární výraz (**REGEX**) pro vyhledávání jmen atletů obsahujících řetězec "louis". Každé jméno soutěžícího bude obsaženo pouze jednou (**DISTINCT**). Tento dotaz až na konci filtruje (**FILTER**) výsledky podle řetězce "louis", takže do té doby dotaz hledá všechny jména atletů, kteří splňují parametry.

#### Výsledný počet řádků odpovědi

1 016

## SPARQL kód

```
PREFIX walls: <http://wallscope.co.uk/ontology/olympics/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbp: <http://dbpedia.org/property/>
```

```
SELECT DISTINCT ?name ?cityName ?seasonName
WHERE {
    ?instance walls:games ?games ;
              walls:athlete ?athlete .

    ?games dbp:location ?city ;
           walls:season ?season .

    ?city rdfs:label ?cityName .

    ?season rdfs:label ?seasonName .

    ?athlete rdfs:label ?name .

    Filter (REGEX(lower(?name), "louis.*"))
}
```

### 5.3.6 SPARQL dotaz 6

#### Slovní popis dotazu

Nejdříve vypiš seznam měst, ve kterých soutěžili atleti ze státu Srbsko a Černé hora. Následně spočítej počet mužů a žen, kteří soutěžili v těchto městech a vrať jejich celkový počet.

#### Komentář k funkcionalitě

Z celé této sady dotazů jde o ten nejnáročnější. Tento dotaz využívá vnořený dotaz, jehož výsledky se následně dále používají. Každý název města je uchováván pouze jednou (**DISTINCT**). Pro výpočet celkového množství atletů rozdělených dle města a pohlaví se používá agregační funkce (**COUNT**). Vypočítané množství atletů je nakonec seskupeno (**GROUP BY**) dle pohlaví atletů.

## Výsledný počet řádků odpovědi

2

## SPARQL kód

```
PREFIX walls: <http://wallscope.co.uk/ontology/olympics/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dbp: <http://dbpedia.org/property/>
PREFIX noc: <http://wallscope.co.uk/resource/olympics/NOC/>
```

```
SELECT ?genderName (COUNT(?athlete) AS ?count)
WHERE {
    ?instance walls:games ?games ;
              walls:athlete ?athlete .

    ?games dbp:location ?city .

    ?athlete foaf:gender ?gender .

    ?gender rdfs:label ?genderName .

    {
        SELECT DISTINCT ?city
        WHERE {
            ?instance walls:games ?games ;
                      walls:athlete ?athlete .

            ?athlete dbo:team ?team .

            noc:SCG dbo:ground ?team .

            ?games dbp:location ?city .
        }
    }
}
```

**GROUP BY** ?genderName

### 5.3.7 SPARQL dotaz 7

#### Slovní popis dotazu

Vypiš název každého sportu, který má uloženy údaje o velikosti týmu uvnitř DBpedia.org, a uveď u každého sportu o jeho velikosti týmu.

#### Komentář k funkcionalitě

Na první pohled se může dotaz jevit jako velmi jednoduchý, ale v rámci vykonávání dotazu dochází k dotazování se na externí zdroj (v tomto případě DBpedia.org) a tím pádem hraje roli v měření tohoto dotazu i odezva sítě.

#### Výsledný počet řádků odpovědi

13

#### SPARQL kód

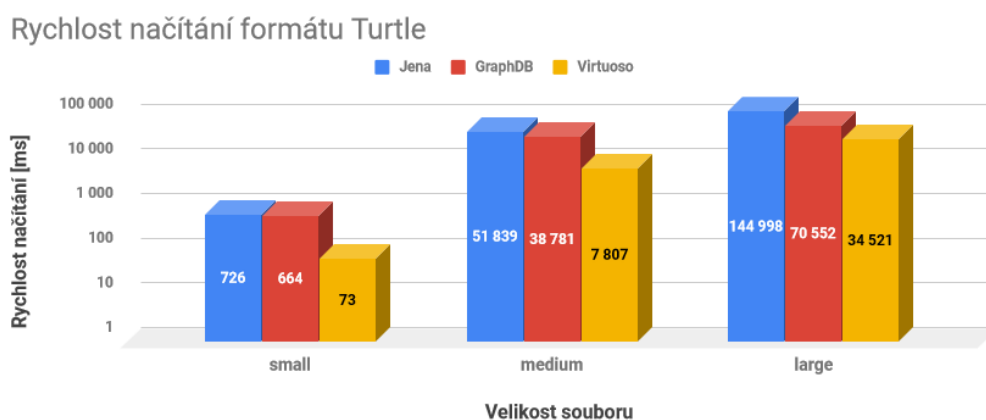
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>
```

```
SELECT ?sportName ?teamSize
WHERE {
    ?sport rdf:type    dbo:Sport ;
           rdfs:label ?sportName .

    SERVICE <http://dbpedia.org/sparql> {
        ?dbsport rdfs:label    ?sportName ;
                 dbo:teamSize ?teamSize .
    }
}
```

## 5.4 Testování rychlosti nahrávání vstupních RDF souborů

První ze série testů bylo testování rychlosti nahrávání různě velikých RDF souborů<sup>4</sup> ve dvou specifických formátech Turtle a N-Triples. V případě Jena a GraphDB bylo nahrávání prováděno přes HTTP klienta a v případě Virtuosa šlo o nahrávání dat pomocí konzole. Každý z těchto souborů v daném formátu byl nahráván celkem třikrát a výsledky byly zprůměrovány běžným aritmetickým průměrem. Výsledky lze vidět na obrázku 5.1 pro formát Turtle a 5.2 pro formát N-Triples.



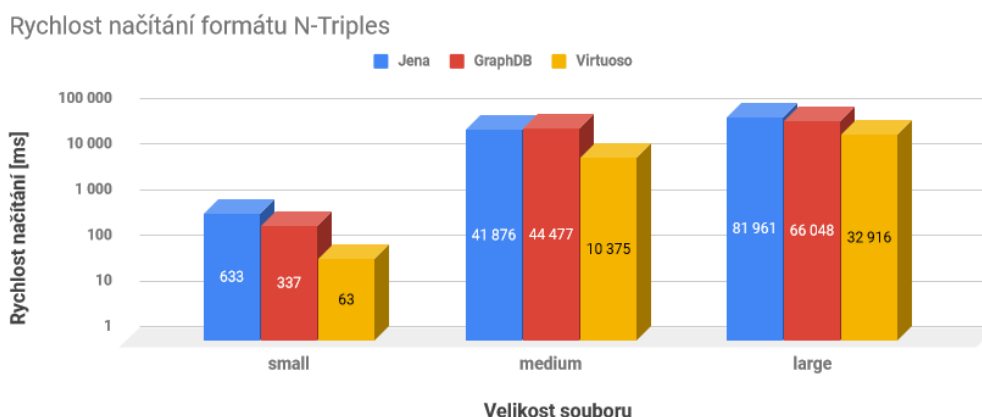
**Obrázek 5.1:** Výsledky testování všech vybraných úložišť, které dosáhly v průměru při načítání Turtle formátu RDF.

Úložiště Jena a GraphDB jsou původně napsány v programovacím jazyce Java a jejich plnění primárně probíhalo přes HTTP klienta, kterého obě úložiště podporují. Tento přístup by částečně mohl vysvětlovat délku plnění těchto úložišť. Naopak u úložiště Virtuoso bylo plnění spouštěno pomocí příkazové řádky<sup>5</sup>, kde samotné Virtuoso je implementováno v programovacím jazyce C, což by naopak mohlo vysvětlit výjimečnou rychlost jeho plnění.

V tomto kole nahrávání bylo vždy úložiště Virtuoso razantně rychlejší než ostatní úložiště, které i na největším datovém souboru dokázalo načíst cca 7 800 000 trojic za necelých 35 sekund. Naopak nejhorším úložištěm bylo v tomto případě Jena Fuseki, kde samotné načítání trvalo v průměru přes 2 minuty 20 vteřin, což je téměř 5x pomalejší než Virtuoso.

<sup>4</sup>Pro podrobnější informace viz tabulky 5.1 a 5.2.

<sup>5</sup>Virtuoso konzole se v prostředí Windows spouští z příkazové řádky spuštěním `iSQL.exe` uvnitř složky `Virtuoso\bin`.



**Obrázek 5.2:** Výsledky testování všech vybraných úložišť, které dosáhly v průměru při načítání N-Triples formátu RDF.

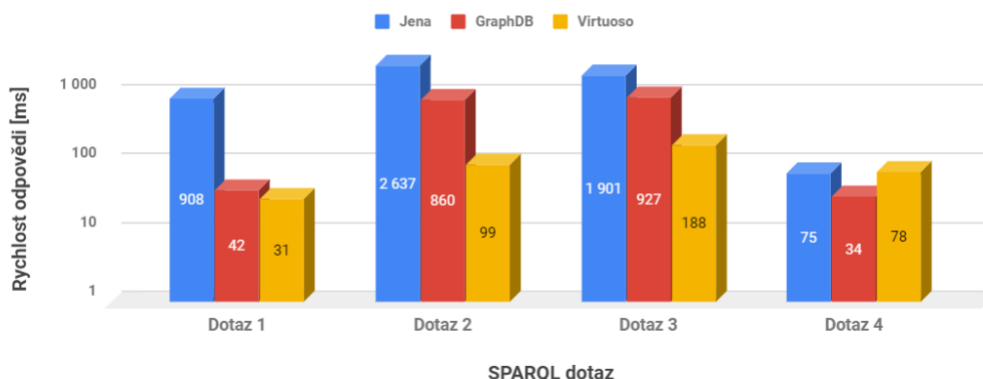
Prvotní očekávání bylo, že by nahrávání Turtle formátu mělo probíhat rychleji než u formátu N-Triples, protože jde celkově o méně dat, která se musejí procházet, ale nebylo tomu tak. Naopak ve většině případů bylo načítání N-Triples formátu rychlejší. Největší zlepšení nahrávacího času bylo u úložiště Jena Fuseki (téměř o minutu), když se jednalo o největší datový soubor, ale i přesto se toto úložiště pyšní stále nejhorší dobou nahrávání. Vybraná úložiště si dokonce zachovala stejné pořadí v rychlosti načítání, tak jako tomu bylo v případě formátu Turtle.

## 5.5 Testování rychlosti dotazování

Podobně jako v případě nahrávání RDF dat, byla rychlost dotazování měřena třikrát pro každý z dotazů a tak byly výsledné a zároveň poměřované hodnoty tohoto testu vypočítány pomocí aritmetického průměru. Výsledky pro první čtyři dotazy lze vidět na následujícím obrázku 5.3.

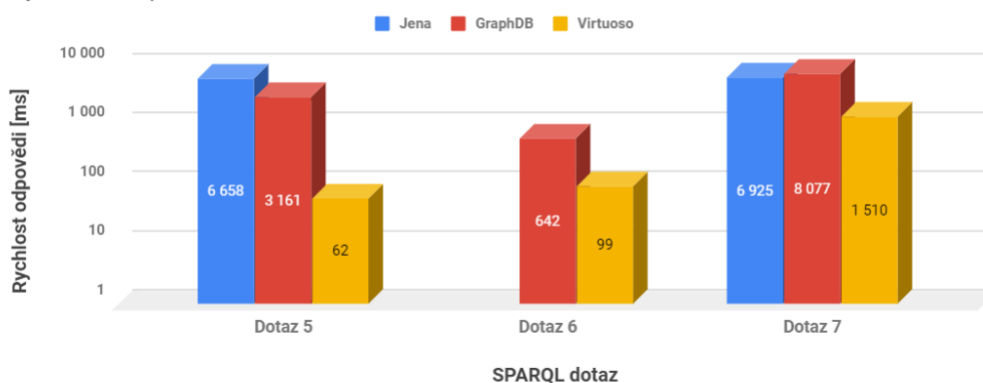
V těchto lehčích dotazech byl opět markantně nejrychlejší Virtuoso, kdy byl dotaz průměrně zodpovězen v rámci stovek milisekund, kde jiná úložiště se pohybovala spíše okolo jedné sekundy. Nejjednoduššími dotazy z této čtveřice byly dotazy 1 a 4, které využili pouze příkazy `DISTINCT` a `MAX`, většina úložišť dokázala odpovědět poměrně rychle (cca do jedné stovky milisekund). U dotazů 2 a 3 se již projevily unikátní způsoby řešení vyhodnocování agregačních funkcí a řazení u jednotlivých úložišť. Dokonce i pořadí úložišť dosažené během nahrávání RDF dat bylo zachováno i v dotazování, až na dotaz číslo 4, kde naopak nejrychlejším úložištěm bylo GraphDB. Na dalším obrázku 5.4 jsou vidět výsledky pro složitější SPARQL dotazy.

Rychlost zodpovídání SPARQL dotazů 1 až 4



**Obrázek 5.3:** Výsledky dosažené při testování rychlosti dotazování úložišť na SPARQL dotazech jedna až čtyři (viz kapitola Použité SPARQL dotazy 5.3), které využívali příkazy na jedinečnost výsledků (DISTINCT) a běžné agregační funkce (AVG, MAX, COUNT). Uvedené doby v milisekundách jsou výsledkem aritmetických průměrů ze tří měření pro každé z uvedených úložišť zaokrouhlenými na celé milisekundy.

Rychlost zodpovídání SPARQL dotazů 5 až 7



**Obrázek 5.4:** Rychlost odpovídání na složitější SPARQL dotazy. Tyto dotazy obsahovali jak vnořené dotazy, dotazování se na externí službu a využívání regulárního výrazu. I v tomto případě vítězí na celé čáře Virtuoso. Uvedené doby v milisekundách jsou výsledkem aritmetických průměrů ze tří měření pro každé z uvedených úložišť zaokrouhlenými na celé milisekundy.

Ve složitých dotazech je opět poznat výpočetní síla a optimalizace úložiště Virtuoso, které i zde dokázalo získat odpověď v rámci stovek milisekund. Naopak u dotazu číslo 6, který využívá vnořené dotazy, nebyla schopna Jena odpovědět vůbec. Nejdelší časy ze všech dotazů jsou však u



dotazu číslo 7, který se v rámci vyhodnocování využívá externí službu (připojení na RDF server DBpedia.org). Tento dotaz je velmi ovlivněn rychlostí připojení k internetu.

## 5.6 Výběr vhodného RDF úložiště

Ze specifických testů rychlosti dosáhl vždy Virtuoso nejlepších výsledků, ale za tuto rychlost úložiště platí v jiných částech aplikace. Autorovi této práce vzhledově více vyhovovalo úložiště GraphDB, jehož zprovoznění a ovládní bylo poměrně intuitivní a jednoduché, které ale bohužel nedosahovalo podobných výsledků jako Virtuoso. Samotné úložiště Virtuoso je rozhodně více vhodné po výpočetní stránce než GraphDB, ale zato se neovládá příliš snadno.

I přes tyto malé nedostatky je však hlavním účelem RDF úložiště skladovat objemná data a dokázat v nich rychle vyhledávat odpovědi, a proto je nejvhodnějším kandidátem pro nasazení do reálného provozu úložiště **Virtuoso**. Však GraphDB je doporučeno autorem této práce jako vhodné úložiště pro testovací a nebo kontrolní účely právě díky svému poměru jednoduchost/výkonnost.

# 6 Nasazení a naplnění vybraného RDF úložiště

Pro nasazení a další testování bylo vybráno úložiště OpenLink Virtuoso Open Source Edition, které bylo zprovozněno v rámci infrastruktury BiN, která se nalézá převážně uvnitř neuroinformatické laboratoře na katedře informatiky a výpočetní techniky.

Toto RDF úložiště bylo zprovozněno na stejném stroji (viz parametry testovacího stroje v kapitole 5.1), jako bylo prováděno předchozí testování. Úložiště bylo staženo z veřejně dostupných zdrojů z internetu z oficiálních stránek autorů tohoto softwaru. Tento software je publikován pod licenčním ujednáním GNU General Public License (GPL) Version 2, která umožňuje i komerční používání, ale pouze za předpokladu, že finální produkt bude také open source (v současné chvíli systém BiN je open source).

Pro stažení tohoto software není potřeba žádná registrace. Program se stáhne jako součást samo-instalačního balíčku (v prostředí Windows). Program byl nainstalován na lokálním pevném disku počítače. Program si pro svůj provoz dodatečně zprovoznil balíček Microsoft Visual C++ Redistributable. Pro instalaci tohoto programu byla potřeba administrátorská práva. Proto je potřeba uvažovat o tomto faktu, když by se toto úložiště zprovožovalo na veřejném univerzitním serveru.

## 6.1 Testování nasazeného řešení

Pro otestování zda úložiště podporuje načítání různých RDF formátů, tak byly vytvořeny celkem čtyři vstupní soubory, které se pouze lišily použitou RDF serializací. Jednalo se o stejná data ze systému BiN, která byla využita již jako reprezentativní vzorek dat v rámci prvního výběru<sup>1</sup> a testování úložišť.

- **RDF/XML**
- **Turtle**
- **N-Triples**
- **Notation3**

V případě Turtle a N-Triples formátu se jednalo o stejný zdrojový soubor, který byl použit při prvotním výběru úložišť. Zbývající formáty byly zvoleny

---

<sup>1</sup>Jednalo se v tomto případě o vstupní souboru small.

na základě podporovaných RDF formátů úložiště Virtuoso.

### 6.1.1 Test nahrávání různých RDF formátů

Prvním krokem bylo nahrávání různých RDF formátů do různě pojmenovaných repositářů, aby bylo možné následně pokládat kontrolní SPARQL dotazy, které by ověřovaly správnost konverze a načtení dat. Zdrojová data, která poslouží k tomuto otestování musejí být pro všechny serializace stejná, aby bylo možné ověřit případné chyby při konverzi mezi RDF serializacemi, či rozdílná data při SPARQL dotazování, a zároveň musí pocházet ze systému BiN. Byla proto vybrána data pořízená v rámci Dnů Vědy a Techniky 2017, protože se jednalo o jednu z nejrozsáhlejších kolekcí dat uvnitř systému BiN, pořízených v rámci dvou měřících dnů (celkem 16 hodin měření). V následující tabulce 6.1 je možné vidět jaké množství trojic bylo při různých serializacích načteno a jak dlouho toto nahrávání trvalo v úložišti Virtuoso.

Serializace	Počet trojic	Doba nahrávání	Velikost souboru
RDF/XML	7 326	62 ms	642 kB
Turtle	7 326	63 ms	298 kB
N-Triples	7 326	63 ms	1 683 kB
Notation3	7 326	62 ms	298 kB

**Tabulka 6.1:** Prezentuje, že každá z uvedených serializací popisovalo stejné množství trojic. Tento fakt ověřuje, že data byla převedena správným způsobem do různých serializací a zároveň byla i správně načtena do úložiště. Bohužel data, která sice obsahují desetitisíce trojic jsou stále ještě moc malá, než aby jejich doba načítání byla razantně rozdílná.

### 6.1.2 Test dotazování se na načtená data v různých serializacích

Způsob kontroly načtených testovacích dat, byl založen na základě poměrování počtu navrácených řádků u každého z uvedených SPARQL dotazů včetně dohledávání dosažených výsledků ze zdrojových CSV dat. Také byla u těchto dotazů sledována doba, která byla potřeba k jejich vykonávání. Testovací dotazy byly specificky vytvořeny pro datový set DvT2017. Jednalo se o celkem tři SPARQL dotazy, kde jejich slovní popis včetně zdrojového kódu lze nalézt v následujících podkapitolách.

## SPARQL testovací dotaz 1

### Slovní popis dotazu

Výpis všech informací (atributů), které byly vyplněny (neobsahovali NaN hodnoty) o účastníkovi pod označením r\_273.

### Komentář k funkcionalitě

Účelem tohoto dotazu, bylo vypsat informace, které lze snadno ověřit ve zdrojovém CSV souboru. Označení r\_273 popisuje v tomto případě jiného účastníka nežli toho, který má podobné označení uvnitř atributu id ve zdrojovém CSV souboru (v tomto případě id = 274).

### Výsledný počet řádků odpovědi

25

### SPARQL kód - upravený pro Virtuoso

```
PREFIX bin: <http://purl.bioontology.org/ontology/  
BIN/id/>
```

```
SELECT ?information ?value  
FROM <[id\_repository\_s\_daty]>  
WHERE {  
    bin:r_273 ?information ?value .  
}
```

## SPARQL testovací dotaz 2

### Slovní popis dotazu

Vypiš všechna ID účastníků a jejich nejhorších časů dolních končetin, kteří měli svůj nejhorší čas dolních končetin nad 1 200 ms (hraniční hodnota byla maximálně 2 000 ms). Výsledky seřaď sestupně podle nejhoršího času dolních končetin.

### Komentář k funkcionalitě

Tento dotaz využívá již konstrukce dotazovacího jazyka SPARQL. Jedná se zejména o klíčová slova FILTER, pro omezení reakční doby nad 1 200 ms a ORDER BY, pro seřazení výsledků

## Výsledný počet řádků odpovědi

160

## SPARQL kód - upravený pro Virtuoso

```
PREFIX bin: <http://purl.bioontology.org/ontology/
BIN/id/>
PREFIX response: <https://bioportal.bioontology.org/
ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpurl
.bioontology.org%2Fontology%2FBIN%2FBIN_00000086#>

SELECT ?participant_id ?slowest_response_time
FROM <[id\_repositare\_s\_daty]>
WHERE {
    ?participant_id response:classPermalinkModal
    ?slowest_response_time .
    FILTER(?slowest_response_time > 1200)
}
GROUP BY ?participant_id
ORDER BY DESC(?slowest_response_time)
```

## SPARQL testovací dotaz 3

### Slovní popis dotazu

Vypiš identifikátory všech účastníků, kteří měli průměrnou reakční dobu horních končetin pod 800 ms včetně a zároveň udělali při měření méně než dvě chyby. Výsledky seřaď vzestupně.

### Komentář k funkcionalitě

Tento dotaz používá složený FILTER. Zahrnutí účastníci musejí mít zároveň vyplněny jak reakční dobu horních končetin, tak počet minutí během testu, aby byly obsaženi ve výsledcích.

## Výsledný počet řádků odpovědi

22

## SPARQL kód - upravený pro Virtuoso

```

PREFIX bin: <http://purl.bioontology.org/ontology/
BIN/id/>
PREFIX reac_hands: <https://bioportal.bioontology.org/
ontologies/BIN/?p=classes&conceptid=http%3A%2F%2Fpurl
.bioontology.org%2Fontology%2FBIN%2FBIN_00000080#>
PREFIX hands_missed: <https://bioportal.bioontology
.org/ontologies/BIN/?p=classes&conceptid=http%3A%2F%2F
purl.bioontology.org%2Fontology%2FBIN%2FBIN_00000081#>

SELECT ?part ?value_time ?value_missed
FROM <[id\_repositare\_s\_daty]>
WHERE {
    ?part reac_hands:classPermalinkModal ?value_time ;
        hands_missed:classPermalinkModal ?value_missed .
    FILTER (?value_time <= 800 AND ?value_missed < 2)
}
GROUP BY ?part
ORDER BY ASC(?value_time)

```

### Výsledek testu dotazování

Výsledky z testu dotazování, který byl proveden pro všechny datové formáty RDF, jsou uspořádány v následující tabulce 6.2.

Serializace	SPARQL 1	SPARQL 2	SPARQL 3
RDF/XML	16 ms	15 ms	16 ms
Turtle	16 ms	16 ms	16 ms
N-Triples	16 ms	16 ms	16 ms
Notation3	15 ms	16 ms	16 ms
<b>Počet řádků</b>	25	160	22

**Tabulka 6.2:** Tak jako v při testování nahrávacích dob, i zde je vidět, že rychlost dotazování se je pro všechny serializace téměř totožná. Tento efekt je nejspíše způsoben tím, že samotná testovací data nejspíše nebyla dostatečně objemná.

Ve všech případech, kdy se testovaly dotazy napříč všemi sledovanými serializacemi, vždy bylo vrácen z úložiště Virtuoso stejný počet očekávaných řádků jednotlivých SPARQL dotazů. Důvodem proč jsou časy dotazování téměř totožné by mohlo být způsobeno tím, že buď Virtuoso uchovává načtená RDF data ve stejné struktuře a nebo nebyla testovací data dostatečně velká, aby se projevil různé způsoby dotazování.

## 6.2 Zhodnocení nasazení a otestování RDF úložiště

Bylo dokázáno, že konverze vstupního CSV souboru do čtyř vybraných RDF formátů (RDF/XML, Turtle, N-Triples a Notation3) probíhá alespoň na odzkoušených příkladech správně. Tyto formáty je opravdu možné nahrát do nasazeného RDF úložiště Virtuoso a je možné se na načtená data, v libovolném z otestovaných formátů, dotazovat.

## 7 Zhodnocení dosažených výsledků

Nejprve byla prozkoumána problematika sběru heterogenních dat v prostředí webové i mobilní aplikace pro sběr dat. Jako zdroj informací byly použity veřejně dostupné články [3], [4] a interní materiály ohledně systému BiN [26]. Byla také prostudována řada informací o fungování a myšlenkách sémantického webu. Řada informací byla velmi podnětná a dobře zpracována (např. články a standardy W3C), některá však nebyla zcela důvěryhodná. Na základě zhodnocení takto získaných informací byl vytvořen obrázek pravděpodobného fungování sémantického webu. Linked data jsou v současné době jedny z nejpoužívanějších technologií, které byly vyvinuty v rámci snahy o realizaci pravého sémantického webu. Linked data je způsob publikování a propojení strukturovaných dat pomocí odkazů, v prostředí webu, která jsou nejen čitelná lidmi, ale i stroji. Tato metoda zajišťuje snazší technické zálohování dat (data na různých strojích) a však nově navozuje otázku důvěryhodnosti a bezpečnosti dat. Vzhledem k tomu, že však je současným trendem nasazení relačních databázových systémů bylo nutno jí pro tuto aplikaci prostudovat a použít.

Nejprve byla prostudována terminologie používaná v rámci systému BiN primárně při sběru dat. Řada potřebných či používaných termínů neobsahovala jednoznačnou definici, a proto bylo přistoupeno ke studiu řady již publikovaných ontologií, či hledání ontologických best practices. Tyto ontologie pochází z dílny odborných pracovišť, je jich celá řada<sup>1</sup>. Tyto ontologie jsou značně rozsáhlé a velmi dobře zpracované, takže posloužili i jako praktické vodítko, jak ontologie správně vytvářet. Podobným způsobem pomohlo prozkoumávání doporučených postupů ve veřejné organizaci OBO Foundry. Nicméně ani jejich obsah nestačil k plnému pokrytí potřeb systému BiN. Proto bylo nutné vyhledat pomoc v obecných slovnících jako je například Webster, Oxford a podobné a s jejich pomocí dotvořit vlastní ontologii, do které bylo přidáno circa 80 nově definovaných termínů. Vzhledem k dobrým mravům, jsem tuto novou ontologii zpřístupnil veřejné obci v rámci systému BioPortal<sup>2</sup> pod Creative Commons licencí verze 4.0.

V prostředí internetu byla nalezena celá řada RDF úložišť v různém

---

<sup>1</sup>Například NCIT byl vytvořen Centrem pro biomedicínskou informatiku a informační technologie se sídlem v Bethesda, U.S.A.

<sup>2</sup>Tato ontologie je dostupná na <http://purl.bioontology.org/ontology/BIN>.



stupni kvality a rozpracovanosti. Některá jsou udržovaná a jiná již značně zastaralá. Během prvotního výběru byl také použit přehled populárních grafových<sup>3</sup> a RDF úložišť<sup>4</sup> a také článek [1], kde autor otestoval některá z těchto populárních úložišť. Vybrána úložiště z těchto recenzí byla odzkoušena a některá z nich i ihned zavržena. Do druhého kola výběru postoupila tři, která se jevila jako perspektivní nebo odpovídala kritériím. Tato úložiště byla podrobně ji otestována z různých hledisek (rychlost dotazování, rychlost načítání, snadnost ovládání/zprovoznění) a ukázalo se, že nelze jednoznačně určit vítěze. Nepodařilo se najít úložiště, které by jednoznačně splňovalo všechny podmínky. Proto je nutné vzít kompromisní řešení z hlediska účelu nasazení a tady jako nejlepší kompromis bylo zvoleno úložiště Virtuoso pro nasazení v systému BiN. Zbylé úložiště GraphDB lze doporučit pro méně náročné aplikace. Jeho výhodou však je podstatně jednodušší správa a uživatelsky přívětivější rozhraní. Pro požadované účely obě úložiště vyhovují. Z hlediska objemu dat jsou BiN databáze v řádech desítek tisíc trojic za dobu jejich sběru. To je poměrně malý objem a nelze předpokládat, že by se za nejbližších pět let razantně zvýšil vzhledem k velikosti pracoviště. Pro otestování výkonu byla použita testovací a veřejně dostupná RDF data, která je o několik řádů větší než BiN sesbíraná data. Výkon těchto úložišť byl pro testovací data více než dostatečný. Takže lze konstatovat, že pro BiN data je úložiště bohatě předdimenzováno, takže i pro poměrně dlouhou budoucnost bude svým výkonem a kapacitou vyhovovat.

Na závěr bylo podrobněji otestováno úložiště Virtuoso z několika různých hledisek a z řadou datových formátů. Výkon tohoto úložiště je tak velký, zejména pak ve spojení 12 jádrovým procesorem. Je problém určit jak dlouho operace opravdu trvají a to i při použití velkých testovacích dat. Vzhledem k tomu, že pracujeme v multi-threadovém systému, kde běží na pozadí řada servisů včetně údržby systému, které není možné rozumně sledovat jsou výsledky podstatně ovlivněny těmito procesy. Jelikož těchto procesů běží současně i několik set jsou časy do stovky milisekund nedůvěryhodné a lze tedy říct, že byly jistě lepší avšak i tyto hodnoty jsou vcelku impozantní. Proto je zcela zjevné, že toto úložiště bohatě překračuje požadavky na něj kladené. Daleko větší časy než doby vlastního zpracování centrální jednotkou budou dány transportem dat po internetu a to zejména v závislosti na transportní rychlosti, agregaci, propustnosti a zatížení sítě. Tyto časy nelze ani změřit ani odhadnout, protože závisí na umístění a okamžitém stavu sítě, každopádně i v tomto případě jsou jak nahrávání, čtení dat, tak databázové dotazy dostatečně rychlé. A to i za předpokladu, že data budou zasílána z

<sup>3</sup>Například stránka <https://db-engines.com/en/ranking/graph+dbms>.

<sup>4</sup>Například stránka <https://db-engines.com/en/ranking/rdf+store>.

velmi vzdálených zemích.

## 8 Závěr

Během této práce byly zmapovány nejen termíny používané při sběru heterogenních dat v rámci systému BiN, ale i autor této práce se osobně zúčastnil samotného procesu měření. Také byly prostudovány dostupné prameny zabývající se sémantickým webem a tyto nově nabyté znalosti byly využity během následných úkolů.

Během vytváření ontologie, která popisuje používanou terminologii v systému BiN, bylo maximalizováno znovupoužití již nadefinovaných termínů z jiných publikovaných ontologií. V případě, že některé definice termínů neexistovaly v již publikovaných ontologiích, tak byly vytvořeny vlastní definice. Jednalo se o circa 80 nových termínů, které bylo potřeba definovat a zveřejnit. Výsledná ontologie BIN je veřejně přístupná na následující webové adrese <https://bioportal.bioontology.org/ontologies/BIN>.

V současné chvíli umožňuje systém BiN export data do formátu CSV, takže byl vytvořen konverzní program v Pythonu, který umožňuje konverzi z formátu CSV do některého z podporovaných RDF formátů (např. RDF/XML, Turtle, N-Triples). Tento modul byl implementován přímo do infrastruktury BiN a je zprovozněn na stejném testovacím stroji jako bylo prováděno prvotní testování úložišť.

Byla vybrána tři vhodná RDF úložiště, která byla následně experimentálně odzkoušena a doporučena podle jejich vlastností a dodatečných parametrů. Během testování byla používána zkonvertovaná data z Dnů Vědy a Techniky 2017 (DVT 2017). Tento zkonvertovaný soubor obsahoval 7 326 trojic. Tato data nebyla sama o sobě dostatečně velká, aby posloužila k výkonnostnímu posouzení testovaného úložiště, a tak ještě byla použita řádově objemnější a veřejně dostupná data obsahující 120 let historie olympijských her (největší soubor obsahoval 7 809 410 trojic).

Doporučeným RDF úložištěm se tak stalo *Virtuoso* Open Source Edition, které bylo ve všech testech na rychlost načítání a dotazování se razantně rychlejší než ostatní testovaná. Toto úložiště bylo následně nasazeno v rámci infrastruktury BiN a bylo znovu otestováno pro načítání různých podporovaných RDF formátů (RDF/XML, Turtle, N-Triples a Notation3). Správnost konverze a naplnění dat, byla následně ověřena pomocí vytvořených SPARQL dotazů.

Byly tak splněny všechny požadované body zadání. V některých případech bylo nabídnuto i několik validních řešení či možností v závislosti na tom, jak se bude aplikace v budoucnosti vyvíjet.

# A Seznam zkratek

Použitá zkratka	Vysvětlení
<b>API</b>	Application Programming Interface
<b>BCI</b>	Brain-Computer interface
<b>BiN</b>	Systém Body in Numbers
<b>BMI</b>	Body Mass Index
<b>CSV</b>	Comma-Separated Values
<b>CSVW</b>	CSV on the Web
<b>ER diagram / ERD</b>	Entity-Relationship diagram
<b>FEV1</b>	Časová vitální kapacita
<b>HTTP</b>	Hypertext transfer protocol
<b>IRI</b>	Internationalized Resource Identifier
<b>JSON</b>	JavaScript Object Notation
<b>JSON/LD</b>	JavaScript Object Notation for Linked Data
<b>KIV</b>	Katedra Informatiky a Výpočetní techniky
<b>NaN</b>	Not a Number
<b>NCIT</b>	National Cancer Institute Thesaurus
<b>OBO</b>	Biomedical ontologies
<b>OLS</b>	Ontology Lookup Service
<b>OWL</b>	Web Ontology Language
<b>PEF</b>	Maximální výdechový proud vzduchu
<b>PURL</b>	Persistent URL
<b>RDF</b>	Resource Description Framework
<b>RDFS</b>	Resource Description Framework Schema
<b>SPARQL</b>	SPARQL Protocol and RDF Query Language
<b>UML</b>	Unified Modeling Language
<b>VC</b>	Vitální kapacita plic
<b>W3C</b>	World Wide Web Consortium

## B Seznam literatury

- [1] ANGUS ADDLESEE. Comparing Linked Data Triplestores. <https://medium.com/wallscope/comparing-linked-data-triplestores-ebfac8c3ad4f>, 2019. Online; [12.05.2019].
- [2] BOHMANN, D. Mobilní aplikace pro sběr medicínských dat. Master's thesis, University of West Bohemia, Czech Republic, 2017.
- [3] BRUHA, P. et al. Exercise and Wellness Health Strategy Framework-Software Prototype for Rapid Collection and Storage of Heterogeneous Health Related Data. *BIOSTEC*. 2017, 2017, s. 477.
- [4] BRŮHA, P. et al. Collection of human reaction times and supporting health related data for analysis of cognitive and physical performance. *Data in brief*. 2018, 17, s. 469–511.
- [5] DB-ENGINES. RDF Stores. <https://db-engines.com/en/article/RDF+Stores>, 2019. Online; [cit. 24.02.2019].
- [6] FACEBOOK. The Open Graph protocol. <http://ogp.me/>, 2017. Online; [cit. 24.02.2019].
- [7] JUPP, S. et al. A new Ontology Lookup Service at EMBL-EBI. In *SWAT4LS*, s. 118–119, 2015.
- [8] OBO FOUNDRY. OBO Foundry principles. <http://www.obofoundry.org/principles/fp-000-summary.html>, 2019. Online; [05.05.2019].
- [9] OPEN SEMANTIC FRAMEWORK. Ontology Best Practices. [http://wiki.opensemanticframework.org/index.php/Ontology\\_Best\\_Practices](http://wiki.opensemanticframework.org/index.php/Ontology_Best_Practices), 2014. Online; [05.05.2019].
- [10] THE NEW YORK TIMES. A 'more revolutionary' Web. <https://www.nytimes.com/2006/05/23/technology/23iht-web.html>, 2006. Online; [cit. 24.02.2019].

- [11] TIM BERNERS LEE, JAMES HENDLER AND ORA LASSILA. The Semantic Web. [https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American\\_%20Feature%20Article\\_%20The%20Semantic%20Web\\_%20May%202001.pdf](https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf), 2001. Online; [cit. 24.02.2019].
- [12] TWO-BIT HISTORY. Whatever Happened to the Semantic Web? <https://twobithistory.org/2018/05/27/semantic-web.html>, 2018. Online; [cit. 24.02.2019].
- [13] VANDENBUSSCHE, P.-Y. et al. Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web, 01 2017.
- [14] W3C. OWL 2 Web Ontology Language Primer (Second Edition). <https://www.w3.org/TR/2012/REC-owl2-primer-20121211/>, 2012. Online; [02.05.2019].
- [15] W3C. OWL 2 Web Ontology Language Profiles (Second Edition). <https://www.w3.org/TR/owl2-profiles/>, 2012. Online; [02.05.2019].
- [16] W3C. OWL Web Ontology Language Overview. <https://www.w3.org/TR/owl-features/>, 2004. Online; [02.05.2019].
- [17] W3C. JSON-LD 1.0 A JSON-based Serialization for Linked Data. <https://www.w3.org/TR/json-ld/>, 2014. Online; [01.03.2019].
- [18] W3C. RDF 1.1 Turtle Terse RDF Triple Language. <https://www.w3.org/TR/turtle/>, 2014. Online; [01.03.2019].
- [19] W3C. RDF 1.1 XML Syntax. <https://www.w3.org/TR/rdf-syntax-grammar/>, 2014. Online; [01.03.2019].
- [20] W3C. OWL 2 Web Ontology Language Document Overview (Second Edition). <https://www.w3.org/TR/owl2-overview/>, 2012. Online; [25.02.2019].
- [21] W3C. Semantic web ontologies. <https://www.w3.org/standards/semanticweb/ontology>, 2015. Online; [05.05.2019].
- [22] WHETZEL, P. L. et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic acids research*. 2011, 39, suppl\_2, s. W541–W545.

- [23] XIANG, Z. et al. OntoFox: web-based support for ontology reuse. *BMC research notes*. 2010, 3, 1, s. 175.
- [24] XIANG, Z. et al. Ontobee: A linked data server and browser for ontology terms. In *ICBO*, 2011.
- [25] YU, L. *A developer's guide to the semantic Web*. Springer Science & Business Media, 2011.
- [26] ŠNEJDAR, P. Vytvoření aplikace pro sběr medicínských dat v rámci projektových dnů. Master's thesis, University of West Bohemia, Czech Republic, 2018.

# C Seznam obrázků

2.1	Pro každou věkovou kategorii může probíhat sběr, předzpracování (případně i výzkum) a publikace dat trochu jinak. Věková rozmezí pro každou z uvedených kategorií jsou následující: starší školní věk je od 11 do 14 let, dospívání je od 15 do 18 let, plná dospělost je od 19 do 30 let, mladý věk je od 31 do 45 let a střední věk je od 46 do 60 let. . . . .	11
2.2	Datový cyklus naměřených dat v projektu BiN. Jednotlivé fáze datového cyklu byly již nastíněny na předchozím obrázku 2.1. . . . .	13
2.3	Definice projektového dne v rámci terminologie systému BiN.	15
2.4	Rozdělení architektury systému BiN dle Docker kontejnerů. Bude vytvořen nový modul do webové aplikace, který bude řešit konverzi dat. V této architektuře je také obsažen sporný modul Logger. . . . .	16
2.5	Uspořádání modulů v rámci webové aplikace BiN včetně přidání modulu pro konverzi naměřených dat do RDF. . . .	17
2.6	Stimulační matice použitá v projektu BASIL. Vždy se aktivuje pouze jeden sloupec a nebo řádek této matice. Na obrázku byl aktivován první sloupec matice, který obsahuje následující činnosti v pořadí od shora dolů: otevřít dveře, rozsvítit světlo a jít na toaletu. . . . .	20
2.7	Příklad datového cyklu z pohledu stanoviště BCI. Experiment se zaměřoval primárně na rozdělení účastníků do kategorií v závislosti na čitelnosti vlny P300. . . . .	21
2.8	Google využívá sémantické technologie například pro vytvoření konceptuálních shrnutí, která se zobrazují vpravo od výsledků vyhledávání. . . . .	24
2.9	Jednoduchý příklad obsahující tři příklady RDF trojic včetně výsledného RDF grafu. . . . .	26
2.10	Příklad formátu RDF/XML. Zdroj: Example 2 [19]. . . . .	27
2.11	Příklad RDF formátu Turtle. Každá z RDF trojic je ukončena tečkou. Tato vzorová data popisují vztah mezi Green Goblmem a Spidermenem. Zdroj: Example 1 [18]. . . . .	28



2.12	Příklad JSON dokumentu obsahujícím jméno, homepage a obrázek. Zdroj: Example 1 [17]. . . . .	29
2.13	Příklad upraveného JSON/LD dokumentu. Změny oproti JSON dokumentu jsou zvýrazněny. Zdroj: Example 2 [17]. . . . .	29
4.1	Vstupního CSV soubor obsahoval následující informace. Základní nastavení skriptu doplňuje před všechny atributy (první řádek CSV souboru) textový řetězec "http://". Tato funkci samozřejmě lze upravit buď ve vstupním souboru a nebo lze tento efekt potlačit pomocí přepínače <code>--schemanamespace</code> . Proto v tomto příkladě neobsahují IRI převzaté z ontologií tento textový řetězec. . . . .	46
4.2	Na tomto obrázku je uveden obsah stejného vstupního CSV souboru jako v předchozím případě pouze v prostředí tabulkového editoru. Modře zvýrazněné hodnoty obsahují běžné textové atributy. Žluté zvýraznění obsahuje atributy, které jsou uvedeny formou IRI (platný odkaz do ontologií) a zelená pole obsahují textový řetězec korespondující s prázdnou hodnotou. V tomto případě se jedná o textový řetězec NaN. . . . .	47
4.3	Zde je uveden výstup z konverzního programu CSV2RDF, kde výsledný RDF soubor je zapsán v notaci Turtle. Tento výsledek byl dosažen pomocí dříve uvedeného spouštěcího příkazu. . . . .	47
4.4	Zde je uveden výstup z konverzního programu CSV2RDF, kde výsledný RDF soubor je zapsán v notaci RDF/XML. Byl použit podobný spouštěcí příkaz jako v předchozím případě na obrázku 4.3, pouze byl vyměněn přepínač <code>--format "turtle"</code> za <code>--format "xml"</code> . . . . .	48
4.5	Příklad SPARQL dotazu v prostředí Jena Fuseki Server, kde byla vypsána uložená testovací data (původně načtena ve formátu Turtle = celkem 3355 trojic). . . . .	50
4.6	Příklad SPARQL dotazu v prostředí Virtuoso Conductor, kde byla vypsána uložená testovací data (původně načtena ve formátu Turtle = celkem 3355 trojic). . . . .	51
4.7	Příklad SPARQL dotazu v prostředí GraphDB, kde byla vypsána uložená testovací data (původně načtena ve formátu Turtle = celkem 3355 trojic). . . . .	52
5.1	Výsledky testování všech vybraných úložišť, které dosáhly v průměru při načítání Turtle formátu RDF. . . . .	62

5.2	Výsledky testování všech vybraných úložišť, které dosáhly v průměru při načítání N-Triples formátu RDF. . . . .	63
5.3	Výsledky dosažené při testování rychlosti dotazování úložišť na SPARQL dotazech jedna až čtyři (viz kapitola Použití SPARQL dotazy 5.3), které využívali příkazy na jedinečnost výsledků ( <b>DISTINCT</b> ) a běžné agregační funkce ( <b>AVG</b> , <b>MAX</b> , <b>COUNT</b> ). Uvedené doby v milisekundách jsou výsledkem aritmetických průměrů ze tří měření pro každé z uvedených úložišť zaokrouhlenými na celé milisekundy. . . . .	64
5.4	Rychlost odpovídání na složitější SPARQL dotazy. Tyto dotazy obsahovali jak vnořené dotazy, dotazování se na externí službu a využívání regulárního výrazu. I v tomto případě vítězí na celé čáře Virtuoso. Uvedené doby v milisekundách jsou výsledkem aritmetických průměrů ze tří měření pro každé z uvedených úložišť zaokrouhlenými na celé milisekundy. . . .	64

# D Seznam tabulek

3.1	Přehled celkem uložených ontologií v jednotlivých ontologických portálech. Některé ontologie jsou uloženy i ve více portálech najednou. Příklad takovéto ontologie může být NCIT (National Cancer Institute Thesaurus), kde tato ontologie je uložena ve všech uvedených portálech. Uvedené počty jsou vztaženy k datu 2019-05-05 (RRRR-MM-DD). . . . .	33
5.1	Zde jsou uvedeny jednotlivé velikosti vstupních souborů formátu Turtle, kde první a nejmenší soubor obsahuje převedená data ze systému BiN. Jedná se data změřená ze Dnů vědy a Techniky 2017 v Plzni. Zbývající data byla převzata z veřejných RDF dat obsahující až 120 let historie olympijských her, která byla případně převedena do Turtle formátu. . . . .	54
5.2	Zde je možné vidět, že data, která obsahují stejnou sémantickou hodnotu a pouze jsou zapsána v jiné RDF serializaci (oproti tabulce 5.1) mají mnohem větší objem. Hlavní rozdíl formátu N-Triples oproti formátu Turtle je v tomto případě velikost vstupního souboru, která je zhruba 2x až 4x větší oproti Turtle formátu. . . . .	54
6.1	Prezentuje, že každá z uvedených serializací popisovalo stejné množství trojic. Tento fakt ověřuje, že data byla převedena správným způsobem do různých serializací a zároveň byla i správně načtena do úložiště. Bohužel data, která sice obsahují desetitisíce trojic jsou stále ještě moc malá, než aby jejich doba načítání byla razantně rozdílná. . . . .	67
6.2	Tak jako v při testování nahrávacích dob, i zde je vidět, že rychlost dotazování se je pro všechny serializace téměř totožná. Tento efekt je nejspíše způsoben tím, že samotná testovací data nejspíše nebyla dostatečně objemná. . . . .	70

# E Terminologie BiN

ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvisí s jiným termínem v BiN ontologii?
1	Measurement	BiN measurement	Měření	Definice všech stanovišť a experimentů s přidělenými pomůckami. Jedná se o upřesnění toho, co a jak se bude v daný projektový den měřit.	A measurement contains the concrete collection of experiments and tools/devices that will be used during the project days.	<a href="http://purl.bioontology.org/ontology/SNO-MEDCT/122869004">http://purl.bioontology.org/ontology/SNO-MEDCT/122869004</a>	An observation, by some objective method, of amount, number, quantity, size, level, extent, or magnitude, resulting in an ordinal or quantitative value	Ne	Ano (Plánování)
2	Experiment		Experiment	Jedno vyšetření účastníka provedené podle příslušného scénáře. Například měření krevního cukru vyžaduje píchnutí do prstu, náběrání krve na měřicí proužek a následně vyhodnocení výsledků glukometrem.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C42790">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C42790</a>	A coordinated set of actions and observations designed to generate data, with the ultimate goal of discovery or hypothesis testing.	Ano	Ano (Stanoviště)
3	Tool		Pomůcka	Přístroje a nebo nástroje, které jsou používány během experimentů. Například tlakoměr pro měření tlaku a tepu, spirometr pro měření kapacity plic nebo metr pro měření výšky.		<a href="http://purl.obolibrary.org/obo/ERO_0001491">http://purl.obolibrary.org/obo/ERO_0001491</a>	A measurement instrument used for precise measurements.	Ano	Ano (Experiment)
4	Station	Measurement station	Stanoviště	Místo určené k měření, které může obsahovat jeden nebo více experimentů.	A location equipped with special equipment and personnel for a particular purpose. In the BiN system, the measurement station also specifies which and how many experiments will take place and what kind of equipment is needed to perform those experiments.	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C67102">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C67102</a>	A facility or location equipped with special equipment and/or personnel for a particular purpose.	Asi ano	Ano (Měření)
5	Planning	BiN planning	Plánování	Založení nového měření, které obsahuje nové definice toho co, jak a pomocí čeho se bude měřit.	Planning defines what kind of experiments and tools / devices will be used during the measurements. This specification provides all the necessary items that will be needed during the experiments.	<a href="http://purl.obolibrary.org/obo/OBI_0000339">http://purl.obolibrary.org/obo/OBI_0000339</a>	a process of creating or modifying a plan specification	Ne	Ne
6	Participant		Účastník	Osoba, na které se provádí měření.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C29867">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C29867</a>	Someone who takes part in an activity.	Ano	Ano (Měření)
7	Device		Přístroj	Nástroj nebo přístroj, který je používán jako pomůcka při realizaci experimentu		<a href="http://purl.obolibrary.org/obo/OBI_0000968">http://purl.obolibrary.org/obo/OBI_0000968</a>	A material entity that is designed to perform a function in a scientific investigation, but is not a reagent.	Ano	Ano (Pomůcka)
8	Unit of measurement		Měřená jednotka	Fyzikální jednotka, ve které dany přístroj měří.		<a href="http://semanticscience.org/resource/SIO_0000074">http://semanticscience.org/resource/SIO_0000074</a>	A unit of measurement is a definite magnitude of a physical quantity, defined and adopted by convention and/or by law, that is used as a standard for measurement of the same physical quantity.	Ano	Ano (Přístroj)
9	Identifier		Identifikátor	Jednoznačné určení přístrojů, záznamů, apod.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25364">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25364</a>	One or more characters used to identify, name, or characterize the nature, properties, or contents of a thing.	Ano	Ano (Přístroj)
10	Device name		Název přístroje	Jméno nebo číselné označení pomůcky/přístroje.		<a href="http://dicom.nema.org/resources/ontology/DCMT/113877">http://dicom.nema.org/resources/ontology/DCMT/113877</a>	The name used to refer to a device; usually locally unique.	Asi Ano	Ano (Přístroj)
11	Experimental design	Experimental design and studies	Kategorie experimentu	Definice, čím se experiment zabývá (měření EEG, pouze měření pomocí váhy, např. používá scénář nebo ne).		<a href="http://edamontology.org/topic_3678">http://edamontology.org/topic_3678</a>	The design of an experiment intended to test a hypothesis, and describe or explain empirical data obtained under various experimental conditions.	Asi Ano	Ano (Experiment)
12	Pilot studies	Pilot studies	Pilotní experiment	Ověření experimentu, ze kterého se mohou vytvořit další experimenty	Small-scale tests of methods and procedures to be used on a larger scale if the pilot study demonstrates that these methods and procedures can work.	<a href="http://scdontology.h3abionet.org/ontology/SCDO_0000905">http://scdontology.h3abionet.org/ontology/SCDO_0000905</a>	Small-scale tests of methods and procedures to be used on a larger scale if the pilot study demonstrates that these methods and procedures can work.	Asi Ano	Ano (Kategorie experimentu)
13	Quasi-experiment	Quasi experimental methods	Quasi-experiment	Experiment, který využívá komplexní scénář pro svoje měření s omezením náhodnosti měření. Například provádíme experiment v kontrolovaném prostředí.	Research conducted in settings or environments where normal or traditional controls are not, or cannot be applied.	<a href="http://ontology.gsa.org/apaonto/terms/only/OUT_5.owl#Quasi_Experimental_Methods">http://ontology.gsa.org/apaonto/terms/only/OUT_5.owl#Quasi_Experimental_Methods</a>	Research conducted in settings or environments where normal or traditional controls are not, or cannot be applied.	Asi Ano	Ano (Kategorie experimentu)
14	Natural experiment	Natural experiment	Natural experiment	Experiment, který využívá komplexní scénář pro svoje měření bez možnosti ovlivnit prostředí.		<a href="http://purl.dataone.org/obo/ECSO_0000507">http://purl.dataone.org/obo/ECSO_0000507</a>	A natural experiment is an empirical study in which individuals (or clusters of individuals) exposed to the experimental and control conditions are determined by nature or by other factors outside the control of the investigators, yet the process governing the exposures arguably resembles random assignment.	Asi Ano	Ano (Kategorie experimentu)
15	Category	Device category	Kategorie přístrojů	Definice, o jaký typ přístroje/nástroje/pomůcky se jedná.	Dividing available devices into separate categories to make it easier to distinguish the purpose of each device within a particular category.			Ne	Ano (Přístroj)
16	Amplifier		Zesilovač	Přístroj, který slouží k zesílení signálů například z změnění pomocí EEG čepice.		<a href="http://purl.obolibrary.org/obo/ERO_0000945">http://purl.obolibrary.org/obo/ERO_0000945</a>	An instrument that changes, usually increases, the amplitude of a signal. The relationship of the input to the output of an amplifier, usually expressed as a function of the input frequency, is called the transfer function of the amplifier, and the magnitude of the transfer function is termed the gain.	Ano	Ano (Kategorie přístrojů)
17	Measuring instrument		Měřicí nástroj	Jedná se o mechanický měřicí přístroj, například váha, siloměr, ...		<a href="http://purl.obolibrary.org/obo/ERO_0000004">http://purl.obolibrary.org/obo/ERO_0000004</a>	Material entity that is designed to have a function and play a role in scientific investigation.	Asi Ano	Ano (pomůcky)
18	Agreement	Agreement to a specified research addendum	Souhlas s účastí v experimentu	Dokument, ve kterém účastník souhlasí s měřením.		<a href="http://purl.obolibrary.org/obo/OBIB_0000489">http://purl.obolibrary.org/obo/OBIB_0000489</a>	An authorization documentation in which a legally authorized individual agrees to a specific research addendum for the person they represent.	Asi Ano	Ano (Měření)
19	Name		Jméno	Jméno (křestní), účastníka experimentu.		<a href="http://purl.bioontology.org/ontology/SNO-MEDCT/703503004">http://purl.bioontology.org/ontology/SNO-MEDCT/703503004</a>	Word or set of words by which a person is known, addressed, or referred to.	Ano	Ano (Účastník)

ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvise s jiným termínem v BIN ontologii?
20	Surname		Příjmení	Příjmení, účastníka experimentu.		<a href="http://publ.bioontology.org/ontology/SNO/MEDCT/39767800">http://publ.bioontology.org/ontology/SNO/MEDCT/39767800</a>	Family name of person.	Ano	Ano (Účastník)
21	Date of birth	Birth Date	Datum narození	Datum narození účastníka.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C68615">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C68615</a>	The calendar date on which a person was born.	Ano	Ano (Účastník)
22	Laterality		Laterality	Dominantní ruka účastníka (např. pro psaní nebo držení přístroje).		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25185">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25185</a>	Dominant use or manifestations of one side of the body versus the other; referring to a side of the body or of a structure.	Ano	Ano (Účastník)
23	Environment		Prostředí	V jakém prostředí probíhá měření, zda je uvnitř budovy, kde můžeme uvlivnit některé rušivé elementy a nebo venku např. na ulici	The external elements and conditions which surround, influence, and affect the life and development of an organism or population.	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C16551">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C16551</a>	The totality of surrounding conditions.	Ne	Ano (Experiment)
24	Controlled environment		Laboratorní prostředí	Provádění uvnitř relativně uzavřeného prostředí, kde jsme schopni omezit určité množství nežádoucích vlivů.		<a href="http://publ.bioontology.org/ontology/MESH/D004780">http://publ.bioontology.org/ontology/MESH/D004780</a>	A state in which the environs of hospitals, laboratories, domestic and animal housing, work places, spacecraft, and other surroundings are under technological control with regard to air conditioning, heating, lighting, humidity, ventilation, and other ambient features. The concept includes control of atmospheric composition. (From Jane's Aerospace Dictionary, 3d ed)	Ano	Ano (Prostředí)
25	Non-Laboratory environment	Uncontrolled environment	Nefixované prostředí	Prostředí, které se náhodně mění vnějšími vlivy, většinou náhodně.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C48932">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C48932</a>	Not controlled, regulated, or constrained.	Asi Ano	Ano (Prostředí)
26	Temperature		Teplota	Průměrná teplota laboratorního a nebo ne-laboratorního prostředí před začátkem měření.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25206">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25206</a>	The property of a body or region of space that determines whether or not there will be a net flow of heat into it or out of it from a neighboring body or region and in which direction (if any) the heat will flow, perceptible by living organism as a somatic sensation of cold or heat. It is a measure of the average translational kinetic energy associated with the disordered microscopic motion of atoms and molecules. Temperature is measured in one of the three standard temperature scales: Celsius, Kelvin, and Fahrenheit. (NCI)	Ano	Ano (Prostředí)
27	Weather		Počasí	Slav vnějšího prostředí.		<a href="http://cerrado.linkeddata.es/ecology/fire#Weather">http://cerrado.linkeddata.es/ecology/fire#Weather</a>	Weather is the state of the atmosphere, to the degree that it is hot or cold, wet or dry, calm or stormy, clear or cloudy. Generally refers to day-to-day temperature and precipitation activity.	Ano	Ano (Prostředí)
28	Humidity		Vlhkost vzduchu	Procentuální zastoupení vody ve vzduchu uvnitř daného prostředí.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C68206">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C68206</a>	A measure of the water vapor content of air.	Ano	Ano (Prostředí)
29	Questionnaire	BIN Questionnaire	Dotazník	Papírový nebo elektronický dokument, který slouží k sesbírání dodatečných metadat před začátkem měření.	A document with a set of printed or written questions with a choice of answers, devised for the purposes of a research study. Another usage of the questionnaire is to collect enough metadata that capture the current physical and psychological state of the participant before the measurement was taken.	<a href="http://publ.obolibrary.org/obo/OBI_0001000">http://publ.obolibrary.org/obo/OBI_0001000</a>	A document with a set of printed or written questions with a choice of answers, devised for the purposes of a survey or statistical study.	Ne	Ano (Měření)
30	Category	Types of survey questions	Kategorie otázek	Na jakou část povahy účastníka, je daná otázka směřována.	They define on which aspect of the participant are the questions being aimed. They can for example focus on the current health status of the participant or their psychological status, eating habits, physical fitness and many more.			Ne	Ano (Dotazník)
31	Psychic	Psychic state questions	Psychika	Otázky ohledně psychického stavu účastníka (nejčastěji současného).	Questions aimed at the current (sometimes past) psychic state of the participant. Those questions can for example be: - If the participant is happy or not. - The amount of stress the participant is currently facing. - Amount of available finances. and many more...	<a href="http://publ.obolibrary.org/obo/NCIT_C16839">http://publ.obolibrary.org/obo/NCIT_C16839</a>	mental health status.	Asi Ano	Ano (Kategorie otázek)
32	Eating habits	Eating habits questions	Stravování	Otázky zaměřené na stravovací návyky účastníka (jídlo, pití, alkohol kouření).	Questions aimed at the eating habits of the participant. Eating habits focused questions can try to answer following things: - Amount of food eating during each phase of the day. - Food composition (what kind of vegetables, how many, what kind of fruits, what kind of meat and so on...) - Alcohol drinking (frequency and amount), and many more...	<a href="http://publ.bioontology.org/ontology/MESH/D005247">http://publ.bioontology.org/ontology/MESH/D005247</a>	Acquired or learned food preferences.	Asi Ano	Ano (Kategorie otázek)
33	Physical fitness	Physical fitness questions	Fyzický stav	Otázky odlehně fyzické zdatnosti a pravidelného pohybu.	Physical fitness questions are trying to describe the current state of the participant's body. Those questions can be aimed at the following things: - Physical body strength - Hand eye coordination - Sight - Balance - Stamina and more...	<a href="http://publ.bioontology.org/ontology/MESH/D010809">http://publ.bioontology.org/ontology/MESH/D010809</a>	The ability to carry out daily tasks and perform physical activities in a highly functional state, often as a result of physical conditioning.	Asi Ano	Ano (Kategorie otázek)

ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvise s jiným termínem v BIN ontologii?
34	Sunny	Sunny weather	Slunečno	Počasí s velkým množstvím slunečního svítu, kdy nebe zároveň může obsahovat malé množství mraků.	The sort of weather where the skies are blue, the grass is bright, and there is lots of sunlight! It is generally warmer when the sun is out, so people in colder climates find it quite pleasant. People located in hotter climates may not be so fond of it.			Ne	Ano (Počasí)
35	Cloudy	Cloudy weather	Zataženo / Oblačno	Počasí s malým množstvím slunečního svítu, kdy nebe obsahuje velké množství mraků.	The sort of weather where the skies are full of or covered with clouds.			Ne	Ano (Počasí)
36	Rainy	Rainy weather	Děšť	Počasí, kdy aktuálně venku prší (nepočítá se i mholení).	Rainy weather is marked by an abundance of clouds that are bringing in a long term rain.			Ne	Ano (Počasí)
37	Snowy	Snowy weather	Sněžení	Počasí, kdy aktuálně venku padá sníh.	A snowy day is a day when a lot of snow has fallen.			Ne	Ano (Počasí)
38	Computer		PC	Osobní počítač.		<a href="http://purl.obolibrary.org/obo/OBI_0400107">http://purl.obolibrary.org/obo/OBI_0400107</a>	A computer is an instrument which manipulates (stores, retrieves, and processes) data according to a list of instructions.	Ano	Ano (Kategorie přístrojů)
39	Electrode		Elektroda	Předmět, který snímá elektrický proud a zároveň ho předává k měřicímu zařízení.		<a href="http://purl.bioontology.org/ontology/CRIS/P/1000-5566">http://purl.bioontology.org/ontology/CRIS/P/1000-5566</a>	conductor by which a current enters or leaves a nonmetallic medium; in electrotherapy, electrodes are instruments with a point or surface from which to transmit electric current to the body of a patient or to another instrument; in electrodiagnosis, they are needles or metal plates.	Ano	Ano (Kategorie přístrojů)
40	Grounding electrode		Zemnicí elektroda	Elektroda, která je spojená se zemním potenciálem.	The primary purpose of the ground electrode is to prevent power line noise from interfering with the small biopotential signals of interest. By design, amplifiers should not be affected by large changes in potential at both the active and reference sites. A ground electrode for EEG recordings is often placed on the forehead (but could be placed anywhere else on the body; the location of the ground on the subject is generally irrelevant).			Ne	Ano (Elektroda)
41	Reference electrode		Referenční elektroda	Elektroda, vůči které se vztahuje naměřený proud.		<a href="http://purl.obolibrary.org/obo/CHMO_0002345">http://purl.obolibrary.org/obo/CHMO_0002345</a>	An electrode through which no appreciable current is allowed to flow and is used to observe or control the potential at a working electrode.	Asi Ano	Ano (Elektroda)
42	Accessories		Příslušenství	Doplňkové části, které se mohou přidat k základní věci a tím zvýšit její užitečnost.	A thing which can be added to something else in order to make it more useful, versatile, or attractive.			Ne	Ano (Pomůcka)
43	Cleaning paste		Pasta na očištění	Pasta, která slouží k odstranění nečistot z povrchu.	A type of paste created and formulated to clean a variety of things.  Inside the BIN project, the cleaning paste is mainly used for removing any impurities from participants forehead and earlobes and thus provides an easier way to fasten referential and grounding electrodes that are necessary for EEG recording.			Ne	Ano (Příslušenství)
44	Conductive paste		Vodivá pasta	Pasta, která slouží ke snížení přechodového odporu, nejčastěji umístěné mezi elektrodou a povrchem lidského těla.	A jelly-like paste capable of effecting a low electrical contact resistance between the skin and an electrode, as used in electrocardiography.			Ne	Ano (Příslušenství)
45	Electrode gel	Electrolytic gel	Elektrode gel	Vodivý gel, který slouží pro usnadnění přenosu elektrického proudu mezi povrchem lidského těla a elektrodou.	Electrolytic gel is a solution or molten substance that conducts electricity.			Ne	Ano (Příslušenství)
46	EEG cap (large)	Large sized eeg cap	EEG čepice (velká)	Obvod hlavy 58 - 62 cm, velká	The large EEG cap is made of elastic spandex fabric with embedded clean electrodes attached to the fabric and is made to accommodate 58-62 cm of head circumference.			Ne	Ano (Drátová EEG zařízení)
47	EEG cap (medium)	Medium sized eeg cap	EEG čepice (střední)	Obvod hlavy 54 - 58 cm, střední	The medium sized EEG cap is made of elastic spandex fabric with embedded clean electrodes attached to the fabric and is made to accommodate 54 - 58 cm of head circumference.			Ne	Ano (Drátová EEG zařízení)
48	EEG cap (small)	Small sized eeg cap	EEG čepice (malá)	Obvod hlavy 50 - 54 cm, malá	The small sized EEG cap is made of elastic spandex fabric with embedded clean electrodes attached to the fabric and is made to accommodate 50 - 54 cm of head circumference.			Ne	Ano (Drátová EEG zařízení)
49	Software		Software	Programové vybavení osobního počítače, které slouží k určitému účelu		<a href="http://purl.bioontology.org/ontology/MESH/H1/D012984">http://purl.bioontology.org/ontology/MESH/H1/D012984</a>	Sequential operating programs and data which instruct the functioning of a digital computer.	Asi Ano	Ano (PC)
50	Recording software		Nahrávací software	Programové vybavení PC primárně určené pro nahrávání EEG.	An EEG recording software is solution for recording EEG time-series data together with all the necessary markers that are sent from the stimulation software.			Ne	Ano (Software)

ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvisí s jiným termínem v BIN ontologii?
51	Stimulation software		Stimulační software	Programové vybavení PC určené pro stimulaci účastníka.	<p>A piece of software that is dedicated to stimulate someone or something to become more active or enthusiastic, or to develop or operate.</p> <p>Inside the BIN project the stimulation software is used to create and present different scenarios to the participant and see how he/she will react to. Stimulus is usually either visual (light flickering, viewing a presentation, ...) or sound (high pitched or low pitched sound).</p> <p>The scenario for the participant usually contains an objective. Such objectives can be:</p> <ul style="list-style-type: none"> <li>- Count the number of occurrences of a high pitched sound.</li> <li>- How many times was the letter Q displayed.</li> <li>and many more...</li> </ul>			Ne	Ano (Software)
52	Celsius		Celsius	Jednotka teploty ve stupních celsia.		<a href="http://ecoinformatics.org/obo/owl/oboe.1.2/oboe-standards.owl#Celsius">http://ecoinformatics.org/obo/owl/oboe.1.2/oboe-standards.owl#Celsius</a>	The degree Celsius is a unit of thermodynamic temperature with the same magnitude as the unit kelvin but where absolute zero (0 Kelvin) is defined as -273.15 degrees Celsius.	Ano	Ano (Teplota)
53	Kelvin		Kelvin	Jednotka absolutní teploty v kelvinech.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C42537">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C42537</a>	A basic unit of thermodynamic temperature, one of the seven base units of the International System of Units (Système International d'Unités, SI). It is 1/273.16th of the thermodynamic temperature of the triple point of water. This sets the size of the kelvin unit for temperature differences and defines the thermodynamic temperature of an equilibrium mixture of waters ice-liquid-vapor as 273.16 K, where 0 K is the lowest possible temperature ("absolute zero").	Ano	Ano (Teplota)
54	Fahrenheit		Fahrenheit	Jednotka teploty v fahrenheitech.		<a href="http://purl.obolibrary.org/obo/UG_0000195">http://purl.obolibrary.org/obo/UG_0000195</a>	In this scale, the freezing point of water is 32 degrees Fahrenheit and the boiling point is 212 degrees, placing the boiling and freezing points of water 180 degrees apart. -40 degrees Fahrenheit is equal to -40 degrees Celsius. A temperature unit which is equal to 5/9ths of a kelvin. Negative 40 degrees Fahrenheit is equal to negative 40 degrees Celsius.	Ano	Ano (Teplota)
55	Hospital		Nemocnice	Instituce sloužící pro řešení medicínských problémů hospitalizačním způsobem.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C16696">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C16696</a>	An institution that provides medical, surgical, or psychiatric care and treatment for the sick or the injured.	Ano	Ano (Ne-laboratorní prostředí)
56	Home environment	Home	Domácí prostředí	Domácí prostředí účastníka.	Home environment is a specific place where one lives permanently, especially as a member of a family or household. A person inside ones home usually feels safe, comfortable and at ease.			Ne	Ano (Ne-laboratorní prostředí)
57	Outdoor environment	Outdoor	Venkovní prostředí	Otevřené prostředí, kde dochází k častému kontaktu s nekontrolovaným okolím.		<a href="http://purl.obolibrary.org/obo/ExO_0000n20">http://purl.obolibrary.org/obo/ExO_0000n20</a>	An exogenous location that is happening or arising or located outside	Asi Ano	Ano (Ne-laboratorní prostředí)
58	Wearable electronics	Wearable electronic devices	Nositelná elektronika	Elektronické přístroje, které mají déle trvající kontakt s pokožkou za účelem měření určitých údajů. Např. SmartWatch.		<a href="http://purl.bioontology.org/ontology/MESH/D000076251">http://purl.bioontology.org/ontology/MESH/D000076251</a>	Electronic implements worn on the body as an implant or as an accessory. Examples include wearable diagnostic devices, wearable ACTIVITY TRACKERS, wearable INFUSION PUMPS, wearable computing devices, SENSORY AIDS, and electronic pest repellents.	Ano	Ano (Přístroj)
59	PsychoPy		PsychoPy	Stimulační software, ve kterém se vytváří stimulační scénář pomocí Pythonu	PsychoPy is an open source software package, written in the Python programming language, for the generation of experiments for neuroscience and experimental psychology.			Ne	Ano (Stimulační software)
60	OpenSesame		OpenSesame	Další stimulační software, ve kterém se vytváří stimulační scénáře.	OpenSesame is free, open-source, and cross-platform software that features a comprehensive and intuitive graphical user interface and supports Python scripting for complex tasks. Additional functionality, such as support for eyetrackers, input devices, and video playback, is available through plug-ins. OpenSesame can be used in combination with existing software for creating stimulation and other kinds of experiments.			Ne	Ano (Stimulační software)
61	Presentation		Presentation	Placený stimulační software, ve kterém se vytváří stimulační scénáře.	Presentation is a stimulus delivery and experiment control program for neuroscience. It runs on any Windows PC, and delivers auditory, visual and multimodal stimuli with sub-millisecond temporal precision. Presentation is powerful enough to handle almost any behavioral, psychological or physiological experiment using fMRI, ERP, MEG, psychophysics, eye movements, single neuron recording, reaction time measures, other performance measures, and more.	<a href="http://vivoweb.org/ontology/core#Presentation">http://vivoweb.org/ontology/core#Presentation</a>	Encompasses talk, speech, lecture, slide lecture, conference presentation	Ne	Ano (Stimulační software)

ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvise s jiným termínem v BIN ontologii?
62	BrainVision Recorder		BrainVision Recorder	Nahrávací software pro EEG signál včetně markerů.	BrainVision Recorder is a multifunctional EEG recording software, created by BrainProducts, designed to provide an extremely versatile and easy-to-use platform for recording setup and execution.			Ne	Ano (Nahrávací software)
63	BrainVision PyCorder		BrainVision PyCorder	Opensource nahrávací software pro zařízení BrainVision actiCHamp amplifier. Software podporuje interakci se zesilovačem pomocí Pythonu.	PyCorder is a real open source acquisition program, based on the Python programming language, for setting up the BrainVision actiCHamp amplifier and storing data.			Ne	Ano (Nahrávací software)
64	EEG devices		EEG zařízení	Zařízení, které se zabývají měřením a nebo nahráváním EEG dat.	An EEG device is used to create a picture of the electrical activity of the brain. It has been used for both medical diagnosis and neurobiological research. In medicine, they are used to diagnose such things as seizure disorders, head injuries, and brain tumors. The data collected from this device can be interpreted by a computer and provides a geometrical picture of the brain's activity.			Ne	Ano (Přístroj)
65	Wired eeg devices		Drátová EEG zařízení	Zařízení, která potřebují připojení pomocí kabelů.	A wired eeg device is either an eeg cap or an eeg recording solution that needs to be connected by a wire in order to fulfill its function.	<a href="http://publ.obolibrary.org/obo/NCIT_C49855">http://publ.obolibrary.org/obo/NCIT_C49855</a>	A long, thin, multistranded rope or rope-like object; a multistranded conductor designed to carry signals over a distance.	Ne	Ano (EEG zařízení)
66	Wireless eeg devices		Bezdrátová EEG zařízení	Zařízení, která nevyžadují připojení pomocí kabelů.	A wireless eeg device is either an eeg cap or an eeg recording solution that is connected through wireless technologies in order to fulfill its function.			Ne	Ano (EEG zařízení)
67	EMOTIV EPOC+ 14	EMOTIV EPOC+	EMOTIV EPOC+ 14	EEG headset obsahující 14 kanálů (elektrod).	Emotiv EPOC+, a wireless, 14-channel mobile EEG system to target professional use in research and industrial applications with upgraded electronics including 9-axis inertial sensors, Bluetooth Smart and an improved power source.			Ne	Ano (Bezdrátová EEG zařízení)
68	NeuroSky	Neurosky mindwave	NeuroSky	EEG headset obsahující pouze 1 elektrodu umístěnou na čele.	NeuroSky MindWave is a wireless EEG headset, which allows to read electrical impulses of human brain. This device can be used as for scientific researches, as well as for entertainment, because it allows to control brainwave activity in real time.			Ne	Ano (Bezdrátová EEG zařízení)
69	Muse	Muse headband	Muse	EEG headset se zaměřením na měření meditace a pozornosti.	Muse is a wearable brain sensing headband. The device measures brain activity via 4 electroencephalography (EEG) sensors. An accompanying mobile app converts the EEG signal into audio feedback that is fed to the user via headphones.			Ne	Ano (Bezdrátová EEG zařízení)
70	Smarting	mBrainTrain Smarting	Smarting	Bezdrátové EEG zařízení, které umožňuje měření EEG.	Smarting offers a mobile EEG recordings with 24 channels. It allows recording of EEG and ERP experiments on both PC and Android devices.			Ne	Ano (Bezdrátová EEG zařízení)
71	BrainAmp DC	BrainVision actiCHamp	BrainAmp DC	Zesilovač EEG signálu od společnosti BrainAmp	The actiCHamp is an amplifier that integrates high-end components for electrophysiological (EEG/ERP, ExG, BCI, etc.) research in one versatile 24-bit amplifier with up to 160 channels.			Ne	Ano (Zesilovač)
72	BrainVision V-AMP 16		BrainVision V-AMP 16	Zesilovač EEG signálu od společnosti BrainAmp	Brain Products smallest amplifier, starting with 8 channels (plus 2 sensor channels) up to 16 channels (plus 2 sensor channels) V-Amp is already equipped with a 9-bit trigger port and offers ample powerful features such as sampling rates up to 20,000 Hz. The amplifier where just a few channels are needed. Suitable for animal research as well as for BCI / Biofeedback applications.			Ne	Ano (Zesilovač)
73	ActiCap	actiCAP Xpress	ActiCap	EEG čepice s aktivními elektrodami.	ActiCAP Xpress is an EEG cap dry electrode system which enables you to also manually adjust the position and height of the electrode holders.			Ne	Ano (Drátová EEG zařízení)
74	EasyCap		EasyCap	EEG čepice bez nutnosti namazání.	Easycap is a cap system with integrated flat electrode holders. In these holders the special donut shaped sintered electrodes are snapped in place. This means that after placement of the cap and electrodes, one can still reach the skin to improve the contact quality.			Ne	Ano (Bezdrátová EEG zařízení)
75	Smartphone		Smartphone	Chytrý telefon, který umožňuje využívat některé funkční vybavení běžně nacházejících se na PC.		<a href="http://publ.bioontology.org/ontology/MESH/D0000068997">http://publ.bioontology.org/ontology/MESH/D0000068997</a>	A cellular phone with advanced computing and connectivity capability built on an operating system.	Ano	Ano (Mobilní zařízení)
76	Tablet		Tablet	Mobilní zařízení, které umožňuje i využívat některé funkční vybavení běžně nacházejících se na PC	A tablet, or tablet PC, is a portable computer that uses a touchscreen as its primary input device. Most tablets are slightly smaller and weigh less than the average laptop. While some tablets include fold out keyboards, others, such as the Apple iPad and Motorola Xoom, only offer touchscreen input.	<a href="http://publ.bioontology.org/ontology/SNO MEDCT/385055001">http://publ.bioontology.org/ontology/SNO MEDCT/385055001</a>	A solid single-dose preparation obtained by compressing uniform volumes of particulate solids or by extrusion or moulding. Tablets may be single layer tablets resulting from a single compression of particles and or multilayer tablets consisting of concentric or parallel layers obtained by successive compressions of particles of different composition.	Ne	Ano (Mobilní zařízení)
77	FitBit		Fitbit	Smart náramek, který měří některé fyzikální veličiny osoby, která ho nosí. Může se jednat například o tep, počet kroků, tlak, a další.	The Fitbit is a wearable computing device. It is a fitness band, worn on a person's wrist and is designed to track physical activity. Fitbit devices are designed to track important health and activity markers, including heart rate, quality of sleep and the number of steps walked. The device is comparable to the Apple Watch and the discontinued Microsoft Band.			Ne	Ano (Nositelná elektronika)
78	Weighing machine	Scale	Váha	Přístroj pro změření hmotnosti účastníka v kg.	An instrument or machine for weighing.	<a href="http://publ.obolibrary.org/obo/MMO_0002117">http://publ.obolibrary.org/obo/MMO_0002117</a>	Determination of the mass of a sample.	Ne	Ano (Měřicí nástroj)
79	Tape measure		Metr	Měřicí nástroj, který je používán pro určení výšky účastníka experimentu v cm.				Ne	Ano (Měřicí nástroj)



ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvise s jiným termínem v BIN ontologii?
80	Handheld dynamometer		Siloměr	Přístroj, který změřil sílu stisku účastníkovy ruky v kg.	A tool used to measure the degree of force used in the contraction of the muscles of the hand, the distal part of the upper limb, consisting of the carpus, metacarpus, and fingers.	<a href="http://publ.obolibrary.org/obo/MMO_0000095">http://publ.obolibrary.org/obo/MMO_0000095</a>	A method to measure the degree of force used in the contraction of the muscles of the hand, the distal part of the upper limb, consisting of the carpus, metacarpus, and fingers.	Ne	Ano (Měřicí nástroj)
81	Spirometer		Spirometr	Přístroj sloužící k měření objemové kapacity plic účastníka.		<a href="http://publ.obolibrary.org/obo/ERO_0000452">http://publ.obolibrary.org/obo/ERO_0000452</a>	A spirometer is an apparatus for measuring the volume of air inspired and expired by the lungs. It is a precision differential pressure transducer for the measurements of respiration flow rates. The spirometer records the amount of air and the rate of air that is breathed in and out over a specified period of time.	Ano	Ano (Měřicí nástroj)
82	Isochromatic tables	Pseudoisochromatic tables	Izochromatické tabulky	Celkem 8 barevných obrázků, které slouží k testování barvocitu účastníka.	Of. relating to, using, or being a set of colored plates that include some which appear isochromatic to individuals with color-vision abnormality and that are used in the Ishihara test for color blindness.			Ne	Ano (Měřicí nástroj)
83	Sphygmomanometer		Tonometr	Přístroj k měření systolického a diastolického tlaku účastníka (mnohdy včetně možnosti měření srdečního tepu)		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C69317">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C69317</a>	An instrument used for non-invasive determination of arterial blood pressure, generally consisting of an inflatable cuff and a pressure readout device, classically, a column of mercury.	Ano	Ano (Měřicí nástroj)
84	Glucometer		Glukoměr	Přístroj k změření hladiny krevního cukru účastníka		<a href="http://publ.obolibrary.org/obo/OBI_0000795">http://publ.obolibrary.org/obo/OBI_0000795</a>	A measurement device with the function to measure and record the level/amount of glucose in a blood sample	Ano	Ano (Měřicí nástroj)
85	Balance	Balance test	Rovnováha	Sada dvou vah, která měří rovnováhu uživatele. Odhaluje, která je účastníkovou dominantní noha	Balance testing is a qualitative measure that assesses balance (physical posture) of the subject.  Human balance testing can for example highlights the preferred leg of the participant and/or how much weight does the participant put on individual legs.	<a href="http://publ.obolibrary.org/obo/OBI_0001091">http://publ.obolibrary.org/obo/OBI_0001091</a>	A balance that is used to compare the weights of two bodies, to determine the difference in mass (or weight).	Ne	Ano (Váha)
86	Stimulation device for cognitive research		Stimulátor horních končetin	Stimulátor, který měří rychlost reakce účastníka na obrazové podněty. Jsou zde měřeny zejména průměrná rychlost reakce (v ms), počet chybných/proměškaných reakcí (počet odpovědí).	A programmable hardware stimulator that was designed and developed for EEG/ERP experiments performed in the neuroinformatics laboratory at the University of West Bohemia in Pilsen. This device is mainly used to measure averaged upper limbs reaction times and to track the number of missed and incorrectly answered stimuli.			Ne	Ano (Stimulátor)
87	Lower limbs stimulator		Stimulátor dolních končetin	Stimulátor, který měří rychlost reakce účastníka na obrazové podněty. Jsou zde měřeny zejména průměrná rychlost reakce, nejrychlejší čas, nejhorší čas a směrodatná odchylka z počtu pokusů.	A software stimulator that was designed and developed for EEG/ERP experiments performed in the neuroinformatics laboratory at the University of West Bohemia in Pilsen. This device is mainly used to measure averaged lower limbs reaction times, to track the standard deviation of the sample and the best/worst time achieved during the stimulation.			Ne	Ano (Stimulátor)
88	Stimulator		Stimulátor	Přístroj, které slouží ke specifickému experimentu a je vytvořen za účelem měřit a nebo otestovat určitou experimentální hypotézu		<a href="http://publ.obolibrary.org/obo/ERO_0000863">http://publ.obolibrary.org/obo/ERO_0000863</a>	An instrument used for nerve and muscle stimulation procedures.	Ano	Ano (Přístroj)
89	Age categories		Věkové kategorie	Rozdělení účastníků do celkem 5 zájmových kategorií podle věku. Uvedené kategorie neobsahují rovnoměrné rozdělení podle věku.	Project BIN separates participants into 5 distinct age categories.  Those categories namely are: - Older school age <11 - 15> - Adolescence (15 - 18> - Full adolescence (18 - 30> - Young age (30 - 45> - Middle age (45 - 60>			Ne	Ano (Účastník)
90	Older school age		Starší školní věk	Věk člověka mezi 11 - 15 let života	A human life stage that begins at eleven years of age and continues until fifteen years of age.			Ne	Ano (Věkové kategorie)
91	Adolescence		Dospívání	Věk člověka mezi 15 - 18 let života	A human life stage that begins at fifteen years of age and continues until eighteen years of age.	<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C39298">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C39298</a>	A human life stage that begins at twelve years of age and continues until twenty-one complete years of age, generally marked by the beginning of puberty and lasting to the beginning of adulthood.	Ne	Ano (Věkové kategorie)
92	Full adolescence		Plná dospělost	Věk člověka mezi 18 - 30 let života	A human life stage that begins at eighteen years of age and continues until thirty years of age.			Ne	Ano (Věkové kategorie)
93	Young age		Mladý věk	Věk člověka mezi 30 - 45 let života	A human life stage that begins at thirty years of age and continues until forty-five years of age.			Ne	Ano (Věkové kategorie)
94	Middle age		Střední věk	Věk člověka mezi 45 - 60 let života	A human life stage that begins at forty-five years of age and continues until sixty years of age.			Ne	Ano (Věkové kategorie)
95	Mobile application		Mobilní aplikace	Programové vybavení smartphonu, které rozšiřuje jeho funkčnost určitým směrem.	A mobile application, most commonly referred to as an app, is a type of application software designed to run on a mobile device, such as a smartphone or tablet computer. Mobile applications frequently serve to provide users with similar services to those accessed on PCs. Apps are generally small, individual software units with limited function.			Ne	Ano (Smartphone)
96	Mobile application for data acquisition		Mobilní aplikace pro sběr dat	Mobilní aplikace, která pomáhá při sběru dat v rámci systému BIN.	The objective of mobile application for data acquisition is to reliably and accurately acquire and record heterogenous data from a variety of external devices.			Ne	Ano (Mobilní aplikace)

ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvise s jiným termínem v BIN ontologii?
97	Mobile application for fitness tracking		Mobilní fitness aplikace	Mobilní aplikace pro sledování denních aktivit člověka. Například sledování jídelníčku a cviků účastníka.	A specific mobile application designed for monitoring of daily activities of the user. Those fitness applications usually enable the user to track and/or balance food composition (for nutritionally balanced meals) and physical activity.			Ne	Ano (Mobilní aplikace)
98	Soundproof cabin		Zvukovotěsná komora	Uzavřená zvukově ztlučený prostor, ve které se odehrávají stimulační experimenty.	An enclosed cabin specifically designed to insulate so as to obstruct the passage of sound and possibly also hinder any unwanted visual stimuli.			Ne	Ano (Laboratorní prostředí)
99	LED panels	LED panel	LED panely	Světelné panely, které jsou používány pro obrazovou stimulaci účastníků.	LEDs are flat panel displays that emit visible light when an electric current passes through the light emitting diodes.			Ne	Ano (Příslušenství)
100	Average reaction time of upper limbs		Průměrná reakční doba horních končetin	Průměrná reakční doba dosažená jako aritmetický průměr ze všech reakcí rukou na optické podněty.	An arithmetic mean of measured reaction times (in ms) of all the correctly answered stimuli from the upper limbs stimulation device.			Ne	Ano (Stimulátor horních končetin)
101	Number of missed stimuli		Zameškané stimuly horních končetin	Počet stimulů, na které nedokázal účastník odpovědět do 2s od stimulace.	Number of stimuli that were not given an answer in a predefined time limit (usually up to 2 seconds).			Ne	Ano (Stimulátor horních končetin)
102	Number of incorrectly answered stimuli		Chybné odpovědi na stimuly horních končetin	Počet stimulů, na které byla vybrána špatná odpověď.	Number of stimuli that were not given a right answer.			Ne	Ano (Stimulátor horních končetin)
103	Average reaction time of lower limbs		Průměrná reakční doba dolních končetin	Průměrná reakční doba dosažená jako aritmetický průměr ze všech reakcí nohou na optické podněty.	An arithmetic mean of measured reaction times (in ms) of all the correctly answered stimuli from the lower limbs stimulation device.			Ne	Ano (Stimulátor dolních končetin)
104	Standard deviation for the reaction time of lower limbs	Standard deviation	Směrodatná odchylka průměrné reakční doby dolních končetin	Směrodatná odchylka dosažená během odpovědi pomocí dolních končetin.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C53322">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C53322</a>	A measure of the range of values in a set of numbers. Standard deviation is a statistic used as a measure of the dispersion or variation in a distribution, equal to the square root of the arithmetic mean of the squares of the deviations from the arithmetic mean.	Asi Ano	Ano (Stimulátor dolních končetin)
105	Fastest response time		Nejrychlejší doba odpovědi dolních končetin	Nejrychlejší doba reakce dolních končetin v rámci stimulace.	The fastest possible response time (usually in milliseconds) during which the participant was able to correctly answer to the visual stimuli.			Ne	Ano (Stimulátor dolních končetin)
106	Slowest response time		Nejhorší doba odpovědi dolních končetin	Nejpomalejší doba reakce dolních končetin v rámci stimulace.	The slowest possible response time (usually in milliseconds and usually caps at two seconds) during which the participant was able to correctly answer to the visual stimuli.			Ne	Ano (Stimulátor dolních končetin)
107	Color blindness test results		Výsledek barvocitu	Počet nesprávně zodpovězených izochromatických tabulek	A collection of problematic plates that were NOT seen by the participant during the color blindness test while using the pseudoisochromatic tables.			Ne	Ano (Izochromatické tabulky)
108	Systolic blood pressure		Systolický tlak	Hodnota systolického (horního) tlaku v lidském těle.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25298">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25298</a>	The maximum pressure exerted into the systemic arterial circulation during the contraction of the left ventricle of the heart.	Ano	Ano (Tlakoměr)
109	Diastolic blood pressure		Diastolický tlak	Hodnota diastolického (dolního) tlaku v lidském těle		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25299">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25299</a>	The minimum pressure exerted into the systemic arterial circulation during cardiac ventricular relaxation and filling.	Ano	Ano (Tlakoměr)
110	Heart rate		Tep	Počet tepů srdce člověka za minutu.		<a href="http://purl.bioontology.org/ontology/MESH/H0006339">http://purl.bioontology.org/ontology/MESH/H0006339</a>	Modulation of HEART RATE. The number of times the HEART VENTRICLES contract per unit of time, usually per minute.	Ano	Ano (Tlakoměr)
111	Blood sugar level		Hladina krevního cukru	Množství krevního cukru v krvi účastníka. Měřeno pomocí nanesení kapky krve na měřicí proužek.		<a href="http://purl.bioontology.org/ontology/CRISP/051-4526">http://purl.bioontology.org/ontology/CRISP/051-4526</a>	levels of glucose found in the blood.	Ano	Ano (Glukoměr)
112	Food before blood sugar measurement		Jídlo před měřením	Komentář jak dlouho před měřením krevního cukru bylo konzumováno jídlo. Případně i s dodatkem o jaké jídlo se jednalo.	Describes how long and what kind of food was consumed before the blood sugar measurement.			Ne	Ano (Glukoměr)
113	Weight		Hmotnost	Změřená hmotnost účastníka v kilogramech.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25208">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25208</a>	The vertical force exerted by a mass as a result of gravity.	Ano	Ano (Váha)
114	Height		Výška	Změřená výška účastníka v centimetrech.		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25347">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C25347</a>	The vertical measurement or distance from the base to the top of an object; the vertical dimension of extension.	Ano	Ano (Metr)
115	Body mass index		BMI	Body mass index vypočítaný měřicí vahou.		<a href="http://purl.bioontology.org/ontology/C16358">http://purl.bioontology.org/ontology/C16358</a>	A general indicator of the body fat an individual is carrying based upon the ratio of weight to height.	Ano	Ano (Fitness váha)
116	Percentage of muscle mass in the body	Percentage of muscle mass in the body	Procentuální zastoupení svalové hmoty v těle	Procentuální zastoupení svalové hmoty v těle účastníka, které jsou vypočítané fitness vahou.	The amount of muscle mass inside the body that is calculated using the bioelectrical impedance analysis.			Ne	Ano (Fitness váha)
117	Percentage of water in the body	Percentage of water in the body	Procentuální zastoupení vody v těle	Procentuální zastoupení vody v těle účastníka, které jsou vypočítané fitness vahou.	The amount of water inside the body that is calculated using the bioelectrical impedance analysis.			Ne	Ano (Fitness váha)
118	Percentage of fat in the body	Percentage of fat in the body	Procentuální zastoupení tuku v těle	Procentuální zastoupení tuku v těle účastníka, které jsou vypočítané fitness vahou.	The amount of fat inside the body that is calculated using the bioelectrical impedance analysis.			Ne	Ano (Fitness váha)
119	Flexibility		Flexibilita	Měření flexibility probíhá postavením účastníka na měřicí stupínky do co nejnižšího předklonu. Je měřena vzdálenost mezi konečky prstů a stupínkem.		<a href="http://purl.obolibrary.org/obo/PATO_0001543">http://purl.obolibrary.org/obo/PATO_0001543</a>	A physical quality inhering in a bearer by virtue of the bearer's disposition to being turned, bowed, or twisted without breaking.	Asi Ano	Ano (Metr)
120	Balance of the right leg	Weight put on the right leg	Rovnováha - váha na pravé noze	Váha kterou při vzpřímeném postavení klade účastník na pravou nohu.	Weight that is put on the right leg during the participants standing still balance test.			Ne	Ano (Váha)
121	Balance of the left leg	Weight put on the left leg	Rovnováha - váha na levé noze	Váha kterou při vzpřímeném postavení klade účastník na levou nohu.	Weight that is put on the left leg during the participants standing still balance test.			Ne	Ano (Váha)

ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvise s jiným termínem v BIN ontologii?
122	Strength of the right arm	Right arm strength	Síla pravá ruka	Síla, kterou účastník vyvine při úchopu siloměru pouze pomocí pravé ruky.	Maximum amount of strength developed by pressing the dynamometer using only the right arm.			Ne	Ano (Předpažení, Volný styl)
123	Strength of the left arm	Left arm strength	Síla levá ruka	Síla, kterou účastník vyvine při úchopu siloměru pouze pomocí levé ruky.	Maximum amount of strength developed by pressing the dynamometer using only the left arm.			Ne	Ano (Předpažení, Volný styl)
124	Vital lung capacity (VC)		Vitální kapacita plic	Maximální objem vzduchu, který je schopen účastník vydechnout z plic na jeden nádech.		<a href="http://publ.bioontology.org/ontology/MESH/D014797">http://publ.bioontology.org/ontology/MESH/D014797</a>	The volume of air that is exhaled by a maximal expiration following a maximal inspiration.	Ano	Ano (Zátěžový stav, Klidový stav)
125	Forced expiratory volume in the first second of exhalation (FEV1)		Vydechnutý objem za první sekundu výdechu	Objem vzduchu, který je vydechnutý při usilovném výdechu za první sekundu;		<a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C38084">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C38084</a>	A test of lung function, the FEV1 is the volume exhaled during the first second of a forced expiratory maneuver started from the level of total lung capacity. It is the most frequently used index for assessing bronchoconstriction or bronchodilatation.	Ano	Ano (Zátěžový stav, Klidový stav)
126	Peak expiratory flow (PEF)		Vrcholový výdechový průtok	Nejvyšší rychlost proudění vzduchu na vrcholu usilovného výdechu.		<a href="http://www.ebi.ac.uk/efo/EFO_00097_18">http://www.ebi.ac.uk/efo/EFO_00097_18</a>	The maximum rate of gas flow, beginning at the point of peak inspiratory capacity, that can be sustained during forced exhalation for a defined period of time. This test is most often used in combination with other studies of lung function to diagnose and monitor obstructive and restrictive lung disease(s).	Ano	Ano (Zátěžový stav, Klidový stav)
127	Fitness weighing machine	Fitness scale	Fitness váha	Váha, která je schopná změřit i další veličiny a ne pouze hmotnost objektu. V tomto případě váha ještě měří BMI, % zastoupení svalové hmoty, % zastoupení vody v těle a % zastoupení tuku v těle	A specialised fitness scale that not only measures weight but also other parameters like BMI, percentages of fat, water and muscle mass in the body through the use of bioelectrical impedance analysis.			Ne	Ano (Váha)
128	Fitness step	Aerobic fitness step	Stupínek	Posilovací náčiní pro simulaci chůze nahoru a dolu na schody. Používá se jako simulátor zátěžového stavu v rámci spirometrie.	A tool that is used during the step aerobics as a raised platform to be stepped on and off. Inside the BIN project the fitness step is used to simulate physical stress on the participants body.			Ne	Ano (Příslušenství)
129	Resting state		Klidový stav	Stav účastníka, ve kterém nevykonává žádnou fyzicky náročnou činnost nejčastěji v sedě	A state in which the participant is refrained from exhibiting any physical exertion during the measurement.			Ne	Ano (Spirometr)
130	Stress state		Zátěžový stav	Stav účastníka, ve kterém vykonává fyzicky náročnou činnost po určitou dobu.	A state in which the participant is exhibiting some physical stress during the measurement usually in a form of an exercise.			Ne	Ano (Spirometr)
131	Freestyle	Dynamometer freestyle measurement	Volný styl	Libovolný způsob uchopení siloměru pomocí jedné ruky za účelem vyvinutí maximálně silného sevření.	The participant is free to choose which way he wants to grab the dynamometer for the measurement. The only restriction that concerns the participant is that he has to use only one hand during the strength measurement.			Ne	Ano (Siloměr)
132	Front shoulder raise	Dynamometer measurement during front shoulder raise	Předpažení rukou	Měření síly stisku účastníka při předpažení rukou.	The participant has to hold the dynamometer using the front shoulder raise position while the measurement takes place.			Ne	Ano (Siloměr)
133	Electrode location file		Soubor s pozicemi elektrod	Soubor, který indikuje umístění a označení všech elektrod EEG čepice na skalpu lidské hlavy.	Electrode location file contains every electrode that is used in a specific type of an EEG cap. This file is usually created by the EEG caps manufacturer and contains the x, y, z coordinates for each electrode.			Ne	Ano (Nahrávací software)
134	Garmin Forerunner 210	Forerunner 210	Garmin Forerunner 210	Smart naramek podobný fitbitu, sleduje počet kroku a spálené kalorie.	Forerunner 210 is available with or without our premium soft strap heart rate monitor to display your heart rate in beats per minute. It has 5 heart rate zones, which you can customize based on your exact zones. You can also set up heart rate alerts to notify you when you are above or below your targets. Forerunner 210 provides heart rate-based calorie computations so you can more accurately track your calories burned.			Ne	Ano (Nositelná elektronika)
135	Mobile Devices		Mobilní zařízení	Přenosný technický prostředek s výpočetními schopnostmi.		<a href="http://ontology.apa.org/apaonto/terms/onlyOUT%20(5).owl#Mobile_Device">http://ontology.apa.org/apaonto/terms/onlyOUT%20(5).owl#Mobile_Device</a>	Portable device with computing abilities.	Ano	Ano (Kategorie přístrojů)
136	Remaining fitness step distance		Vydálenost od stupínku	Pokud účastník dosáhl dále než ke vztážené rovině (rovina stupínku), tak se jedná o kladný počet centimetrů. Pokud účastník nedosáhl dále než ke vztážené rovině (rovina stupínku), tak jsou uvedeny záporné centimetry.	The remaining distance that is being measured between participants fingertips and the top of the aerobic fitness step. The remaining distance that is being measured between participants fingertips and the top of the aerobic fitness step.  Flexibility is measured using the aerobic fitness step. The participant gets on top of the fitness step and slowly leans forward. Attention is paid only to the distance between the participants fingertips and the top of the fitness step.  This distance is usually measured in centimeters.  If the participant managed to touch under the top of the fitness step, then this distance would be positive (signaling that the participant is flexible), in the other case the distance is a negative number stating how many centimeters were left between the participants fingertips and the top of the fitness step.			Ne	Ano (Metr)
137	Measurement date		Datum měření	Datum, kdy proběhlo měření.	A specified date, preferably in the format of YYYY-MM-DD, that the measurement was taken on.			Ne	Ano (Měření)

ID	Term (EN)	Updated term (EN)	Termin	Moje definice (CZ)	Moje definice (EN)	Alternativní zdroj	Definice alt. zdroje	Jsou si definice podobné?	Souvisí s jiným termínem v BIN ontologii?
138	Measurement place		Místo měření	Místo, na kterém proběhlo měření.	Defined address of a location usually consisting of either the number of the house, name of the road, and name of the town where the measurement took place, or GPS coordinates.			Ne	Ano (Měření)
	Legenda		Původní (abstraktní) terminologie BIN			Ne	Definice neexistují a nebo nejsou použity ve stejném významu	-- tzn. potřeba vytvořit novou definici	
			Dodatečné (související) termíny			Asi Ano	Definice jsou si v určitém smyslu podobné. Potřeba posoudit jestli je definice vhodná v původním znění.	-- a nebo bude potřeba vytvořit vlastní definici	
						Ano	Definice si jsou podobné. Převzáni původní definice terminu.	-- vlastní i cizí definice se schodují	

