

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra geomatiky

Diplomová práce
Metodika pro doplňování bodů
do databáze Smart Points of Interest

ZÁPADOČESKÁ UNIVERZITA V PLZNI
Fakulta aplikovaných věd
Akademický rok: 2017/2018

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Lucie KARASOVÁ**
Osobní číslo: **A16N0157P**
Studijní program: **N3602 Geomatika**
Studijní obor: **Geomatika**
Název tématu: **Metodika pro doplňování bodů do databáze Smart Points of Interest**
Zadávací katedra: **Katedra geomatiky**

Z á s a d y p r o v y p r a c o v á n í :

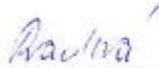
1. Popis vhodných zdrojů bodů pro Smart Points of Interest v České republice.
2. Postup sběru informací (kombinace sběru dat v terénu a existujících datových zdrojů).
3. Návrh efektivní a jednoduché evidence dat pro běžné uživatele.
4. Sběr dat minimálně ve dvou lokalitách (podklad pro návrh metodiky, ověření metodiky).

Rozsah grafických prací: **dle potřeby**
Rozsah kvalifikační práce: **cca 45 stran**
Forma zpracování diplomové práce: **tištěná/elektronická**
Seznam odborné literatury:

- Čerba, O., Charvát, K., Mildorf, T., Berzinš, R., Vlach, P., & Musilová, B. (2016). SDI4Apps Points of Interest Knowledge Base. In Progress in Cartography (pp. 229-237). Springer International Publishing.
- Čerba, O., & Mildorf, T. Smart Points of Interest: Big, Linked and Harmonized Spatial Data.
- Elwood, S. (2008). Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. GeoJournal, 72(3-4), 133-135.
- Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In Crowdsourcing geographic knowledge (pp. 105-122). Springer Netherlands.

Vedoucí diplomové práce: **Ing. Mgr. Otakar Čerba, Ph.D.**
Katedra geomatiky

Datum zadání diplomové práce: **2. října 2017**
Termín odevzdání diplomové práce: **25. května 2018**


Doc. Dr. Ing. Vlasta Radová
děkanka




Doc. Ing. Václav Čada, CSc.
vedoucí katedry

V Plzni dne 2. října 2017

Prohlášení

Prohlašuji, že jsem diplomovou práci na téma

„Metodika pro doplňování bodů do databáze Smart Points of Interest“

vypracovala samostatně pod odborným vedením vedoucího práce s použitím zdrojů a literatury v práci uvedených.

V Plzni dne 25. 5. 2018

.....
Bc. Lucie Karasová

Poděkování

Mé poděkování patří Ing. et Mgr. Otakarovi Čerbovi, Ph.D. za odborné vedení diplomové práce a také za cenné informace, věcné připomínky, čas a trpělivost, které mi v průběhu zpracování věnoval.

Anotace

Diplomová práce se zabývá databází Smart Points of Interest (SPOI), doplňováním bodů zájmu Point of Interest (POI) do databáze SPOI a návrhem metodiky pro toto doplňování. Návrhu metodiky předcházela rešerše dostupných materiálů zabývajících se databází SPOI a také věcmi souvisejícími jako například Linked data, Open data, standardy a datové sady, atributy a objektové vazby vhodné pro databázi SPOI, datový model, harmonizace dat, jazyk SPARQL a další. Návrh metodiky se zabývá doplňováním bodů do databáze SPOI od jejich ověřování, sběru a získávání, přes evidenci dat až po jejich kontrolu a aktualizaci. Cílem práce tedy bylo navrhnout metodiku pro doplňování bodů do databáze SPOI, která by byla použitelná i pro širokou veřejnost. Její hodnocení je uvedeno v poslední kapitole této práce.

Klíčová slova

Smart Points of Interest, databáze, atribut, objektová vazba, metodika, Point of Interest, doplňování bodu, ověřování, sběr, získávání, evidence dat, kontrola, aktualizace

Abstract

The diploma thesis deals with the Smart Points of Interest database (SPOI), adding new points named Points of Interest (POIs) to the SPOI database and the proposal of the methodology for adding of these points to the SPOI database. The proposal of the methodology was preceded by a literature research of available materials dealing with the SPOI database and related issues such as Linked data, Open data, standards and data sets, annotations and object relations suitable for the SPOI database, data model, data harmonization, SPARQL Query Language and so on. The proposal of the methodology deals with adding new points to the SPOI database. It starts from points verification, collection and acquisition, through record of data, to their control and update. The aim of this thesis is to propose the methodology for adding new points to the SPOI database which would also be usable for the general public. Its evaluation is stated in the last chapter of this thesis.

Keywords

Smart Points of Interest, database, annotation, object relation, methodology, Point of Interest, point adding, verification, collection, acquisition, record of data, control, update

Obsah

Obsah	7
Seznam obrázků	8
Seznam tabulek	8
Seznam použitých zkratk	9
Úvod.....	11
1. Smart Points of Interest	13
1.1. Databáze SPOI	13
1.2. Doplnění databáze SPOI	15
1.2.1. Druhy sbíraných dat – popis vhodných bodů	15
1.2.2. Zdroje získávaných dat	17
1.2.3. Druhy vazeb mezi sbíranými daty	26
1.2.4. Datový model.....	32
1.2.5. Harmonizační procesy	39
2. Metodika.....	42
2.1. Způsob získávání dat – postup sběru informací	44
2.1.1. Ověřování.....	44
2.1.2. Sběr a získávání	52
2.2. Evidence dat	61
2.3. Možnosti kontroly a aktualizace	69
2.3.1. Kontrola	69
2.3.2. Aktualizace	72
3. Ověřování metodiky	74
3.1. Způsob ověřování metodiky.....	74
3.2. Výsledky dotazníku.....	75
3.2.1. Silné a slabé stránky metodiky	77
3.3. Možnosti vylepšení	79
Závěr	80
Seznam použité literatury	82
Obsah příloženého CD.....	88

Seznam obrázků

- Obr. 1: Datový model SPOI k 3. 4. 2017 (graficky upravený)
- Obr. 2: Rozšířený datový model SPOI dle Ing. Martina Kiráľa
- Obr. 3: Diagram postupu doplňování databáze SPOI
- Obr. 4: Snímek mapového okna s vybraným bodem zájmu (Lokál Pod Divadlem)
- Obr. 5: Snímek hlavní stránky Smart tourism guide
- Obr. 6: Snímek Smart tourism guide po vyhledání POI bodů v městě Plzni
- Obr. 7: Snímek dotazovacího okna Virtuoso SPARQL Query Editor
- Obr. 8: Snímek výstupu dotazu ve Virtuoso SPARQL Query Editor – tabulka
- Obr. 9: Snímek z konvertoru Convertcsv.com (CSV – GeoJSON)
- Obr. 10: Vytvoření nové vrstvy ze souboru CSV v programu QGIS
- Obr. 11: Konverze souboru CSV v programu CSV Converter
- Obr. 12: ETL proces
- Obr. 13: Snímek vizualizovaných dat v editoru Mapshaper
- Obr. 14: Snímek informací zobrazených v editoru Mapshaper
- Obr. 15: Snímek prázdného listu v aplikaci Microsoft Excel Online
- Obr. 16: Snímek prázdného listu v aplikaci Tabulky Google
- Obr. 17: Snímek prázdného listu v aplikaci WPS Office – Spreadsheets
- Obr. 18: Publikování vrstev na serveru GeoServer.org
- Obr. 19: Stažená data z webové služby WFS z GeoServer.org v programu QGIS
- Obr. 20: Hlavička tabulky pro evidenci POI bodů – povinné atributy
- Obr. 21: Tabulka s měněnými hodnotami

Seznam tabulek

- Tab. 1: Přehled příkladů lokálních zdrojů v České Republice se základními informacemi

Seznam použitých zkratek

CIP – Competitiveness and Innovation Framework Programme
CSV – Comma-Separated Values
ČÚZK – Český úřad zeměměřický a katastrální
DCMI – Dublin Core Metadata Initiative
EÚS – Evropská územní spolupráce
FOAF – Friend of a Friend
GDB – Garmin GPS Database
GeoJSON – a geospatial data interchange format based on JSON
GeoSPARQL – a geographic query language for RDF data
GIS – Geographic Information System
GML – Geography Markup Language
GPS – Global Positioning System
GPX – GPS Exchange Format
HTML – HyperText Markup Language
IATA – International Air Transport Association
ICT-PSP – Information and Communication Technologies Policy Support Programme
IISPP – Integrovaný informační systém památkové péče
ISA – Interoperability Solutions for European Public Administrations
ISAD – Informační systém o archeologických datech
ISKN – Informační systém Katastru nemovitostí
ISÚI – Informační systém územní identifikace
JSON – JavaScript Object Notation
KML – Keyhole Markup Language
KN – Katastr nemovitostí
MIS – Metainformační systém
NPÚ – Národní památkový úřad
ODbL – Open Database License
OGC – Open Geospatial Consortium
OSD – Open Software Description file (Microsoft Corporation)
OSM – Open Street Map
OWL – Web Ontology Language
PDF – Portable Document Format

POI – Point of Interest
PSČ – Poštovní směrovací číslo
RDF – Resource Description Framework
RÚIAN – Registr územní identifikace, adres a nemovitostí
SHP – ESRI Shapefile – spatial data format
SKOS – Simple Knowledge Organization System
SPARQL – SPARQL Protocol And RDF Query Language
SPOI – Smart Points of Interest
SQL – Structured Query Language
URI – Uniform Resource Identifier
URL – Uniform Resource Locator
UTF – Unicode Transformation Format
VDP – Veřejný dálkový přístup
VFR – Výměnný formát RÚIAN
W3C – World Wide Web Consortium
WFS – Web Feature Service
WGS 84 – World Geodetic System 1984
XLS – Excel Microsoft Spreadsheet
XLSX – Excel Microsoft Office Open XML Format Spreadsheet file
XML – Extensible Markup Language

Úvod

Databáze Smart Points of Interest (dále SPOI) je datovou sadou obsahující body zájmu, které se nazývají Points of Interest (POI). Tyto body jsou zajímavé převážně v oblasti turismu, kde je propojena moderní technologie a chytrá zařízení s turistikou, cyklistikou, dopravou a dalšími oblastmi zájmu skrze vizualizaci bodů v digitální mapě – mapový průvodce, v budoucnu multimediální uživatelské zařízení atd. Především díky možnosti využití zájmových bodů pro turismus má také aktivita spojená s databází SPOI souvislost s přírodou a tím i odpočinkem a relaxací. Jelikož přispěvatelem do databáze SPOI může být také široká veřejnost, bylo by vhodné jí toto téma přiblížit a zpřístupnit. Z těchto důvodů, a také protože jsem během svého studia spolupracovala na projektu a evidovala body pro doplňování databáze SPOI, zvolila jsem si jako zaměření své diplomové práce téma „Metodika pro doplňování bodů do databáze Smart Points of Interest“.

Práce se bude zabývat teorií okolo databáze Smart Points of Interest, jejím fungování, charakteristikou dílčích částí a základním popisem souvisejících věcí. Bude zde popsána samostatná databáze a její doplňování body – jaké jsou druhy sbíraných dat a jejich zdroje i s uvedením příkladů, jak jsou body propojeny, jak vypadá datový model a jednotlivé vazby a atributy, taktéž je nahlédnuto do procesů harmonizace dat. Na teorii navazuje praktická část, kde bude navržen postup pro sběr a získání bodů do databáze SPOI a konkrétní možnosti, jak doplňování bodů provádět – ověřování, sběr a získávání, evidence, kontrola a aktualizace dat. Metodika je určena především uživatelům z řad široké veřejnosti nebo uživatelům z institucí či úřadů, kteří mají na doplňování bodů vlastní zájem, tj. pracovníci informačních center, městských úřadů nebo poskytovatelé služeb.

Cílem práce je seznámit uživatele s vlastnostmi a celkovou problematikou doplňování bodů do databáze SPOI a poté navrhnout metodiku, jakožto návod, jak postupovat při obohacování databáze SPOI od ověření bodů a rozhodnutí, které z nich sbírat či získávat, až po poskytnutí dat k vložení do databáze a jejich aktualizaci. Výše zmíněného cíle návrhu metodiky bude dosaženo pomocí přesného popisu aktivit a jednotlivých kroků „sběrače“. Tento přístup především sjednocuje způsob, jakým by se měli v budoucnu uživatelé při rozšiřování databáze SPOI řídit.

V závěrečné kapitole práce bude uvedeno hodnocení navržené metodiky a posouzení, zda byl splněn její cíl. Metodika a také celková práce bude hodnocena na základě zpětné vazby

získané od dotazovaných respondentů, čímž se ověří její použitelnost. Otázky zodpovídané respondenty budou uvedeny taktéž v závěrečné kapitole této práce.

1. Smart Points of Interest

1.1. Databáze SPOI

Databáze Smart Points of Interest (SPOI) je databáze používaná v projektu Peregrinus Silva Bohemica¹ (dále jen Peregrinus), který celý výzkum SPOI zajišťuje. Projekt Peregrinus se zabývá tvorbou multimediálního a digitálního turistického průvodce pro přeshraniční historické cesty v Bavorském lese a na Šumavě. Pomocí propagace, mapování a dokumentace přírodních zdrojů a historických památek má za cíl zvýšit atraktivitu příhraničního regionu a tím zvýšit povědomí o jeho kulturním a přírodním dědictví. Dalším plánem projektu Peregrinus je pak zaměřit pozornost rozhodujících orgánů či potenciálních financujících subjektů k této problematice. Oslovování turistů se děje pomocí moderních prezentačních prostředků (mapový průvodce, 3D modely...) a v budoucnu multimediálního uživatelského prostředí (web, chytré telefony), které prezentují kulturní a přírodní bohatství.¹ Právě k prezentaci kulturního a přírodního bohatství se využívá databáze Smart Points of Interest (SPOI), kde jsou záznamy (body) reprezentující danou památku, zajímavost aj. jednotlivě uloženy.

Databázi bodů SPOI vytvořil v rámci pilotní aplikace Open Smart Tourist Data² tým projektu SDI4Apps³. Tato databáze je bezešvá, což znamená, že není rozdělena na samostatné části, ale je propojená vztahy (vazbami) mezi jednotlivými prvky (body) a také s jinými zdroji (databázemi) – Linked Data (více dále), otevřená – Open Data (viz dále) a je vytvářena kombinací volně dostupných globálních a lokálních dat. Projekt SDI4Apps byl spravován Západočeskou univerzitou v Plzni⁴ a financován prostředky z Evropské unie, které byly získány z Programu na podporu politiky informačních a komunikačních technologií (ICT-PSP)⁵ jako součásti Rámcového programu pro konkurenceschopnost a inovace (CIP)⁶.

Databáze Smart Points of Interest obsahuje (jsou zde shromažďovány) jednotlivé body zájmu POI (Point of Interest), což jsou konkrétní místa (body), která mohou být pro někoho zajímavá nebo užitečná (turistika, cestovní ruch a propagace regionu či dopravní navigace). Pro vizualizaci v mapě lze takovou datovou sadu používat spolu s podkladovou mapovou

¹ <https://kgm.zcu.cz/aktualni-projekty/peregrinus/>

² <http://sdi4apps.eu/project-information/pilot-applications/pilot-2-open-smart-tourist-data/>

³ <http://sdi4apps.eu/project-information/>

⁴ <http://www.zcu.cz/>

⁵ <http://ec.europa.eu/cip/ict-isp/>

⁶ <http://ec.europa.eu/cip/>

vrstvou. Mapový klient⁷, který propojuje tyto POI body s mapovým podkladem, je vytvořený na podkladě knihovny HS Layers⁸ a poskytuje klasické funkce, kterými jsou zoom, posun, přepínání vrstev včetně podkladové mapy nebo vyhledávání ve vrstvách a změna jejich pohledu. Dále lze přidávat a editovat body nebo také mapu sdílet jako statický obrázek či v dynamické podobě.

Datová sada SPOI se začala vyvíjet v rámci projektu Peregrinus pro rozvoj a popularizaci míst nejen pro turisty, ale také pro zvýšení pozornosti orgánů v území, jak bylo psáno výše, a taktéž jako potřeba zmapovat a doplnit body zájmu. Jinak řečeno obohatit a zahustit síť již vedených bodů dalšími zajímavými či zájmovými body a místy. Na území České republiky se jednalo zprvu konkrétně o Šumavu a poté celkově Plzeňský kraj, avšak sběr bodů probíhá prakticky po celém světě, přičemž nejvíce bodů SPOI je ve vyspělých zemích jako jsou například USA a Velká Británie. (Čerba a Mildorf 2016) Ideálním stavem datové sady SPOI je zmapování veškerých zájmových bodů na celé Zemi a poté „pouhé“ udržování a aktualizace informací a doplňování nově vzniklých bodů a míst.

Databáze obsahuje kolem 28 000 000 bodů zájmu (k 1. 5. 2018)⁹ publikovaných jako 5* Linked Open Data^{10 11}. Open Data¹² znamená, že datová sada SPOI je otevřená a data jsou tak volně dostupná pod určitou licenci, lze je editovat a vkládat do sady nové záznamy (body). Licence, která toto splňuje a je užívána databází SPOI, je Open Database License (ODbL)¹³, jež umožňuje uživatelům také data svobodně sdílet a používat databázi za podmínky poskytnutí stejné svobody ostatním uživatelům.¹⁴ Linked Data¹⁵ jsou propojená data, která odkazují na externí zdroje dat a umožňují, aby na ně odkazovala další data a informace. Uživatelé tedy mohou tyto body dále využívat – stahovat, vyhledávat nebo je používat pro vlastní aplikace. Formát, který je v souladu s požadavky¹⁶ na tzv. pětihvězdičková data, je RDF (Cyganiak et al. 2014). Data SPOI jsou uložena jako grafová struktura ve formě RDF trojic (více v kapitole 1.2.3.). (Čerba et al. 2017a)

⁷ <http://ng.hsayers.org/examples/geosparql/>

⁸ <http://ng.hsayers.org/>

⁹ V nejbližší době bude zveřejněno více bodů.

¹⁰ <https://www.w3.org/DesignIssues/LinkedData.html>

¹¹ <https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html#linked-open-data>

¹² <http://opendatahandbook.org/guide/en/what-is-open-data/>

¹³ <https://opendatacommons.org/licenses/odbl>

¹⁴ https://cs.wikipedia.org/wiki/Open_Database_License

¹⁵ <https://www.w3.org/standards/semanticweb/data>

¹⁶ Požadavek na čtveřhvězdičková data je identifikace datových položek pomocí URI a zápis dat v univerzálním RDF formátu.

1.2. Doplnování databáze SPOI

1.2.1. Druhy sbíraných dat – popis vhodných bodů

Prvotními sbíranými daty jsou body užitečné a vhodné pro cestovní ruch a související oblasti jako například doprava, logistika, reklama atp. Body vhodné pro turistiku nebo cyklistiku patří vzhledem k projektu Peregrinus mezi hlavní a to z důvodu, že se zabývá multimediálním a digitálním turistickým průvodcem pro přeshraniční historické cesty Bavorském lese a na Šumavě. Mezi sbíraná data může patřit prakticky jakýkoliv bod či místo zájmu. Každý člověk považuje za zajímavé něco jiného, má jiné preference, a proto by se do databáze SPOI měly zařadit body všechny, byť mohou být pro někoho málo zajímavé nebo zcela bezvýznamné.

Současná klasifikace bodů SPOI je rozdělena do jednotlivých kategorií (tříd) a ty jsou uvedeny ve SPOI ontologii¹⁷. SPOI ontologie obsahuje 290 tříd a jejich základní struktura byla převzata z klasifikačního systému Waze¹⁸ používaného pro data v této aplikaci, ostatní třídy byly odvozeny z OpenStreetMap (OSM)¹⁹ Map Features²⁰ nebo vytvořeny na zakázku. Příkladem kategorií dle SPOI ontologie jsou kultura a volný čas, jídlo a pití, ubytování, krajnotvorné prvky, volný čas venku, nákupy a služby, doprava, veřejné služby atd. (Čerba 2017)

Body lze dále dělit dle nejrůznějších kritérií. Jedním z mnoha způsobů je dělení bodů a míst do kategorií podle toho, kdo je hledá, koho zajímají a pro koho mají význam. Tento způsob je užitečný pro laickou veřejnost z důvodu lepší orientace v tom, které body sbírat a doplňovat do databáze SPOI. POI body lze rozdělit na body zajímavé pro turisty (kategorie Turistické body) a body praktické pro stávající (místní) obyvatele území (kategorie Body vyhledávané místními obyvateli). Je samozřejmé, že obě skupiny se mohou překrývat a že turista může hledat nemocnici a lékárnu nebo obyvatel města bankomat a restauraci, i když se předpokládá, že takové věci bude hledat spíše právě turista, který to v dané lokalitě příliš nezná. Proto je vhodné přidat další skupinu druhů sbíraných bodů, kterou jsou body užitečné pro všechny (kategorie Všeobecně zájmové body).

¹⁷ <http://gis.zcu.cz/spoi/doc/ontology.owl>

¹⁸ <https://wazeopedia.waze.com/wiki/Czech/M%C3%ADsta>

¹⁹ <https://www.openstreetmap.org/>

²⁰ https://wiki.openstreetmap.org/wiki/Map_Features

Zde je uvedeno rozdělení druhů bodů do kategorií i s konkrétními příklady, avšak jedná se o individuální záležitost – pro každého je důležité a zajímavé něco jiného, každý uživatel by mohl jednotlivé příklady zařadit do jiné kategorie, ale pro ukázkou, obecné pochopení a pro usnadnění rozhodování, jaké body sbírat, je uvedeno rozdělení bodů dle mého osobního mínění:

Turistické body:

- **obec/město:** muzea, kostely, zámky, turistická infocentra, informační tabule, památníky, rozcestníky, ubytování (hotel, penzion atp.), plovárny, sportovní areály atd.,
- **příroda:** studánky, zbořeniště, zříceniny, stezky, rozhledny, boží muka, jezera, hory, přírodní památky, památné stromy, naleziště chráněných druhů rostlin, ptačí oblasti, vyhlídky atd.

Body vyhledávané místními občany:

- úřady a jiné instituce, vzdělávací instituty, knihovny, pošty, banky atd.

Všeobecně zájmové body:

- restaurace, hospody, bary a kavárny, sportoviště a rekreační podniky, obchody a služby, bankomaty, nemocnice a polikliniky, lékárny, parkoviště, čerpací stanice, autobusové, vlakové a jiné stanice a zastávky, kino a divadlo atd.

1.2.2. Zdroje získávaných dat

Databáze bodů POI je vytvářena kombinací volně dostupných globálních a lokálních dat. Mezi globální zdroje dat patří taková data, která jsou celosvětově dostupná jako například OpenStreetMap, Natural Earth²¹ či GeoNames.org²². Mezi lokální zdroje dat pak patří taková data, jež jsou zaměřena na menší celky (lokální charakter) a lze je získat volně na Internetu nebo jsou poskytována partnery projektů Citadel on the Move²³, SDI4apps (oblasti Zemgale, Pošumaví), FOODIE²⁴ (Belluno) a OpenTransportNet²⁵ (Antverpy, Issy). Partnerem projektu Peregrinus a tedy i přispěvatelem do databáze SPOI může být také široká veřejnost, která bude sbírat a získávat (doplňovat) nové body zájmu nebo aktualizovat již zanesené (vedené) body zájmu v databázi. Takto získaná data se poté převádějí do společného datového modelu pomocí harmonizačních procesů (viz kapitola 1.2.5.). (Čerba et al. 2017)

Mezi lokální zdroje dat lze zařadit také jakékoliv databáze památek, přírodních zajímavostí, firem, společností atd., mapové služby a portály se zanesenými body nebo webové stránky jednotlivých obcí a měst či jiných územních celků. U všech zdrojů musí ovšem odpovídat licence, tj. informace musí jít volně přebírat a dále šířit.

Existuje řada lokálních zdrojů v České republice, které by mohly být vytěžovány, tj. z kterých by mohly být body zájmu stahovány, ukládány a poté vkládány do databáze SPOI (příp. čerpány k nim další informace). Největším problémem u přebírání dat obecně je licence, tj. podmínky, za jakých jsou data přístupná a dále použitelná. V případě, že data lze použít, ale licence je jiná než ODbL musí být u bodu uveden zdroj a licence (atribut dc:source a dc:rights, více v kapitole 1.2.4.), pod kterou je publikovaný. Příklady lokálních zdrojů působících v České republice rozdělené dle jejich použitelnosti jsou uvedeny níže, jejich přehledné porovnání s hlavními údaji je v tabulce pod popisem (Tab. 1). Data uvedených zdrojů lze získat mimo dále zmíněné způsoby taktéž pomocí vytěžování skriptem (strukturované soubory či HTML stránky [Faulkner et al. 2017] – viz kapitola 2.1.2.) – netýká se pravděpodobně formátu CSV (Shafranovich 2005), který je ke stažení na stránkách Českého úřadu zeměměřického a katastrálního (ČÚZK) v aplikaci Nahlížení do Katastru nemovitostí (KN²⁶, jedná se o data Registru územní identifikace, adres a nemovitostí

²¹ <http://www.naturalearthdata.com/>

²² <http://www.geonames.org/>

²³ <http://www.citadelonthemove.eu/>

²⁴ <http://www.foodie-project.eu/>

²⁵ <http://opentransportnet.eu/>

²⁶ <http://nahlizeniidokn.cuzk.cz/StahniAdresniMistaRUIAN.aspx>

[RÚIAN]²⁷ – viz popis níže). U jednotlivých zdrojů i v tabulce jsou informace uvedené pouze pro přístupné (veřejné) části, nikoliv pro části určené výhradně specifickým skupinám (např. orgány státní správy aj.) a registrovaným uživatelům.

Využitelné lokální zdroje s jednoznačnou licencí:

eSTUDANKY²⁸

- Jedná se o Národní registr pramenů a studánek, jehož správcem jsou Mladí ochránci přírody²⁹ a jež podporují mimo jiné Lesy ČR, s. p.³⁰, Ministerstvo životního prostředí ČR³¹, Hlavní město Praha³², Mapy.cz, s. r. o.³³ atd.
- Data jsou vedena jako text nebo body v mapě, také je lze jednotlivě stáhnout v souboru GPX (Foster), v němž jsou informace o souřadnicích a výšce a je zde uveden název, odkaz, popis a datum aktualizace.
- Data lze získat stažením jednotlivých GPX souborů nebo ručním přepsáním a jsou vedena pod licencí Creative Commons: CC BY-NC-ND 3.0 CZ³⁴ (Uveďte původ-Neužívejte komerčně-Nezpracovávejte). Pro nekomerční účely a za podmínky uvedení autorství umožňuje tato licence data převzít, ale nesmí se nijak modifikovat.
- Registr má také slovenskou verzi³⁵.

Drobné památky³⁶

- Neoficiální databáze drobných památek České republiky (pomníky, kapličky, boží muka atd.).
- Na stránkách jsou uvedeny 2 rozdílné licence – není-li v konkrétním případě uvedeno jinak, mají fotografie a texty licenci Creative Commons CC BY-NC 3.0³⁷ (Uveďte původ-Zachovejte licenci) anebo CC BY-NC 4.0³⁸ (Uveďte původ-Nepoužívejte komerčně). Obě z těchto licencí povolují pro nekomerční účely přebírání a používání dat za podmínek uvedení autorství a v případě jejich modifikace uvést změny a zachovat licenci.

²⁷ [https://www.cuzk.cz/Uvod/Produkty-a-sluzby/RUIAN/RUIAN-\(1\).aspx](https://www.cuzk.cz/Uvod/Produkty-a-sluzby/RUIAN/RUIAN-(1).aspx)

²⁸ <http://www.estudanky.eu/>

²⁹ <http://www.emop.cz/>

³⁰ <https://lesy-cr.cz/>

³¹ <https://www.mzp.cz/>

³² <http://www.praha.eu/>

³³ <https://mapy.cz/>

³⁴ <https://creativecommons.org/licenses/by-nc-nd/3.0/cz/>

³⁵ <http://sk.estudanky.eu/>

³⁶ <http://www.drobnepamatky.cz/>

³⁷ <https://creativecommons.org/licenses/by-nc/3.0/cz/>

³⁸ <https://creativecommons.org/licenses/by-nc/4.0/deed.cs>

- Data jsou vedena jako text (se souřadnicemi) nebo jako body v mapě (bez uvedených souřadnic) a lze je získat ručním přepisováním. Vhodné použití je také pro vazby jako odkazy pro doplňující informace či webové stránky pojednávající o daném prvku (bodu zájmu).

Lokální zdroje s nejednoznačnou licencí a použitelností:

Registr územní identifikace, adres a nemovitostí (RÚIAN)

- Jeden ze čtyř základních registrů poskytující data ve formě souborů obsahujících data RÚIAN nebo data Informačního systému územní identifikace (ISÚI)³⁹ ve **výměnném formátu RÚIAN (VFR)**⁴⁰ anebo ve formátu **CSV**.
- RÚIAN je veřejný a jeho data jsou tak volně k dispozici pro veřejnou i komerční sféru (otevřená data). Jejich licenční podmínky a tedy i získávání je ovšem nejasné – v „Podmínkách poskytování prostorových dat ČÚZK“⁴¹ je uvedeno, že data RÚIAN ve formátu VFR lze vytěžovat. Dle pracovníka správy RÚIANu lze data (týká se dat ve formátu VFR i CSV) vytěžovat, pokud to nebude časté, tj. dle jeho slov „s rozumem“. Na druhou stranu v „Podmínkách používání aplikace Nahlížení do KN“⁴² je, že se údaje nesmí automaticky vytěžovat a automaticky stahovat.
- Soubory VFR je možné stahovat prostřednictvím aplikace **Veřejný dálkový přístup (VDP)**⁴³.
 - Umožňuje nahlížet a získávat data základního registru RÚIAN a také některá data ISÚI a Informačního systému katastru nemovitostí (ISKN)⁴⁴.
 - Pro přístup do aplikace VDP není potřeba žádné registrace a poskytovaná data z VDP jsou zdarma, nejsou ovšem referenční, tj. mají pouze informativní charakter.
 - Soubory VFR jsou ve formátu XML (Bray et al. 2006) – konkrétně se jedná o formát GML 3.2.1 (OGC: Geography Markup Language), což je XML určené pro geografické aplikace.
 - Data jsou poskytována formou měsíčních stavových souborů (generované automaticky vždy poslední den v měsíci a stažitelná den následující) a denních

³⁹ https://www.cuzk.cz/Uvod/Produkty-a-sluzby/RUIAN/1-Editacni-agendovy-system-ISUI/Editacni-agendovy-system-ISUI.aspx#ui_1

⁴⁰ [https://www.cuzk.cz/Uvod/Produkty-a-sluzby/RUIAN/2-Poskytovani-udaju-RUIAN-ISUI-VDP/Vymenny-format-RUIAN/Vymenny-format-RUIAN-\(VFR\).aspx](https://www.cuzk.cz/Uvod/Produkty-a-sluzby/RUIAN/2-Poskytovani-udaju-RUIAN-ISUI-VDP/Vymenny-format-RUIAN/Vymenny-format-RUIAN-(VFR).aspx)

⁴¹ <http://www.cuzk.cz/Predpisy/Podminky-poskytovani-prostor-dat-a-sitovych-sluzeb/Podminky-poskytovani-prostorovych-dat-CUZK.aspx>

⁴² http://nahlizenidokn.cuzk.cz/Napoveda/index.htm?id=idh_podminkyuzivaniaplikace

⁴³ <http://vdp.cuzk.cz/>

⁴⁴ <https://www.cuzk.cz/Katastr-nemovitosti/O-katastru-nemovitosti/Informacni-system-katastru-nemovitosti-ISKN.aspx>

změnových souborů (generovány každý den s výjimkou víkendů a státních svátků, kdy nedochází ke změnám v systému základního registru).

- Data ve formátu CSV jsou v podobě **Seznamu adresních míst RÚIAN**⁴⁵.
 - Soubory s daty jsou rozděleny a lze je stáhnout po obcích, přičemž stažení je zdarma na webových stránkách ČÚZK⁴⁶ v aplikaci Nahlížení do KN.
 - Soubor obsahuje souřadnice, adresy míst s číslem popisným nebo číslem evidenčním, PSČ a také kód obce a jejích částí.
 - Data jsou aktualizována měsíčně, ale je zde problém se způsobem jejich získávání – viz výše. Stahování souborů s daty je tedy zaručeně možné ručně (manuálně), ale není zřejmé, zda lze automaticky.

Lokální zdroje jednoznačně licencované, ale pravděpodobně nevyužitelné:

Národní památkový ústav (NPÚ)⁴⁷

- Národní památkový ústav je odbornou a výzkumnou organizací státní památkové péče v České republice⁴⁸.
- Žádný z jednotlivých informačních systémů Portálu integrovaného informačního systému památkové péče (IISPP)⁴⁹ NPÚ ani původní aplikace MonumNet⁵⁰ nemá dosud stanovenou konkrétní licenci, pod kterou jsou data publikována. Data jsou volně dostupná, ale pro jejich případné použití je nutno respektovat Autorský zákon (zákon č. 121/2000 Sb., 2000) a vyžádat si od NPÚ souhlas s případným využitím dat pro daný účel. Nicméně, NPÚ plánuje jednotlivé systémy a jejich data do budoucna publikovat pod konkrétními licencemi (nejlépe pod jednotnou).
- **Portál integrovaného informačního systému památkové péče**
 - **Aplikace Památkový katalog**⁵¹
 - Systém evidence všech památek (nemovitých i movitých), památkově chráněných území i ochranných pásem. Součástí evidence je také evidence „objektů památkového zájmu“ (objekty, které nebyly prohlášeny za památky) a monitoring aktuálního stavu, který umožňuje generování seznamů ohrožených a zachráněných památek.
 - Aktualizace veřejného katalogu je prováděna průběžně.

⁴⁵ http://vdp.cuzk.cz/vymenny_format/csv/ad-csv-struktura.pdf

⁴⁶ <http://www.cuzk.cz/>

⁴⁷ <https://www.npu.cz/cs>

⁴⁸ <https://www.mkcr.cz/pamatkova-pece-19.html>

⁴⁹ <https://iispp.npu.cz/>

⁵⁰ <http://monumnet.npu.cz/monumnet.php>

⁵¹ <http://pamatkovykatalog.cz/>

- Ve veřejné části je umožněno vyhledávání a prezentace údajů (jako text) a také možnost stažení výsledků ve formátu XLSX ([MS-XLSX] 2017), ale bez odkazů a souřadnic. Dalším způsobem, jak získat údaje, je jejich ruční přepisování.
- **Informační systém o archeologických datech (ISAD)**⁵²
 - Jednotný sklad všech archeologických dat získávaných v procesu péče o archeologický fond, a to jako nezbytný předpoklad zajištění správy a aktualizace těchto dat a jejich efektivního využití.
 - Systém je průběžně doplňován a aktualizován.
 - Data jsou rozdělena do dvou databází (seznamů) – Státní archeologický seznam ČR⁵³, kde jsou informace veřejné, ale jednotlivé prvky jsou zobrazovány jako seznamy pouze ve formě textu a lze je tedy získat výhradně ručním přepisem, a Významné archeologické lokality⁵⁴, kde jsou informace také veřejné, ale data lze získat navíc pomocí skriptu, jelikož má HTML stránka strukturovaný text (více v kapitole 2.1.2.).
- **Dokumentační fondy a knihovny – webový katalog Tritius 1.0**⁵⁵
 - Katalog umožňuje vyhledávat informace o knihách, časopisech a dalších dokumentech v celém knihovním a dokumentačním fondu všech pracovišť NPÚ.
 - Ve fondu jsou uloženy odborné publikace, fotografie, zprávy, mapy, plány a další dokumenty z oboru památkové péče, dějin umění, architektury, konzervace a restaurování památek.
 - Vhodné použití těchto dat je pro vazby jako odkazy s dalšími informacemi k objektu (bodů zájmu).
 - Seznam je doplňován průběžně a je veřejný, ale jednotlivé informace je možné z výpisu (seznamu) získat pouze ručním přepsáním, nelze je stáhnout.

⁵² <http://isad.npu.cz/>

⁵³ <http://isad.npu.cz/ost/archeologie/ISAD/free/>

⁵⁴ <http://isad.npu.cz/ost/archeologie/val-free/>

⁵⁵ <https://iispp.npu.cz/carmen/>

- **Metainformační systém (MIS)**⁵⁶
 - Systém zajišťuje jednotné ukládání, popis a zpřístupnění digitálních nebo digitalizovaných dokumentů (fotografie, mapy, plány, textové dokumenty apod.) týkajících se objektů zájmu památkové péče nebo obecně odborné činnosti NPÚ.
 - Tato část IISPP je nejlépe využitelná pro vazby jako odkazy s dalšími informacemi k objektu (bodu zájmu).
 - Ve veřejně přístupné části je umožněno vyhledávání a zobrazení základních popisných údajů a náhledů veřejných dokumentů.
 - Informace lze získat pouze ručním přepsáním, nelze je nijak stáhnout, kromě konkrétních dokumentů ve formátu PDF (ISO 32000-1:2008 2008).
 - Doplnování systému probíhá průběžně, přičemž zveřejněné dokumenty jsou autorská díla, která lze užít pouze na základě písemného souhlasu autora nebo nositele autorských práv.
- **Aplikace MonumNet**
 - Jedná se o původní portál IISPP.
 - Roku 2016 byla ukončena aktualizace dat Ústředního seznamu kulturních památek České republiky⁵⁷ v systému MonumNet, ale systém zůstává nadále k dispozici jako archivní a jsou v něm údaje o památkách prohlášených do 31. 12. 2016.
 - Je to dočasně paralelně udržovaná evidence památek (nemovité památky, světové dědictví, národní kulturní památky, chráněná území, seznam ohrožených nemovitých památek, operativní průzkumy a dokumentace,...).
 - Jednotlivé položky databází (evidencí) jsou vedeny jako veřejné seznamy ve stylu tabulky, tedy XLS ([MS-XLS] 2017) a jako text.
 - Data lze stáhnout v tabulce (XLS), ale nejsou zde uvedeny souřadnice, nebo je lze získat ručním přepisováním.

⁵⁶ http://iispp.npu.cz/mis_public/homepage.htm

⁵⁷ https://cs.wikipedia.org/wiki/%C3%9Ast%C5%99edn%C3%AD_seznam_kulturn%C3%ADch_pam%C3%A1tek_%C4%8Cesk%C3%A9republiky

Webové stránky krajů

- Mezi další vhodné lokální zdroje by mohly patřit webové stránky krajů. Záleží ovšem na možnosti využití vedených dat, které je nutné domluvit s provozovatelem daných webových stránek a to z důvodu autorského práva (copyright).
- Příkladem může být webová stránka pro Plzeňský kraj, konkrétně **Turistický portál Plzeňského kraje – Turistů ráj**⁵⁸, nebo pro kraj Jihočeský, kde informace věnované turistice, památkám a dalším událostem lze nalézt v **Informačním systému cestovního ruchu Jihočeského kraje – Jižní Čechy**⁵⁹. Ostatní krajské stránky lze využít stejným způsobem.
- Data jsou vedena ve formě textu a jejich získání je tedy možné pouze ručním přepisem.

Sportovní a turistické webové stránky

- Stejným typem zdroje, co se týče přebírání a dalšího použití dat, je webová stránka **Šumavské okruhy – Prožijte Šumavu jinak!**⁶⁰.
- Byla vytvořena nadšenými sportovci a nabízí cyklistické, běžecké a lyžařské okruhy na Šumavě, pro něž lze stáhnout jejich GPS souřadnice⁶¹ ve formátu GPX, KML (OGC: KML) či GDB. Ostatní informace jsou pouze jako text a jediná možnost, jak je získat, je ruční přepsání.
- Mezi podobné webové stránky patří např. **Křivoklátské okruhy**⁶² a **Javornické okruhy**⁶³.

⁵⁸ <http://www.turisturaj.cz/>

⁵⁹ <https://www.jiznicechy.cz/>

⁶⁰ <http://www.sumavskeokruhy.cz/>

⁶¹ https://cs.wikipedia.org/wiki/Zem%C4%9Bpisn%C3%A9_sou%C5%99adnice

⁶² <http://www.krivoklatskeokruhy.cz/>

⁶³ <http://javornickeokruhy.cz/>

Přehledné porovnání příkladů lokálních zdrojů působících v České republice s hlavními údaji.

Databáze / seznam / registr	Poskytovatel	Licence	Formát	Stahování / získání dat	Aktualizace
Využitelné lokální zdroje s jednoznačnou licencí					
Národní registr pramenů a studánek	eSTUDANKY	CC BY-NC-ND 3.0 CZ	text, mapa, GPX	GPX, ručně	různá
Neoficiální databáze drobných památek	Drobné památky	CC BY-NC 3.0 nebo CC BY-NC 4.0	text, mapa	ručně	různá
Lokální zdroje s nejednoznačnou licencí a použitelností					
Veřejný dálkový přístup	RÚIAN	není jasně stanovena	Výměnný formát RÚIAN (GML)	Výměnný formát RÚIAN (GML)	průběžná – měsíční stavová a denní změnová
Seznam adresních míst	RÚIAN	není jasně stanovena	CSV	CSV	měsíční
Lokální zdroje jednoznačně licencované, ale pravděpodobně nevyužitelné					
IISPP Památkový katalog	NPÚ	copyright	XLSX, text	XLSX, ručně	průběžná
IISPP ISAD	NPÚ	copyright	text	ručně	průběžná
IISPP Trinitus	NPÚ	copyright	text	ručně	průběžná
IISPP MIS	NPÚ	copyright	text	ručně	průběžná
MonumNet	NPÚ	copyright	XLS, text	XLS, ručně	již neaktualizována
Turistický portál	Plzeňský kraj – Turistů ráj	copyright	text	ručně	nepravidelná
Informační systém cestovního ruchu	Jihočeský kraj – Jižní Čechy	copyright	text	ručně	nepravidelná
Šumavské okruhy	Šumavské okruhy – Prožijte Šumavu jinak!	copyright	GPX, KML a GDB, text	GPX, KML a GDB, ručně	nepravidelná

Tab. 1: Přehled příkladů lokálních zdrojů v České Republice se základními informacemi
Zdroj: vlastní

Budoucí možný zdroj dat a informací:

Zlatá stezka – síť turistických cest na „Zelené střeše Evropy“

- Jedná se o projekt Evropské unie, České republiky (Jihočeského⁶⁴, Plzeňského⁶⁵ a Karlovarského⁶⁶ kraje) a Svobodného státu Bavorsko⁶⁷ – Program přeshraniční spolupráce Česká republika – Svobodný stát Bavorsko Cíl EÚS 2014–2020⁶⁸.
- Cílem projektu je vytvoření paralelní pěší Zlaté stezky vedoucí územím Jihočeského a Plzeňského kraje zrcadlově k již fungující Zlaté stezce v Bavorsku, tj. vznik nových turistických tras a jejich propojení se stávajícími.
- Datum ukončení projektu je 30. 6. 2019 a měla by být vytvořena pěší stezka na území České republiky s připojením na bavorské pěší stezky prostřednictvím propojovacích tras a zprovozněny také internetové stránky.
- Jako další zájmové body do databáze SPOI lze využít nově vzniklé prvky v rámci Zlaté stezky (samotná stezka, rozcestníky, informační tabule, odpočinkové zastávky atd.).

⁶⁴ <https://www.jccr.cz/projekty/zlata-stezka>

⁶⁵ <http://www.turisturaj.cz/clanek/projekt-zlata-stezka-sit-turistickyh-cest-na-zelene-strese-evropy>

⁶⁶ http://www.kr-karlovarsky.cz/region/dotace/Stranky/preshranicni_programy/Bavorsko.aspx

⁶⁷ <https://www.efre-bayern.de/europaeische-territoriale-zusammenarbeit/grenzuebergreifende-zusammenarbeit/bayern-tschechien/>

⁶⁸ <https://www.by-cz.eu/cz/aktuality/>

1.2.3. Druhy vazeb mezi sbíranými daty

Sbíraná data (body) mají určité závislosti a souvislosti (vztahy) mezi sebou. Pomocí atributů mezi body lze tyto vztahy vyjádřit, propojit tak data (Linked data) a vytvořit síť provázaných bodů, které na sebe odkazují. Takovéto atributy se nazývají vazby. Propojení SPOI je realizováno standardizovanými vazbami, jimiž jsou:

- **sémantická/slovníková vazba** – SPOI Ontologie (viz kapitola 1.2.1.) – jedná se o zařazení bodu do určité kategorie (třídy) dle jeho významu – např. kosmetický salon, nehtové studio i kadeřnictví patří do skupiny „beauty“,
- **vazba na identické objekty** – DBpedia⁶⁹ (strukturovaná data vytěžená z Wikipedie⁷⁰ a Wikidat⁷¹ zpřístupněná na Internetu jako součást sémantického webu⁷²), GeoNames.org (stažitelná geografická databáze všech zemí), OpenStreetMap (volně využitelná mapa světa) – propojení bodu s identickým bodem v jiné databázi, mapě – např. pro Karlův most existuje ekvivalent Charles Bridge⁷³ (DBpedia), Karlův most⁷⁴ (GeoNames.org) a Karlův most⁷⁵ (OpenStreetMap),
- **topologická vazba (administrativní jednotky)** – DBpedia, GeoNames.org – pro body nacházející se v určité administrativní jednotce – např. Katedrála svatého Bartoloměje v Plzni⁷⁶ (GeoNames.org), v okrese Plzeň – město⁷⁷ (GeoNames.org), v Plzeňském kraji⁷⁸ (DBpedia), v České Republice⁷⁹ (DBpedia),
- **vazba na rozšiřující data a informace** (obrázky, webové stránky, mapy, dokumenty atd.)⁸⁰.

Všechny atributy (vlastnosti) vycházejí z různých dokumentů, standardů a datových sad (dále jen dokumentů), kde jsou popisovány. Zde jsou uvedeny jednotlivě, přičemž zkratka (řetězec znaků) uvedena u každého dokumentu definujícího jednotlivé atributy je typický prefix (viz dále) a odkazy jsou konkrétní jmenné prostory, které se doplňují do hlavičky dokumentu s datovou sadou a ze kterých se implementují objektové atributy.

⁶⁹ <http://wiki.dbpedia.org/>

⁷⁰ <https://cs.wikipedia.org/wiki/Wikipedie>

⁷¹ https://www.wikidata.org/wiki/Wikidata:Main_Page

⁷² https://cs.wikipedia.org/wiki/S%C3%A9mantick%C3%BD_web

⁷³ http://dbpedia.org/page/Charles_Bridge

⁷⁴ <http://www.geonames.org/6269522/karluv-most.html>

⁷⁵ <https://www.openstreetmap.org/way/119016167#map=17/50.08663/14.41056>

⁷⁶ <http://www.geonames.org/3068160/pilsen.html>

⁷⁷ <http://www.geonames.org/3068158/okres-plzen-mesto.html>

⁷⁸ http://dbpedia.org/page/Plze%C5%88_Region

⁷⁹ http://dbpedia.org/page/Czech_Republic

⁸⁰ https://www.researchgate.net/publication/317956781_Smart_Points_of_Interest

- Dokument „**RDF Schema 1.1 – W3C**“ (Brickley a Guha 2014)
 - Doporučení, které je ale používáno jako standard.
 - Ontologický jazyk vyvinutý pod dohledem konsorcia W3C (World Wide Web Consortium)⁸¹.
 - Jazyk je součástí rodiny specifikací RDF a rozšiřuje jazyk RDF o možnost vytvářet vlastní ontologie a RDF slovníky.
 - Dokument slouží pro rozlišení typů zdrojů, základní práci s třídami a vlastnostmi.
 - rdfs:
 - <<http://www.w3.org/2000/01/rdf-schema#>>

- Dokument „**OWL 2 Web Ontology Language Quick Reference Guide (Second Edition) – W3C**“ (Bao et al. 2012)
 - Doporučení, které je také používáno jako standard.
 - Značkovací jazyk vyvinutý konsorciem W3C.
 - Slouží pro tvorbu ontologií využitelných v prostředí sémantického webu – rozšíření možnosti popisu ontologií.
 - Dokument popisuje jazyk OWL a odkazuje na jednotlivé třídy, vlastnosti, individuály a hodnoty a také na další dokumenty týkající se OWL.
 - owl:
 - <<http://www.w3.org/2002/07/owl#>>

- Dokument „**FOAF Vocabulary Specification 0.99**“ (Khandelwal et al. 2014)
 - Specifikace FOAF je součástí projektu FOAF⁸² a slouží jako hlavní dokumentace pro slovní zásobu RDF/XML a formáty dokumentů označované jako „FOAF“.
 - Tato specifikace popisuje jazyk FOAF definovaný jako slovník vlastností a tříd používající technologii RDF pro propojení lidí, aktivit a informací o nich – celý název zkratky projektu je „Friend of a Friend“.
 - Dokument popisuje jednotlivé třídy a vlastnosti.
 - foaf:
 - <<http://xmlns.com/foaf/0.1/>>

⁸¹ <https://www.w3.org/>

⁸² <http://www.foaf-project.org/>

- Dokument „**The Organization Ontology**“ (Reynolds 2014)
 - Jedná se o doporučení konsorcia W3C.
 - Ontologie byla původně vyvinuta a publikována společností Epimorphics⁸³, ale rozšířena a dále rozvíjena byla v rámci skupiny W3C Government Linked Data Working Group⁸⁴.
 - Dokument popisuje základní ontologii organizační struktury zaměřenou na publikování propojených dat organizačních informací uvnitř jedné nebo mezi vícero organizacemi.
 - org:
 - <<http://www.w3.org/ns/org#>>

- Dokument „**SKOS Simple Knowledge Organization System Reference**“ (Miles a Bechhofer 2009)
 - Dokument je doporučením konsorcia W3C.
 - Definiuje Simple Knowledge Organization System (SKOS) a společný datový model pro reprezentaci, sdílení a propojení „systémů organizace znalostí“ prostřednictvím webu.
 - SKOS umožňuje sdílení dat a technologií v různých aplikacích.
 - skos:
 - <<http://www.w3.org/2004/02/skos/core#>>

- Standard „**GeoSPARQL – A Geographic Query Language for RDF Data**“ (Perry a Herring 2012)
 - Standard Open Geospatial Consortium (OGC)⁸⁵
 - GeoSPARQL podporuje reprezentaci a dotazování prostorových dat na sémantickém webu.
 - Definiuje slovní zásobu pro reprezentaci prostorových dat v RDF a definuje rozšíření dotazovacího jazyka SPARQL (Harris a Seaborne 2013) pro zpracování prostorových dat.
 - geos:
 - <<http://www.opengis.net/ont/geosparql#>>

⁸³ <http://epimorphics.com/public/vocabulary/org.html>

⁸⁴ https://www.w3.org/2011/gld/wiki/Main_Page

⁸⁵ <http://www.opengeospatial.org/>

- Datová sada **GeoNames.org** (Vatant a Wick 2012)
 - Geografická databáze, která obsahuje více než 11 milionů geografických názvů, jež pokrývají všechny státy a které jsou k dispozici ke stažení zdarma.
 - Ontologie GeoNames.org přidává geografické sémantické informace na World Wide Web.
 - Všechny toponyma mají jedinečnou adresu URL⁸⁶ s příslušnou webovou službou RDF a jednotlivé prvky na sémantickém webu GeoNames.org jsou vzájemně provázány.
 - gn:
 - <<http://www.geonames.org/ontology/>>

- Datová sada **DBpedia** (DBpedia 2008)
 - Projekt, který těží z Wikipedie a Wikidat strukturovaná data a zpřístupňuje je na webu jako součást webu sémantického v podobě RDF záznamů.
 - Ontologie DBpedia poskytuje třídy a vlastnosti používané v datové sadě DBpedia.
 - dbo:
 - <<http://dbpedia.org/ontology/>>

- Dokument „**DCMI Metadata Terms**“ (DCMI Usage Board 2012a)
 - Dublin Core Metadata Initiative (DCMI)⁸⁷ doporučení.
 - Dokument je aktuální specifikací všech metadat, která jsou udržována organizací DCMI, včetně vlastností, tříd, systému kódování slovníků a systému kódování syntaxe.
 - Zahrnuje také 15 termínů z Dublin Core Metadata Element Set, které byly také publikovány jako IETF RFC (Kunze a Baker 2017), ANSI/NISO Standard (ANSI/NISO Z39.85-2012 2012) a ISO Standard (ISO 15836-1:2017(en) 2017).
 - dcterms:
 - <<http://purl.org/dc/terms/1.1/>>

⁸⁶ https://aleph.nkp.cz/F/?func=direct&doc_number=000000568&local_base=KTD

⁸⁷ <http://dublincore.org/>

- Dokument „**Dublin Core Metadata Element Set, Version 1.1: Reference Description**“ (DCMI Usage Board 2012b)
 - Dokument je výňatkem z obsáhlejšího dokumentu „DCMI Metadata Terms“.
 - Poskytuje zkrácenou referenční verzi patnácti popisů prvků (vlastností pro popis zdrojů), které byly formálně schváleny v normách ISO Standard (ISO 15836-1:2017(en) 2017), ANSI / NISO Standard (ANSI/NISO Z39.85-2012 2012) a IETF RFC (Kunze a Baker 2017).
 - Více viz dokument „DCMI Metadata Terms“.
 - dc:
 - <<http://purl.org/dc/elements/1.1/>>

- Dokument „**ISA Programme Location Core Vocabulary**“ (Perego a Lutz 2015)
 - Dokument poskytuje minimální soubor tříd a vlastností pro popis jakéhokoliv místa z hlediska jeho jména, adresy nebo geometrie.
 - Slovní zásoba je určena konkrétně pro zveřejňování údajů (dat), které jsou interoperabilní se směrnicí EU INSPIRE⁸⁸.
 - locn:
 - <<http://www.w3.org/ns/locn#>>

- Datová sada **SPOI**⁸⁹
 - Linked Open Database – otevřená, propojená a bezešvá datová sada Smart Points of Interest.
 - Datová sada obsahuje přes 28 000 000 POI bodů.
 - Dokument o datovém modelu „Smart Points of Interest: Data Model“ (Čerba 2017).
 - poi:
 - <<http://www.openvoc.eu/poi#>>

Jak již bylo napsáno výše, propojení dat (bodů) mezi sebou se děje pomocí vazeb (přiřazování odkazů). Tyto vazby (platí také pro všechny ostatní atributy) sestávají ze 3 základních komponentů – jedná se o subjekt neboli popisovaný bod (zdroj), predikát neboli vlastnost (atribut) subjektu (více v následující kapitole) a objekt neboli hodnota atributu (odkaz, textový řetězec). Pokud bychom tuto definici převedli do věty, jednalo

⁸⁸ <http://inspire.ec.europa.eu/>

⁸⁹ <http://sdi4apps.eu/spoi/>

by se o podměť, přísudek a předmět. Jinak řečeno bod jako zdroj má konkrétní vlastnosti (atributy), které určují jeho charakter a tyto atributy mají odpovídající hodnoty (odkaz či konkrétní údaj – textový řetězec). Atribut jako vazba zároveň vyjadřuje vzájemný vztah mezi subjektem a objektem. V odborné terminologii se mluví o RDF trojici (RDF Triple) (Cyganiak at al. 2014).⁹⁰

Každý atribut je definován a popsán v jemu odpovídajícím dokumentu, standardu či datové sadě (viz výše), jehož název (typ) lze většinou vyčíst z prefixu (zpravidla část nebo zkratka názvu dokumentu). Prefix mají definovány všechny dokumenty a je to zkratka pro URI⁹¹ adresu, která je určena a napsána na začátku dokumentu (kódu popisujícím jednotlivé body) a odkazuje na daný dokument. Zkratka URI adresy je kód (řetězec znaků) sloužící k identifikaci dokumentu nebo přesné specifikaci jiného zdroje informací na Internetu.⁹² Po prefixu následuje název atributu (vlastnosti), který je oddělen dvojtečkou (:), a který spojuje subjekt a objekt (např. *restaurace se nachází ve městě*).

⁹⁰ https://cs.wikipedia.org/wiki/Resource_Description_Framework

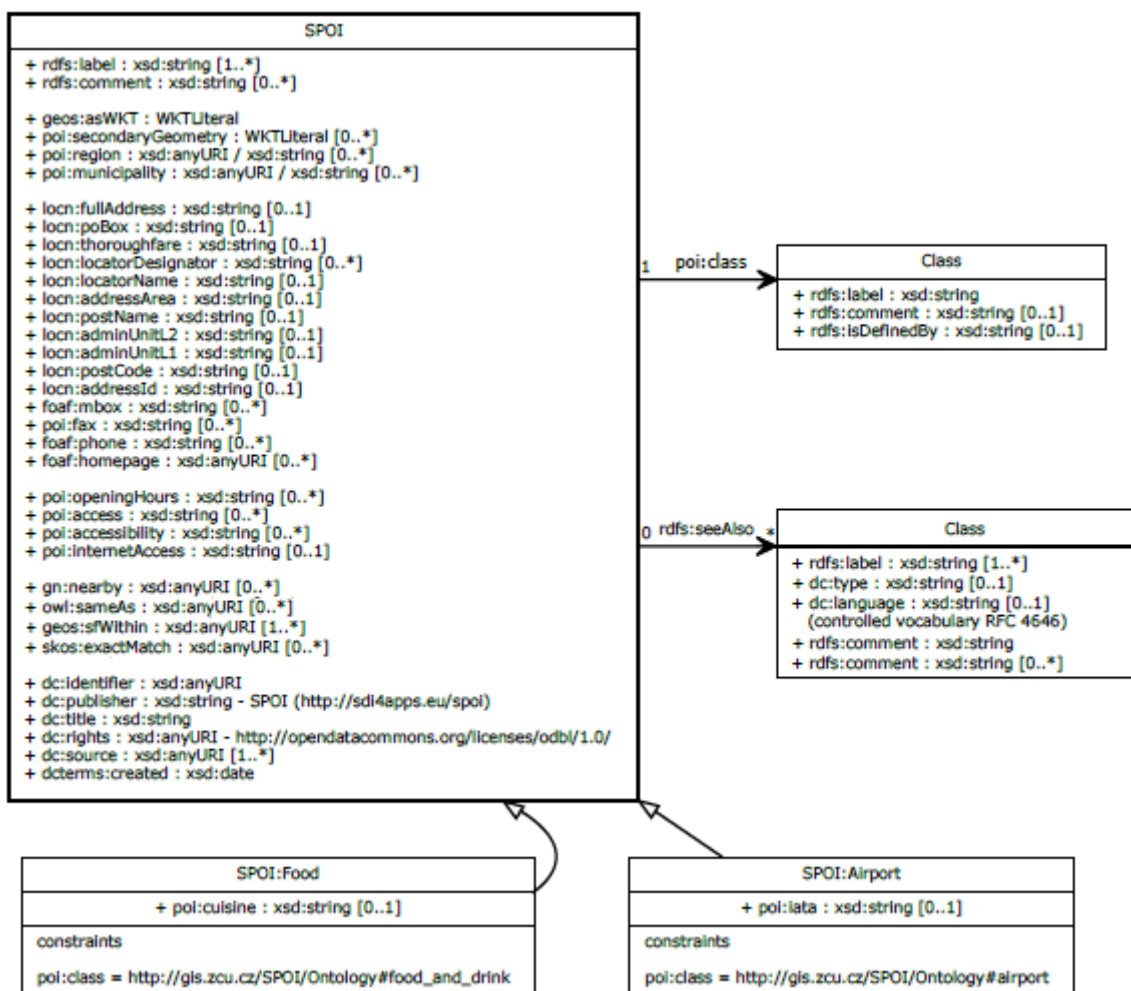
⁹¹ https://aleph.nkp.cz/F/?func=direct&doc_number=000000567&local_base=KTD

⁹² https://cs.wikipedia.org/wiki/Uniform_Resource_Identifier

1.2.4. Datový model

V současné době se databáze SPOI nachází ve stavu, kdy obsahuje velké množství atributů, které jsou například významem stejné nebo jsou použity u malého množství bodů (v některých případech pouze u jednoho bodu). Toto je způsobeno tím, že některé atributy ze starších verzí datového modelu nejsou zatím opraveny, což by mělo změnit nahrání aktuální nové verze. U SPARQL dotazů (viz kapitola 2.1.1.) se navíc propagují další atributy z jiných Linked data zdrojů, tj. mají stejný význam jako použité atributy v databázi SPOI, ale vychází z jiného dokumentu (viz předchozí kapitola).

Datový model SPOI⁹³ (platný k 3. 4. 2017 – Obr. 1) je uniformní (jednotný) a je založen na standardech, sémantickém popisu a Linked Data přístupu. (Čerba a Mildorf 2016) Je také ale otevřený a flexibilní, což znamená, že ho lze přizpůsobit požadavkům uživatele.



Obr. 1: Datový model SPOI k 3. 4. 2017 (graficky upravený)

Zdroj: SDI4APPS – SPOI

⁹³ http://sdi4apps.eu/spoi/doc/POI_data_model.pdf

Linked Data jsou propojená data pocházející z různých zdrojů. Toto propojení je uskutečněno pomocí vazeb (viz předchozí kapitola). Všechny atributy včetně vazeb se dělí do několika kategorií podle povahy (druhu) informace, kterou nesou, resp. na kterou odkazují. Kategorie i s výčtem atributů, které zahrnují, vypadají následovně (podtržené atributy jsou povinné):

- **popis:** rdfs:label, rdfs:comment,
- **lokalizace:** geos:asWKT, poi:secondaryGeometry, poi:region, poi:municipality,
- **klasifikace:** poi:class,
- **kontaktní informace:** locn:fullAddress, locn:poBox, locn:thoroughfare, locn:locatorDesignator, locn:locatorName, locn:addressArea, locn:postName, locn:adminUnitL2, locn:adminUnitL1, locn:postCode, locn:addressId, foaf:mbox, poi:fax, foaf:phone, foaf:homepage,
- **turistické informace:** poi:openingHours, poi:iata, poi:access, poi:accessibility, poi:cuisine, poi:internetAccess,
- **odkazy:** rdfs:seeAlso, skos:exactMatch, owl:sameAs, geos:sfWithin, gn:nearby,
- **metadata:** dc:identifier, dc:publisher, dc:title, dc:rights, dc:source, dcterms:created.

Povinné atributy jsou pro bod zájmu nezbytné (pro jeho identifikaci, lokalizaci a další věci spojené s databází nebo legislativou), neznamená to ale, že jsou všechny uváděny (vyplňovány) „sběračem“ – některé tyto atributy jsou doplňovány až při vkladu do databáze SPOI (viz kapitola 2.1.2.). Níže je u některých z nich uveden způsob zápisu hodnoty atributu – URI (internetová adresa – odkaz) nebo textový řetězec (řetězec znaků – písmena, číslice, symboly). V případě, že není uvedena ani jedna z možností či jejich kombinace, musí být hodnota zapsána doslovně (jednoznačně – např. nepsat zkratky názvů v adrese, pokud tam nejsou, nezaokrouhlovat souřadnice, psát celý [oficiální] název státu atd.). Toto neplatí pro dc:publisher a dc:rights, kde je hodnotou vždy odkaz na SPOI (<http://sdi4apps.eu/spoi>), resp. na OdbL (<http://opendatacommons.org/licenses/odbl/1.0/>) a dcterms:created, kde je uváděno datum zaevidování.

Popis jednotlivých atributů:

- **rdfs:label** – jméno/název POI bodu (textový řetězec),
- **rdfs:comment** – podrobnější popis bodu (textový řetězec),
- **geos:asWKT** – souřadnice bodu (zeměpisná délka a zeměpisná šířka),
- **poi:secondaryGeometry** – informace o sekundární geometrii – linie nebo polygon,
- **poi:region** – odkaz na region, do kterého bod spadá (okres, kraj, stát,...) nebo jeho název (URI nebo textový řetězec),

- **poi:municipality** – odkaz, do které obce bod spadá (kde se nachází), nebo její název (URI nebo textový řetězec),
- **poi:class** – kategorie, základní třída, která označuje bod a která vychází ze SPOI ontologie (anglicky) – je na ní odkazováno (URI),
- **locn:fullAddress** – kompletní adresa bodu ve formě textového řetězce,
- **locn:poBox** – číslo poštovní schránky (jednotky), které slouží k jednoznačné identifikaci bodu v budově s více jednotkami,
- **locn:thoroughfare** – součást adresy, jež představuje identifikaci bodu na úrovni ulice,
- **locn:locatorDesignator** – číslo nebo řada znaků, které jednoznačně identifikují místo (bod) v rámci příslušné oblasti (číslo budovy),
- **locn:locatorName** – jako locn:locatorDesignator, ale jedná se o vlastní jméno, které neobsahuje číslice (název nemovitosti, komplexu, budovy nebo části budovy, jméno místnosti uvnitř budovy,...),
- **locn:addressArea** – jméno geografické oblasti, která sdružuje mnoho adres, ale není administrativní jednotkou (část města, sousedství,...),
- **locn:postName** – název pošty,
- **locn:adminUnitL2** – název kraje,
- **locn:adminUnitL1** – název státu,
- **locn:postCode** – poštovní směrovací číslo,
- **locn:addressId** – globálně jedinečný identifikátor pro každou adresu (např. některé státy Evropské unie mají implementována ID do adresních registrů),
- **foaf:mbox** – e-mailová adresa na bod či správce/vlastníka/provozovatele bodu (textový řetězec),
- **poi:fax** – fax (číslo) na bod či správce/vlastníka/provozovatele bodu (textový řetězec),
- **foaf:phone** – telefonní číslo na bod či správce/vlastníka/provozovatele bodu (textový řetězec),
- **foaf:homepage** – odkaz na webové stránky bodu – hlavní/základní stránky pojednávající o bodu (URI),
- **poi:openingHours** – otevírací doba, pokud ji zájmový bod má (textový řetězec),
- **poi:iata** – kód letiště IATA⁹⁴ (textový řetězec),

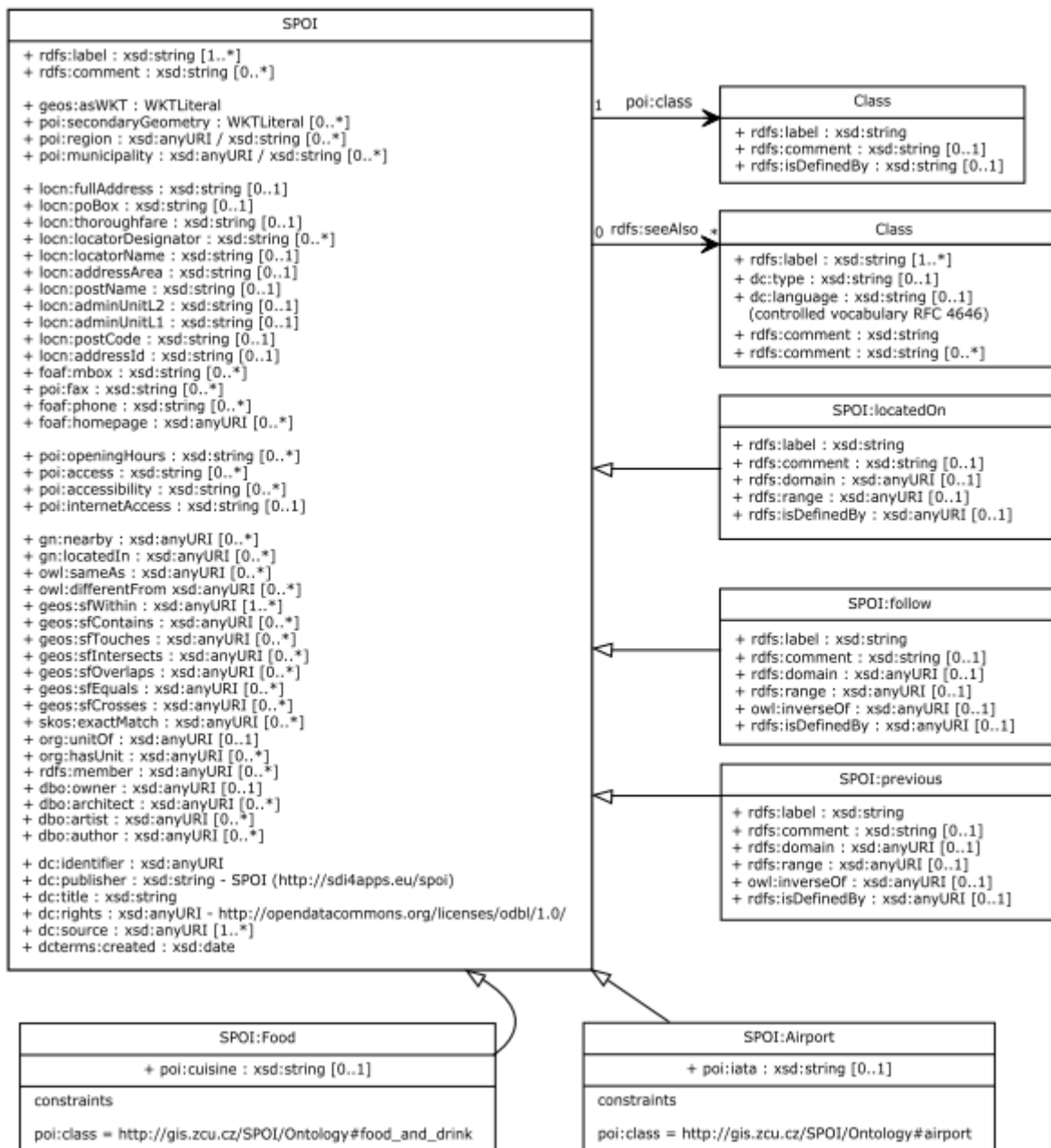
⁹⁴ <http://www.iata.org/publications/Pages/code-search.aspx>

- **poi:access** – popis způsobu přístupu k bodu (textový řetězec),
- **poi:accessibility** – informace o přístupnosti bodu např. pro handicapované (textový řetězec),
- **poi:cuisine** – styl kuchyně restaurace či baru (textový řetězec),
- **poi:internetAccess** – možnost internetového připojení (textový řetězec),
- **rdfs:seeAlso** – odkaz na externí doplňující informace o bodu (URI),
- **owl:sameAs** – odkaz na externí propojená data, která reprezentují ten samý bod – synonyma (URI),
- **skos:exactMatch** – stejně jako owl:sameAs (URI),
- **geos:sfWithin** – topologická vazba, která se používá pro vyjádření, kdy je jeden objekt prostorově uvnitř objektu druhého – např. kostel v dané obci nebo odkaz na příslušný stát, ve kterém se bod zájmu nachází, pomocí dat GeoNames.org a DBpedia (URI),
- **gn:nearby** – vazba, která ukazuje na blízké prvky, jež ale nejsou sousedící, více viz GeoNames.org Ontology⁹⁵ (URI),
- **dc:identifier** – kopie jedinečného identifikátoru objektu (URI),
- **dc:publisher** – vydavatel dat (SPOI – <http://sdi4apps.eu/spoi>),
- **dc:title** – první nebo hlavní označení bodu zájmu – nejdůležitější název (textový řetězec),
- **dc:rights** – odkaz na licenci zájmového bodu v databázi SPOI (ODbL – <http://opendatacommons.org/licenses/odbl/1.0/>),
- **dc:source** – odkaz na původní zdroj dat (URI),
- **dcterms:created** – datum zaevidování bodu (datum).

Do kategorie odkazů byly ještě přidány další standardizované a také nově navržené vazby, o kterých pojednává v diplomové práci Ing. Martin Királ' (2017). Upravený datový model lze vidět na Obr. 2. Přidané vazby jsou:

- **standardizované:** owl:differentFrom, geos:sfContains, geos:sfTouches, geos:sfIntersects, geos:sfOverlaps, geos:sfEquals, geos:sfCrosses, org:unitOf, org:hasUnit, rdfs:member, dbo:owner, dbo:architect, dbo:artist, dbo:author, gn:locatedIn,
- **nově navržené**⁹⁶: poi:locatedOn, poi:follow, poi:previous.

⁹⁵ <http://www.geonames.org/ontology/documentation.html>



Obr. 2: Rozšířený datový model SPOI dle Ing. Martina Királ'a

Zdroj: Diplomová práce – Objektové atributy pro prostorová Linked Data (Királ' 2017)

Popis jednotlivých vazeb navržených Királ'em (2017):

- **owl:differentFrom** – označení rozdílného bodu stejného typu v jednom místě – 2 různé body (URI),
- **geos:sfContains** – propojení bodů, kdy jeden bod obsahuje bod druhý (nachází se v něm) – inverzní k vazbě geos:sfWithin (URI),
- **geos:sfTouches** – pro sousedící body – jejich geometrie se prostorově dotýkají (URI),

⁹⁶ Prefix a tedy i jmenný prostor vycházející z datové sady SPOI platí pro tyto nově navržené vazby až ve chvíli, kdy budou začleněny do datového modelu SPOI (viz dále). V současné době mají tyto nově navržené vazby prefix SPOI a jejich jmenným prostorem je <http://gis.zcu.cz/SPOI/OntologyRelations#>.

- **geos:sfIntersects** – pro protínající se či dotýkající se prostorové body/objekty – geometrie objektů mají společný prvek – např. bod nacházející se v lese (URI),
- **geos:sfOverlaps** – objekty se stejnou prostorovou dimenzí, které se navzájem překrývají, jeden přesahuje přes druhý – např. rybník v lese (URI),
- **geos:sfEquals** – ukazuje na prostorově stejně rozložené objekty – jejich geometrie se prostorově rovnají – např. Hlavní město Praha a kraj Praha (URI),
- **geos:sfCrosses** – pro objekt (bod), který křížuje objekt jiný – např. most na řece (URI),
- **org:unitOf** – pro body spadající pod jiné – odkazuje na nadřazené (obvykle větší) celky (body), jichž je tato jednotka (bod) součástí (např. pobočka vůči hlavní organizaci/sídlu) – inverzní vazba k org:hasUnit (URI),
- **org:hasUnit** – pro body, které mají podřízené části – dělí se na menší podřízené objekty, na které vazba odkazuje (jsou součástí tohoto nadřazeného (většího) celku – např. hlavní organizace/sídlu a její pobočky) – inverzní vazba k org:unitOf (URI),
- **rdfs:member** – odkaz na zdroj, jehož je objekt členem – pro spojení více bodů do jednoho – např. sochy v sousoší (URI),
- **dbo:owner** – odkaz na vlastníka daného objektu (URI),
- **dbo:architect** – odkaz na architekta daného objektu (památky) na data DBpedia, Wikidata,... (URI),
- **dbo:artist** – odkaz na umělce, který vytvořil dílo, na data DBpedia, Wikidata (URI),
- **dbo:author** – odkaz na autora daného díla na data DBpedia, Wikidata,... (URI),
- **gn:locatedIn** – odkaz na prvek, v němž je bod zájmu lokalizován – např. v konkrétní budově (URI),
- **poi:locatedOn** – pro body umístěné na jiném objektu – např. bankomat, upomínková deska na fasádě budovy (URI),
- **poi:follow** – vazba pro spojení následující – odkaz na objekt následující po daném bodu zájmu (např. naučné tabule na turistických trasách) – inverzní k poi:previous (URI),
- **poi:previous** – vazba pro spojení předcházející – odkaz na objekt předcházející danému bodu zájmu (např. naučné tabule na turistických trasách) – inverzní k poi:follow (URI).

V současné době se připravuje nová verze datového modelu, ve kterém budou vypuštěny vazby owl:sameAs, poi:region, poi:municipality a povinná vazba poi:class bude nahrazena

povinnou standardizovanou vazbou owl:Class, jejíž význam, datová sada i URI zůstává stejné. Nové vazby navržené v datovém modelu dle Kiráľa (2017) budou zrevidovány a poté budou do datového modelu SPOI implementovány.

Jako další standardizované atributy vhodné pro doplnění do datového modelu SPOI jsou dcterms:creator, dcterms:contributor, dcterms:license a dcterms:modified. Všechny atributy jsou přebírány z dokumentu „DCMI Metadata Terms“ (DCMI Usage Board 2012a) a jejich jmenným prostorem je tedy <http://purl.org/dc/terms/1.1/>.

Popis jednotlivých vhodných atributů pro doplnění:

- **dcterms:creator** – jméno toho, kdo body zaevidoval – autor (URI),
- **dcterms:contributor** – jméno toho, kdo přispěl (doplnil informace) k bodu – editor (URI),
- **dcterms:license** – originální (původní) licence zájmového bodu – originální licence (URI),
- **dcterms:modified** – datum poslední editace či aktualizace – datum změny (datum).

Atribut autor jakožto „sběrač“, který body sbírá, získává a eviduje, a editor jako ten, kdo body opravuje či aktualizuje, mají význam z důvodu znalosti uživatelů, kteří se podílejí na obohacování databáze SPOI, a pro jejich možné kontaktování. V případě, že by tyto dvě položky byly povinnými atributy, mohlo by to odradit eventuální zneužití, tj. úmyslné zanesení nesprávných informací do databáze SPOI. Systém registrace autorů, resp. editorů je z hlediska toho, že databáze SPOI je open databáze, zbytečný. Nicméně, by se tyto atributy měly vyplňovat a to pouze odkazy na oficiální profily (např. OpenID⁹⁷, LinkedIn⁹⁸, pracovní profily atd.). Atribut originální licence slouží pro doplnění odkazu na licenci dle původního zdroje bodu.

Metadatová položka datum změny je vhodná pro data (body), která se opravují či aktualizují, aby bylo pro ostatní uživatele zřejmé, kdy byla naposledy provedena změna a tedy, jak staré, resp. nové jsou údaje u těchto dat (bodů). Lze toho využít například při rozhodování, která data kontrolovat. Další využití (jak atributu datum změny, tak editor) je u samotné aktualizace a opravy bodu (více v kapitolách 2.1.2., 2.2. a 2.3.2.). Informace o autorovi bodu je využitelná přímo u sběru a získávání dat.

⁹⁷ <http://openid.net/>

⁹⁸ <https://cz.linkedin.com/>

1.2.5. Harmonizační procesy

Vstupní formáty zdrojových dat, které musí být v konečné fázi sjednocené, jsou různé – od jakéhokoliv textu, přes tabulky (CSV, XLS), geografické formáty (SHP [ESRI Shapefile Technical Description 1998], GML, KML, GeoJSON [Butler et al. 2016]), JSON (Bray 2015), XML, OWL až po formát OSM binary⁹⁹. Protože je sbíraných dat velké množství a zdroje jsou velmi heterogenní (rozmanité), je nutné sesbíraná a získaná data pro další práci upravit, tj. sjednotit. Cílem je tedy vytvořit datovou sadu, která bude harmonizovaná a propojená.

Pro harmonizaci se využívá 5krokový harmonizační rámec – framework (Janecka et al. 2013). Tento proces se skládá ze dvou fází, které mají jednotlivé dílčí kroky:

- kognitivní fáze:
 - Harmonizační teorie,
 - Znalost vstupních dat,
 - Znalost výstupních dat,
 - Návrh a vývoj jednotlivých harmonizačních kroků,
- technická fáze:
 - Praktická realizace harmonizace dat.

Krok 1, **Harmonizační teorie používaná pro SPOI**, je založený na 3 zdrojích, kterými jsou:

- dokumenty zaměřené na publikování Linked (Open) Data (např. Bizer et al. 2008; Bizer et al. 2009; Bauer a Kaltenböck 2011),
- texty zabývající se Linked Data s ohledem na prostorová data (např. Auer et al. 2009; Kuhn et al. 2014),
- výzkumy spojené s modelováním a harmonizací prostorových dat (např. Goodchild 1992; Shekhar et al. 1997; Longley et al. 2001; Janecka et al. 2013).

V kroku 2, **Znalost vstupních dat**, jsou získávány podklady pro analýzu a „čištění“ heterogenních dat – je tedy nutné nalézt informace, které jsou důležité z pohledu integrace dat, aby byla práce s nimi co nejjednodušší a nejefektivnější.

Krok 3, **Znalost výstupních dat**, se zabývá specifikací výstupních dat. Mimo požadavku na Linked Data jsou požadavky na data dány především datovým modelem (které informace budou či nebudou poskytnuty). Data (body) SPOI využívají tyto slovníky:

⁹⁹ https://wiki.openstreetmap.org/wiki/WhereAmI/OSM_Binary_Format

- Web Ontology Language – OWL,
- Simple Knowledge Organization System – SKOS,
- Friend of a Friend – FOAF,
- GeoSPARQL,
- Dublin Core Metadata Initiative,
- ISA Programme Location Vocabulary.

Krok 4, **Návrh a vývoj jednotlivých harmonizačních kroků**, se konkrétně věnuje modelování dat, které obnáší mapování mezi atributy ve zdrojových a cílových datech, spojení se slovníky a dalšími sémantickými zdroji a také konverzí do formátu RDF. Data SPOI (body) jsou propojena s externími datovými zdroji (DBpedia, LinkedGeoData¹⁰⁰, GeoNames.org a Wikidata), publikována pod ODbL licencí (viz kapitola 1.1.) a jsou dostupná v mapovém klientu nebo pomocí stahovací služby SPARQL endpoint¹⁰¹ (koncový bod, viz kapitola 2.1.1.).

Poslední krok 5, **Praktická realizace harmonizace dat**, spočívá v tom, že se pomocí skriptu spustí navržené harmonizační procesy v kroku 4. (Čerba a Mildorf 2016) Zjednodušeně, dle současného harmonizačního procesu¹⁰², lze postupy harmonizace zapsat takto:

- stahování a rozbalení, vytváření dočasných souborů – týká se dat, která (zatím) nejsou ve formátu a stavu potřebném pro další zpracování (např. ZIP¹⁰³),
- přepis do strukturovaného formátu – týká se nestrukturovaných dat (např. text do tabulky – CSV),
- úpravy a kompletování dat,
- změna kódování znaků či souřadnicového systému – pro data v jiném souřadnicovém systému než WGS 84¹⁰⁴ nebo s jiným kódováním než je UTF-8¹⁰⁵,
- filtrování dat a atributů,
- přidávání odkazů a metadat, reklasifikace,
- transformace do datového modelu SPOI a export do formátu RDF – týká se všech dat¹⁰⁶.

¹⁰⁰ <http://linkedgeo.org/About>

¹⁰¹ <https://www.w3.org/wiki/SparqlEndpoints>

¹⁰² http://sdi4apps.eu/spoi/doc/Data_harmonization.svg

¹⁰³ <https://www.iana.org/assignments/media-types/application/zip>

¹⁰⁴ http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html

¹⁰⁵ <https://tools.ietf.org/html/rfc3629>

¹⁰⁶ <https://drive.google.com/file/d/0B4P8WefNVti-LUQ3cnhKVW13VjA/view>

Jelikož harmonizační proces není jednoduchou záležitostí, je tedy vhodné, aby nově sbíraná či získávaná data měla všechny náležitosti (viz kapitola 2.1.2. a 2.2.) a vstupní formát byl nejlépe takový, který je snadno upravitelný a převeditelný do databáze SPOI (viz kapitola 2.2.).

2. Metodika

Databáze Smart Points of Interest (SPOI) je databáze používaná v projektu Peregrinus Silva Bohemica¹⁰⁷, která byla vytvořena v rámci projektu SDI4Apps¹⁰⁸, jež byl spravován Západočeskou univerzitou v Plzni¹⁰⁹. Databáze SPOI obsahuje jednotlivé body zájmu POI (Point of Interest), což jsou konkrétní místa (body), která mohou být pro někoho zajímavá nebo užitečná. Tyto body jsou publikovány jako 5* Linked Open Data^{110 111} a v mapě je lze vizualizovat spolu s podkladovou mapovou vrstvou. Linked Data¹¹² znamená, že mezi jednotlivými body jsou vazby a jsou také propojeny s externími zdroji (databázemi). Open Data¹¹³ znamená, že je databáze otevřená, tj. volně dostupná pod licencí ODbL¹¹⁴, jež umožňuje data editovat, vkládat, sdílet a používat za podmínky poskytnutí stejné licence ostatním uživatelům. Databáze SPOI je vytvářena kombinací volně dostupných globálních a lokálních dat (více viz první kapitola).

Metodika pro doplňování bodů do databáze Smart Points of Interest (dále jen metodika) je určena pro každého, kdo by chtěl přispívat a doplňovat body do databáze SPOI. Její cílovou skupinou je tedy především laická veřejnost, která by se chtěla zapojit a být nápomocná při rozšiřování databáze. Předpokládá se, že veřejnost, která tuto metodiku bude používat (číst), není zatím do sběru či projektu nijak výrazně zainteresována, tj. zatím neví určité (všechny) informace o tom, jak body zájmu sbírat, získávat a obohacovat tak databázi. Mimo široké veřejnosti (obyvatelé měst a obcí) se konkrétně jedná například o pracovníky informačních center, městských úřadů nebo poskytovatele služeb (dále jen uživatelé).

Níže popsané jednotlivé kroky sběru a získávání dat udávají, jak postupovat a čím se řídit při doplňování do databáze (diagram postupu doplňování databáze SPOI lze vidět na Obr. 3). Je zde popsán jak způsob sběru a získávání dat, tak jejich evidence a jak tato data kontrolovat a aktualizovat.

¹⁰⁷ <https://kgm.zcu.cz/aktualni-projekty/peregrinus/>

¹⁰⁸ <http://sdi4apps.eu/project-information/>

¹⁰⁹ <http://www.zcu.cz/>

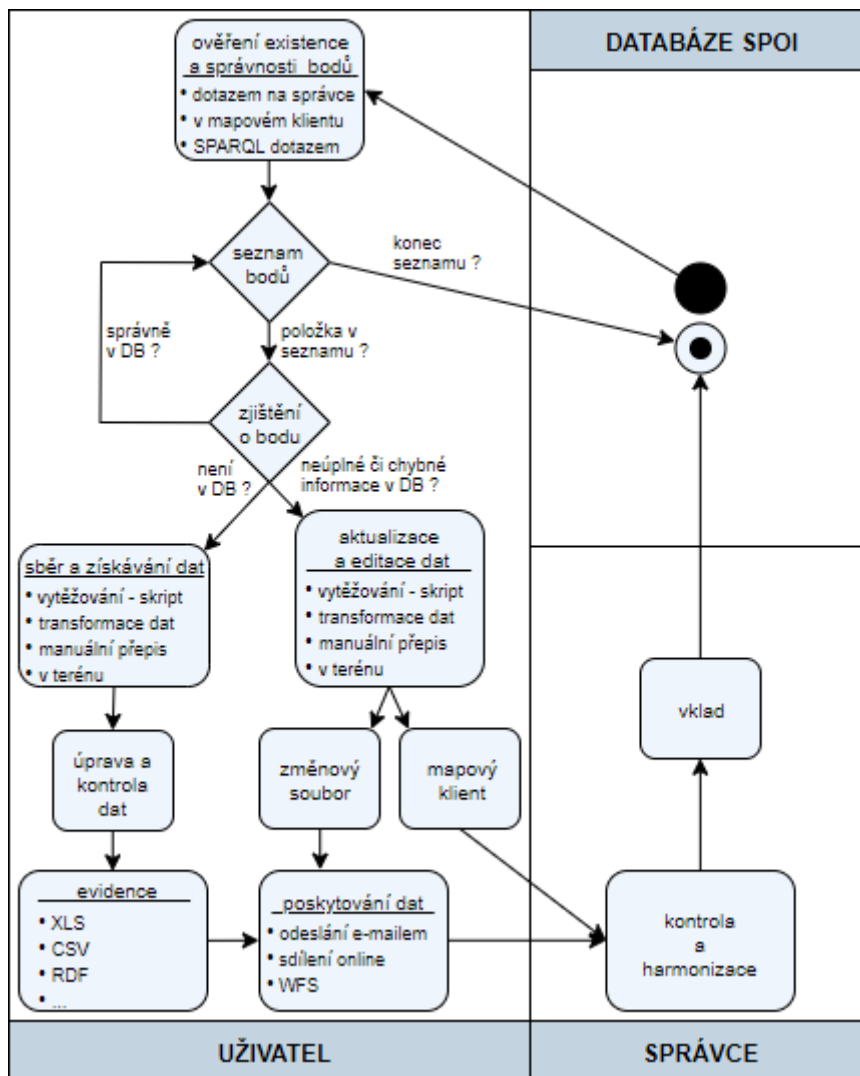
¹¹⁰ <https://www.w3.org/DesignIssues/LinkedData.html>

¹¹¹ <https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html#linked-open-data>

¹¹² <https://www.w3.org/standards/semanticweb/data>

¹¹³ <http://opendatahandbook.org/guide/en/what-is-open-data/>

¹¹⁴ <https://opendatacommons.org/licenses/odbl>



Obr. 3: Diagram postupu doplňování databáze SPOI

Zdroj: vlastní

2.1. Způsob získávání dat – postup sběru informací

2.1.1. Ověřování

Je důležité si uvědomit a vybrat, jaká data nebo místa zájmu mají být sbírána. Po výběru lokality či zájmových bodů/míst si musí „sběrač“ zjistit, zda jím vybrané body a místa pro sběr nebo celé lokality již nejsou v databázi SPOI zaneseny, tj. sesbírány a zaevidovány někým jiným. Ověření zanesení bodů v databázi SPOI lze provést několika způsoby, které by se daly rozdělit dle zkušeností „sběrače“. Procházení databáze a dat v ní je možné skrze:

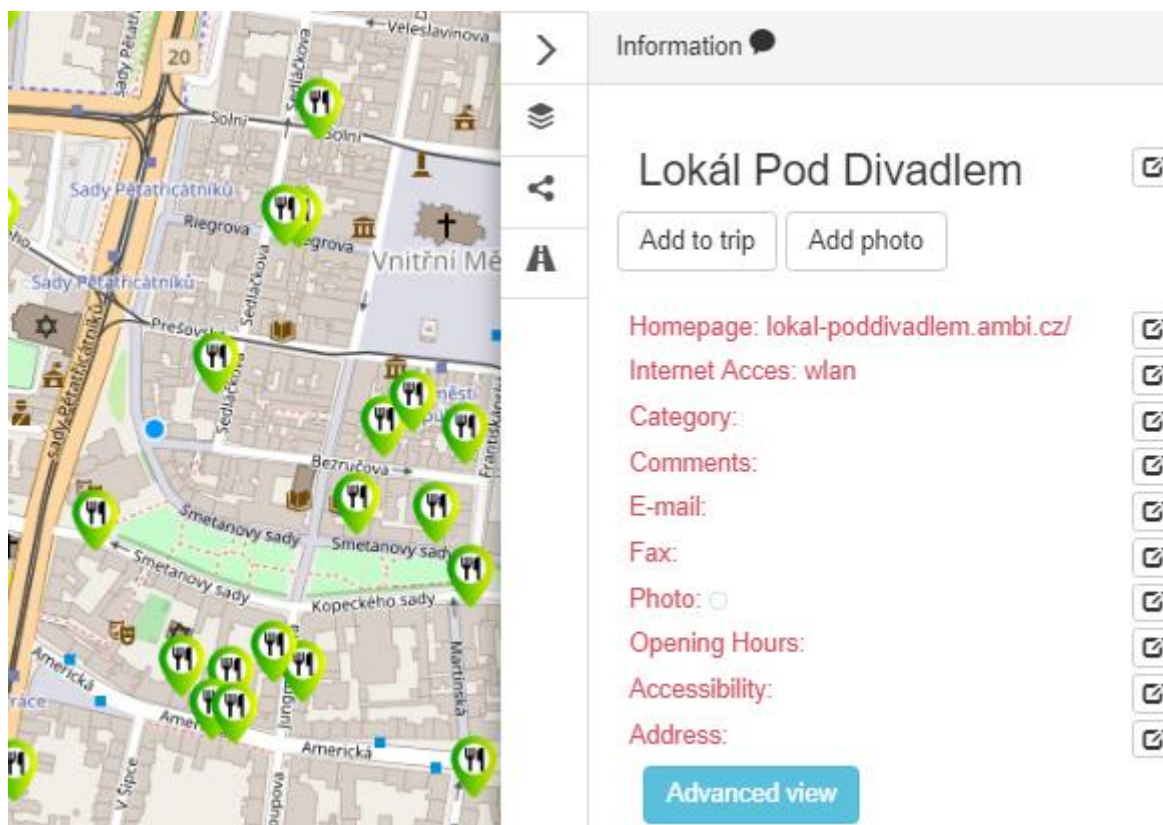
- správce,
- mapové okno (klient),
- SPARQL dotazy.

Kontaktovat správce¹¹⁵ je sice velmi jednoduchá forma procházení databáze a zjišťování, která data (body zájmu) jsou v databázi SPOI již zanesena, nicméně je ale také velmi nevhodná. Je to především z důvodu pracovní kapacity správce, jelikož není v jeho silách ani možnostech ověřovat veškeré požadavky, lokality a body, které by mu uživatelé zasílali v dotazech.

Vizuální procházení databáze skrze mapového klienta¹¹⁶ je možností, jak snadno a rychle zjistit, zda bod zájmu v databázi SPOI již je nebo není. Problém nastává v případě, kdy se jedná o větší oblast (lokalitu), ve které se mají body ověřovat – prohledávání se poté stane časově velmi náročným a také více náchylným k chybám. Ukázka vyhledávání v mapovém okně je vidět na Obr. 4. Jsou zde zobrazeny pouze body zájmu vedené v kategorii restaurace, po kliknutí na znak se objeví informace o dané restauraci (bodů zájmu Lokál Pod Divadlem) a místo znaku se objeví modré kolečko.

¹¹⁵ Veškeré kontaktní informace na správce jsou dostupné na webové stránce projektu SPOI – <http://sdi4apps.eu/spoi/>.

¹¹⁶ <http://ng.hslayers.org/examples/geosparql/>



Obr. 4: Snímek mapového okna s vybraným bodem zájmu (Lokál Pod Divadlem)
Zdroj: Mapový klient

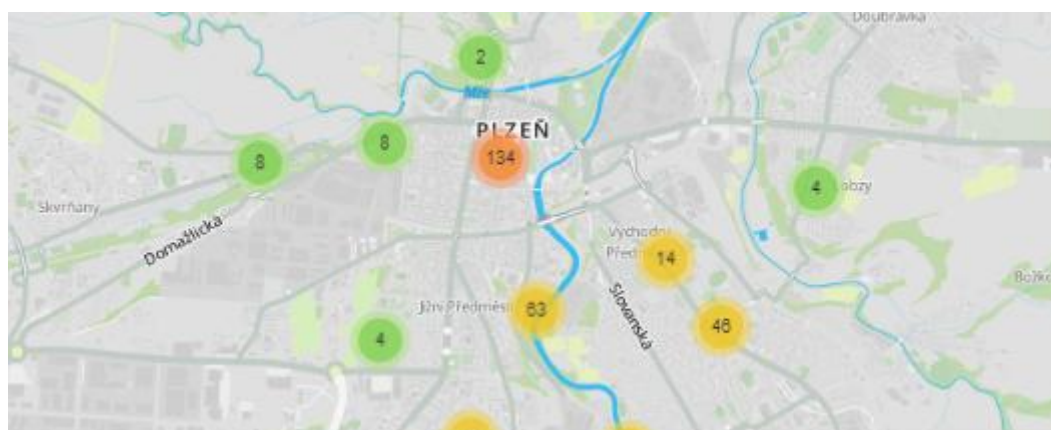
Pro vyhledávání a zobrazování bodů POI je vhodný také turistický průvodce Smart tourism guide¹¹⁷ (Obr. 5 a 6), který vytvořili Bc. Jan Macura a Ing. Jáchym Kellar s pomocí přispěvatelů serveru GitHub¹¹⁸. Průvodce umožňuje vyhledávat body zájmu pomocí lokace jak názvem, tak souřadnicemi.

¹¹⁷ <http://portal.sdi4apps.eu/guide/>

¹¹⁸ <https://github.com/>



Obr. 5: Snímek hlavní stránky Smart tourism guide
Zdroj: Smart tourism guide



Results for 49.7475 13.3776 3 km

Error obtaining weather forecast: parsererror

Name	Category (Filter)
Lezecká Stěna Hannah	Culture And Entertainment, Climbing Wall
Climbing	Culture And Entertainment, Climbing Wall
Poliklinika	Hospital, Professional And Public

Obr. 6: Snímek Smart tourism guide po vyhledání POI bodů v městě Plzni
Zdroj: Smart tourism guide

Nejobtížnější z hlediska náročnosti a pochopení je procházení a ověřování databáze SPOI pomocí SPARQL dotazů. Na druhou stranu je to nejvhodnější a nejkompaktnější možnost pro mnoho různých požadavků (dotazů) a také jejich kombinace.

Dotaz SPARQL

SPARQL (Harris a Seaborne 2013) je sémantický dotazovací jazyk pro data uchovávaná ve formátu RDF a je standardem konsorcia W3C. Jazyk je podrobně popsán právě na jím se zabývajících webových stránkách konsorcia W3C¹¹⁹, kde lze také nalézt, jakou strukturu a význam mají dotazy a jak je psát. Samotná struktura dotazu je pak velmi podobná jazyku SQL (ISO/IEC 9075-1:2011(E) 2011).

Data SPOI jsou uložena v prostředí Virtuoso¹²⁰ jako RDF trojice a jsou publikována (možnost dotazování a stahování dat) skrze SPARQL endpoint¹²¹. SPARQL endpoint je adresa URI, pod kterou patří datová kolekce RDF, což je množina dokumentů RDF, nad kterou se jednotlivé dotazy provádí. Trojice je základní informační prvek v RDF skládající se z podmětu (popisovaný bod), vlastnosti (atributu) a předmětu (hodnota atributu, více v kapitole 1.2.3.). Z tohoto důvodu se SPARQL dotaz skládá z množiny trojic, kde každý prvek trojice může být proměnná. Během dotazování jde o to najít shodné trojice v databázi a tím výsledky k proměnným v dotazu (např. bod, který se nachází v Plzni).

Dotazování se provádí pomocí napsání dotazu na webových stránkách nástroje Virtuoso SPARQL Query Editor¹²² (Obr. 7), kde je odeslán a poté přijímán jeho výsledek. Na této stránce je okno (s předdefinovaným dotazem pro výběr všech atributů použitých v databázi SPOI), do kterého je zadáván SPARQL dotaz na vyhledání dat a ověření, zda jsou již v databázi zaneseny či nikoliv, jaké informace k nim jsou vedeny atd. Výstupem může být JSON, XML, RDF apod., které lze uložit, ale nejvhodnějším (nejvíce přehledným) pro běžného uživatele je výstup ve formě tabulky (Obr. 8) na nové HTML stránce, jež je také výchozí výstup, který lze změnit pod oknem. Dále lze pod editorem nastavit, jaká data budou pro dotaz použita (nastaveno na lokální data), časový limit na provedení dotazu a další možnosti nastavení.

¹¹⁹ <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>

¹²⁰ <https://virtuoso.openlinksw.com/>

¹²¹ <https://www.w3.org/wiki/SparqlEndpoints>

¹²² <http://data.plan4all.eu/sparql>

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where {[] a ?Concept} LIMIT 100
```

Sponging:

Results Format:

Execution timeout: milliseconds (values less than 1000 are ignored)

Options: Strict checking of void variables Log debug info at the end of output (has no effect on some queries and output formats)

(The result can only be sent back to browser, not saved on the server, see [details](#))

Obr. 7: Snímek dotazovacího okna Virtuoso SPARQL Query Editor
zdroj: Virtuoso SPARQL Query Editor

popis	souradnice
"Zámek Štábeň"	"POINT(14.0031393 49.2673768)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Doma"	"POINT(14.5149494 49.9347371)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"zámek Vokšice"	"POINT(15.3228225 50.426977)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Zámek Rudolfovo (nepř.)"	"POINT(14.5360155 48.9927915)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"zámek"	"POINT(14.4164475 50.7985427)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Salhausenský zámek"	"POINT(14.4127594 50.7999174)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Felsenburg Falkenstein"@de	"POINT(14.4053986 50.8548853)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Hrad Falkenštejn"	"POINT(14.4053986 50.8548853)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Bělá pod Bezdězem"	"POINT(14.8078326 50.5013433)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Weißwasser"@de	"POINT(14.8078326 50.5013433)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Jevín"	"POINT(14.3339527 50.3453396)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Prácheň"	"POINT(13.6814964 49.3158933)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Cimburk"	"POINT(16.7328755 49.7087986)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Jaroměřice"	"POINT(16.7543736 49.6270561)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>
"Zámek Doksy"	"POINT(14.6526163 50.5648038)"^^<http://www.openlinksw.com/schemas/virtrdf#Geometry>

Obr. 8: Snímek výstupu dotazu ve Virtuoso SPARQL Query Editor – tabulka
zdroj: Virtuoso SPARQL Query Editor

Příklady dotazů na databázi SPOI

Zelenou barvou jsou části, které lze změnit dle uživatelského zájmu a tím dostat nový dotaz.

1. Body v České Republice spadající do třídy 'castle':

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX geo:  <http://franz.com/ns/allegrograph/3.0/geospatial/>
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
PREFIX poi:  <http://www.openvoc.eu/poi#>
SELECT DISTINCT ?bod ?navez ?souradnice
WHERE
{
  ?bod geos:sfWithin <http://dbpedia.org/resource/Czech_Republic> .
  ?bod poi:class <http://gis.zcu.cz/SPOI/Ontology#castle> .
  ?bod rdfs:label ?navez .
  ?bod geos:asWKT ?souradnice
}
ORDER BY 2
LIMIT 100
```

2. Všechny body v České Republice, které mají v adrese slovo 'Černohorská':

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
SELECT DISTINCT ?bod ?adresa
WHERE
{
  ?bod geos:sfWithin <http://dbpedia.org/resource/Czech_Republic> .
  ?bod locn:fullAddress ?adresa .
  FILTER regex(?adresa, 'Černohorská', 'i')
}
LIMIT 100
```

3. Body ve vzdálenosti do 5 km od bodu (13.7977, 49.6284) a jejich souřadnice a název:

Tento dotaz nefunguje zcela správně – nelze zaručit, že vypíše všechny body v okruhu 5 km od vybraného bodu. Důvodem je nejspíše velké množství bodů, které se prohledává (i v případě omezení oblasti výběru na Českou republiku). Počet vyhledaných bodů závisí na hodnotě, do jaké vzdálenosti se mají hledat body průniku (`bif:st_intersects`), a také na hodnotě, do jaké vzdálenosti mají být body hledány od daných souřadnic

(bif:st_distance). V případě zadání nízkých hodnot u obou těchto prostorových dotazů, je velmi pravděpodobné, že bude nalezeno minimální množství záznamů a jen od určité vzdálenosti od souřadnic anebo nebude nalezen žádný záznam. Kolik bodů je vyhledáno, závisí na tom, jaké hodnoty pro prostorové dotazy jsou použity (v tomto případě je vyhledáno pouze 10 záznamů od 2 do 5 km od souřadnic, při zvýšení hodnoty vzdálenosti průniku na 50 km je nalezeno 22 bodů v okruhu 1–5 km od souřadnic). Pokud se první širší výběr bodů (průnik) provede ještě ve větším okruhu, je nalezeno mnohem více záznamů a zároveň mnohem více bližších záznamů daným souřadnicím. Na druhou stranu nelze toto číslo (vzdálenost průniku) zvětšovat libovolně, od určité hodnoty již dotaz nevyhledá žádné body (v tomto případě je to např. číslo 300), pokud je ale poté zadáno číslo o něco menší a pak opět větší, funguje to. Při zadání ještě vyšší hodnoty průniku (v tomto případě 346 a vyšší) se dotaz vůbec neprovede a je hlášena chyba: „Virtuoso S0002 Error SQ200: No table rdf_obj“. Mezní hodnota vzdálenosti průniku se liší v závislosti na zadaných souřadnicích a prohledávané vzdálenosti od nich (velikost okruhu okolo bodu). Dalším negativem je, že při zadání vysoké hodnoty průniku, ale v mezích, kdy není hlášena chyba (zde do 345), je nalezeno velké množství záznamů, přičemž se v nich ale vyskytuje mnoho duplicit z hlediska souřadnic (odkaz i název bodu je u některých jiný), které nelze odstranit ani použitím vnořených dotazů s omezením pro duplicitní záznamy (DISTINCT).

```
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT ?bod (bif:st_distance (?souradnice, bif:st_point
(13.7977, 49.6284))) AS ?vzdalenost ?navez ?souradnice
WHERE
{
  ?bod rdfs:label ?navez .
  ?bod geos:asWKT ?souradnice .
  FILTER (bif:st_intersects (?souradnice, bif:st_point (13.7977,
49.6284), 15)) .
  FILTER (bif:st_distance (?souradnice, bif:st_point (13.7977,
49.6284)) < 5)
}
ORDER BY ASC 2
```

Byly zkoušeny také prostorové dotazy pomocí vyhledávání bodů funkcemi BOX či POLYGON (cca 10 km od daných souřadnic), ale bohužel nebyl nalezen žádný záznam. V příkladu je

použit prostorový dotaz pro vyhledávání pomocí funkce BOX a je zde „zakomentovaný“ (pomocí znaku #) prostorový dotaz pomocí funkce POLYGON.

```
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema>
SELECT DISTINCT ?bod (bif:st_distance (?souradnice, bif:st_point
(13.7977, 49.6284))) AS ?vzdalenost ?nazev ?souradnice
WHERE
{
  ?bod geos:sfWithin <http://dbpedia.org/resource/Czech_Republic> .
  ?bod geos:asWKT ?souradnice .
  ?bod rdfs:label ?nazev .
  FILTER (bif:st_intersects (?souradnice , bif:st_geomfromtext(
"BOX(13.707868 49.538421, 13.887532 49.718379)"))
  #FILTER (bif:st_intersects (?souradnice , bif:st_geomfromtext(
"POLYGON((13.707868 49.538421, 13.707868 49.718379, 13.887532
49.718379, 13.887532 49.538421, 13.707868 49.538421)"))))
}
ORDER BY ASC 2
```

2.1.2. Sběr a získávání

Po ověření dat a zjištění, že určité body či lokality v databázi SPOI ještě nejsou zaneseny, přichází na řadu sběr a získávání nových dat. Body, které v databázi jsou zaevidované, ale mají chybné, zastaralé či chybějící informace, je nutné opravit a aktualizovat (více v kapitole 2.3.). Data pro zanesení do databáze SPOI lze získat jak na Internetu, tak sběrem bodů v terénu.

Data na Internetu

Data a informace na Internetu lze hledat procházením webových stránek obce či města a databází firem, služeb, institucí, památek a jiných zajímavostí. Další možností jsou mapové služby a portály s již zanesenými body, ale s odpovídající licencí. Obecně lze hledání dat (bodů a informací) do databáze SPOI na Internetu a poté jejich zaevidování (soubor s uloženými informacemi a daty – viz následující kapitola) provést několika způsoby. Jsou jimi:

- vytěžování pomocí skriptu,
- transformace dat,
- manuální (ruční) přepis dat.

Ovšem ne všechny výše uvedené možnosti jsou vhodné pro laickou veřejnost, které je tato metodika určena. Jedná se konkrétně o první možnost – vytěžování pomocí skriptu, které spočívá v napsání kódu, kterým by se data automaticky přepisovala a stahovala ze zdrojových HTML stránek. Tento způsob je možné využít v případě, že se jedná o HTML se strukturovaným textem či tabulkou. Pokud by uživatel neznalý této problematiky stál o postup vytěžování pomocí skriptu, jedinou možností je objednat si odbornou firmu.

Druhý způsob pomocí transformace dat lze využít pouze tehdy, jsou-li zdrojová data uchovávána v souboru, který je také strukturovaný – tedy, že se nejedná o prostý text, ale jsou vedeny např. v tabulkách (CSV soubor). Taková data lze transformovat (konvertovat) pomocí nejrůznějších programů a nástrojů, které jsou volně přístupné, bez poplatku a online, např. do formátů JSON, XML, GeoJSON atd. Jedním z těchto programů (aplikací) je online konvertor [Convertcsv.com](http://www.convertcsv.com/)¹²³, který nabízí mnoho formátů a dalšího nastavení (Obr. 9) pro transformaci menšího objemu dat.

¹²³ <http://www.convertcsv.com/>

Step 1: Select your input

Option 1 - Choose a CSV/Excel file Soubor nevybrán Encoding

Option 2 - Enter an URL

Option 3 - paste into Text Box below

Enter or paste CSV here

Input Records- Header: false
Data: Separator: , Fields: 0 Records: 0

Step 2: Choose input options (optional) ▼

Step 3: Choose output options (optional) ▼

Step 4: Generate output

Result Data:

Output Results

Save your result: EOL:

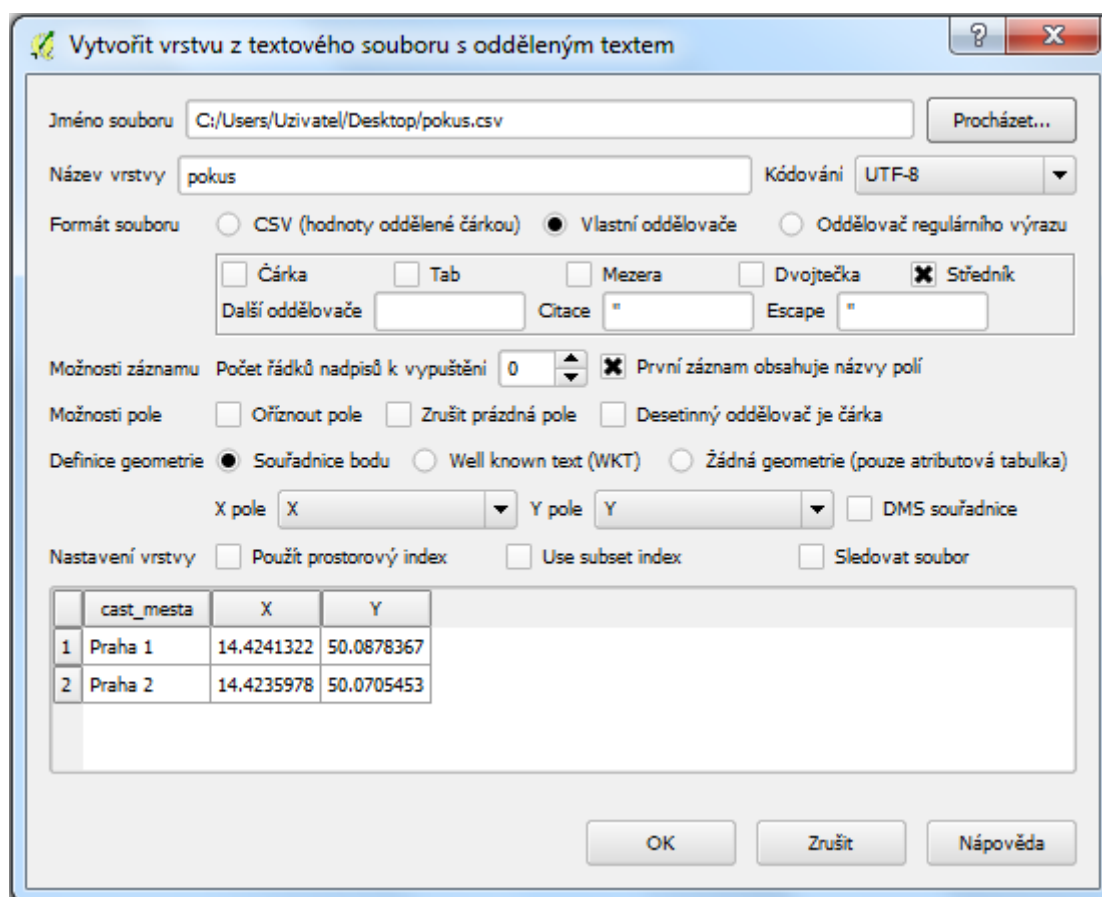
Obr. 9: Snímek z konvertoru Convertcsv.com (CSV – GeoJSON)

Zdroj: Convertcsv.com

Pro velký objem dat je vhodnější využít konverzi dat pomocí open source GIS (Geographic Information System) programů či jiných open source offline programů, tj. programů, které lze volně stáhnout a používat na desktopovém (stolním) počítači či notebooku. Open source GIS je volně dostupný a stažitelný geografický informační systém, což je program, který umožňuje ukládat, spravovat a analyzovat prostorová data. Takovým programem je například QGIS (neboli Quantum GIS)¹²⁴, ve kterém lze mimo prohlížení, tvorby a editace rastrových a vektorových vrstev, zpracování GPS dat, tvorby mapových výstupů atd. také konvertovat

¹²⁴ <https://qgis.org/en/site/>

formáty souborů. Vhodné je pomocí programu QGIS konvertovat např. soubor CSV do formátu SHP, který je využitelný pro publikování dat skrze stahovací služby (více v následující kapitole). Konverze formátu CSV je řešena pomocí vytvoření nové vrstvy, která se poté uloží v požadovaném formátu (vytvoření nové vrstvy ze souboru CSV na Obr. 10). Program ke stažení¹²⁵ a manuál¹²⁶ je dostupný na webových stránkách QGISu, mnoho dalších materiálů ve formě návodů je na jiných webových stránkách, taktéž existuje velké množství výukových videí (například také s postupem pro konverzi z formátu CSV¹²⁷), které jsou dostupné na stránkách YouTube.com¹²⁸.



Obr. 10: Vytvoření nové vrstvy ze souboru CSV v programu QGIS

Zdroj: vlastní

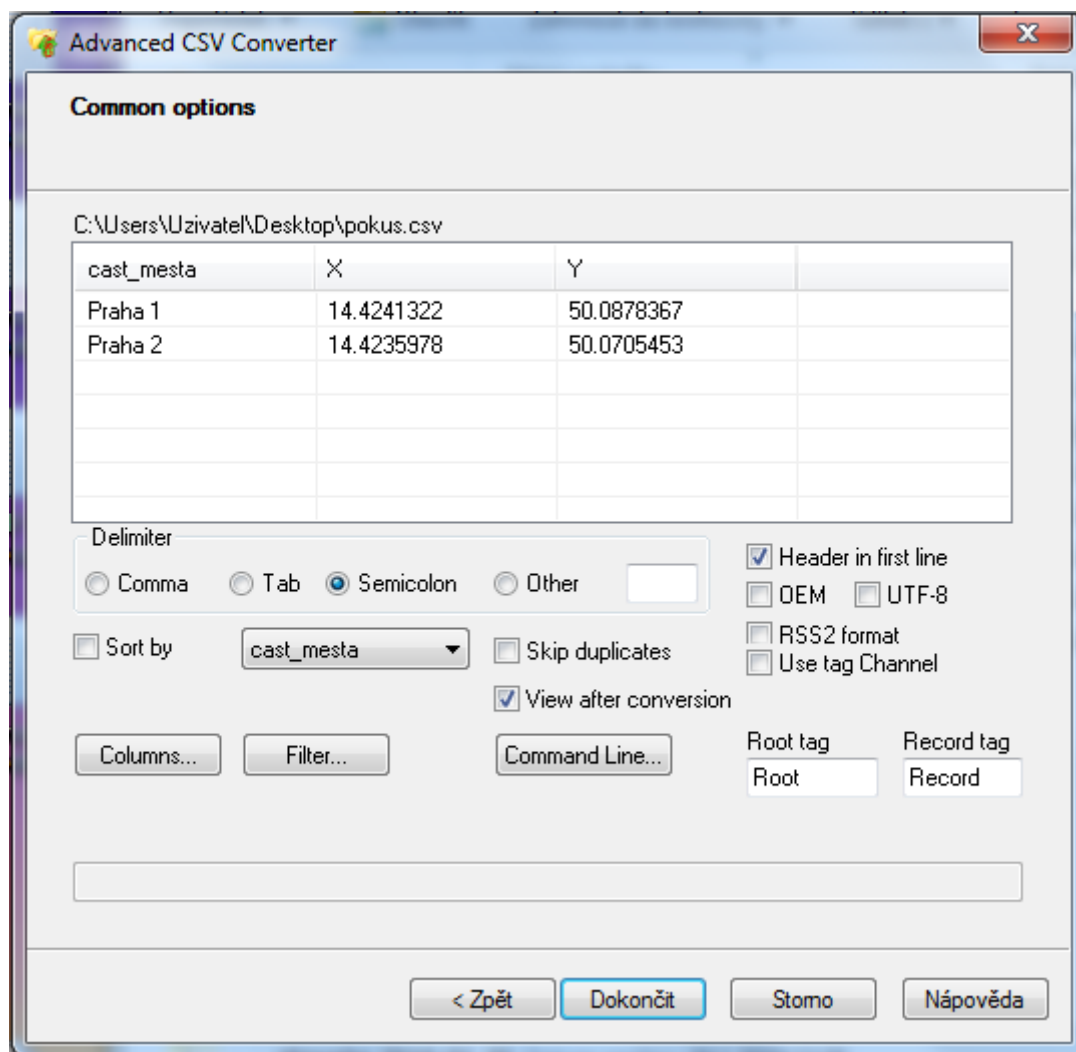
¹²⁵ <https://www.qgis.org/en/site/forusers/download.html>

¹²⁶ https://docs.qgis.org/2.18/en/docs/user_manual/

¹²⁷ <https://www.youtube.com/watch?v=k9M6hbHXHrI>

¹²⁸ <https://www.youtube.com/>

Příkladem jiného offline programu pro konverzi formátu CSV je program CSV Converter¹²⁹, který převede soubor na jeden z více než 10 formátů a umožňuje i vybrat základní nastavení vlastností (Obr. 11).

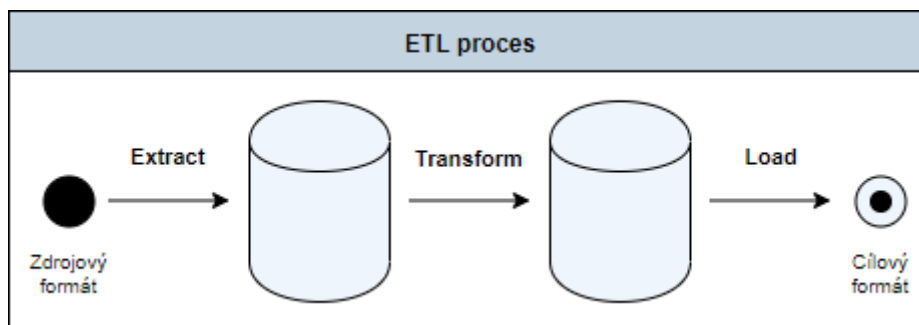


Obr. 11: Konverze souboru CSV v programu CSV Converter
Zdroj: vlastní

Pro konverzi jednoho formátu do druhého lze využít také ETL nástroje, kde ETL znamená Extract-Transform-Load – tři procesy sloučené do jednoho nástroje (Obr. 12). Nástroj načte data (soubor v určitém formátu), transformuje (konverzuje) je a uloží je do nového formátu. Tento nástroj, jakožto program, si může uživatel buď sám graficky navrhnout a poté vygenerovat spustitelný kód anebo použít nástroj dostupný na trhu, který je většinou zabudovaný jako balík s dalšími nástroji a prostředky v rámci určitého programu – jedná

¹²⁹ <http://www.dbf2002.com/csv-converter/>

se např. o produkty společnosti IBM¹³⁰, Oracle¹³¹, Microsoft¹³² atd. Více o ETL procesech a nástrojích se lze dočíst v bakalářské práci „ETL procesy“ (Husáriková 2006).



Obr. 12: ETL proces
Zdroj: vlastní

Jedním z výše uvedených způsobů transformované soubory s daty lze poté zobrazit a vizualizovat např. v online editoru pro mapová data Mapshaper¹³³ nebo dále propojovat s jinými daty (např. mapový podklad, další tabulky,...). Vizualizace adresních míst vedených v RÚIAN (černé čtverečky, stažené ve formátu CSV a dále převedené pomocí konvertoru do formátu GeoJSON) spolu s katastrální mapou (zeleně a pouze hranice parcel bez vnitřní kresby, stažená ve formátu SHP na stránkách ČÚZK¹³⁴) je na Obr. 13 (oba soubory byly staženy pro obec Hlubyně). Informace vedené u jednoho bodu, tj. adresního místa (číslo popisné 1 zobrazené červenou barvou) ze souboru CSV lze vidět na Obr. 14. Offline program, ve kterém lze data zobrazit a dále s nimi pracovat, je např. program QGIS.

¹³⁰ <https://www.ibm.com/cz-cs/>

¹³¹ <https://www.oracle.com/cz/index.html>

¹³² <https://www.microsoft.com/cs-cz>

¹³³ <http://mapshaper.org/>

¹³⁴ <http://services.cuzk.cz/shp/ku/epsg-5514/>



Obr. 13: Snímek vizualizovaných dat v editoru Mapshaper
Zdroj: vlastní

Kód ADM	12903019
Kód obce	564605
Název obce	Hlubyně
Kód MOMC	..
Název MOMC	..
Kód MOP	..
Název MOP	..
Kód části obce	39705
Název části obce	Hlubyně
Kód ulice	..
Název ulice	..
Typ SO	č.p.
Číslo domovní	1
Číslo orientační	..
Znak čísla orientačního	..
PSČ	26272
Platí Od	2013-12-04T00:00:00

Obr. 14: Snímek informací zobrazených v editoru Mapshaper
Zdroj: vlastní

Nejjednodušší, avšak časově nejnáročnější, je třetí způsob – manuální (ruční) přepis dat. Tuto možnost má každý uživatel bez nutnosti jakékoliv znalosti skriptovacích jazyků, kódování nebo formátů uložení dat. Daní je ovšem časová náročnost v případě objemných dat, tj. velkého množství bodů a záznamů k nim.

Data z terénu

Posledním způsobem, jak doplňovat data do databáze SPOI, je sběr bodů v terénu, což znamená sbírat data (body) pomocí mobilní aplikace (nebo GPS přístroje v případě jiného [ručního] zapisování informací, tj. použití pouze pro získání souřadnic bodu). Z ní je poté nutné exportovat (uložit) data v určitém formátu a dále je konvertovat, upravovat a vkládat do databáze SPOI – posílat, sdílet či poskytnout ke stažení správci (viz následující kapitola). K tomu je již zapotřebí mobilní zařízení s připojením k Internetu a staženou vhodnou aplikací s GPS lokátorem. Jak vybrat vhodnou mobilní aplikaci pro účely sběru dat (bodů), lze vyčíst v metodice „Nejvhodnější aplikace pro získání souřadnic zájmových bodů pro databázi Smart Points of Interest“ (Karasová 2018).

Doplňování atributů

Povinnými sbíranými informacemi u všech případů sběru a získávání dat, tj. atributy pro body zájmu do databáze SPOI (všechny atributy viz kapitola 1.2.4.), které doplňuje (vyplňuje) „sběrač“, jsou:

- **název** (rdfs:label) – oficiální jméno bodu, z názvu by mělo být vždy jasné, o jaký bod zájmu se jedná (druh památky, servis, kavárna, škola,...),
- **souřadnice** (geos:asWKT) – rozděleny na **zeměpisnou délku** a **šířku** bodu,
- **typ objektu** (poi:class, nově owl:Class) – kategorie, základní třída vycházející ze SPOI ontologie – anglický ekvivalent k základnímu názvu bodu v takové podobě, aby z něj bylo jasné, o co se jedná (kaple – chapel, kostel – church, restaurace – restaurant, lékárna – pharmacy,...), zapisuje se tak, jak je uveden v ontologii,
- **umístění** (geos:sfWithin) – rozděleno na **stát**, **kraj**, **okres** a **obec** – vyplňuje se zde odkaz na lokalitu (oblast), kde se bod zájmu nachází – data GeoNames.org a DBpedia,
- **zdroj** (dc:source) – odkaz na původní zdroj (v případě, že se nejedná o body sbírané přímo v terénu, tj. nepřebírané),
- **datum** (dcterms:created) – datum zaevidování bodu, tj. datum sběru, nebo datum stažení dat.

Mezi povinné informace – atributy uváděné u POI bodu dle datového modelu databáze SPOI (viz kapitola 1.2.4.), ale vytvářené (zadávané) až během vkládání do databáze (nedoplňuje je tedy „sběrač“), patří:

- **identifikátor** (dc:identifier) – kopie jedinečného identifikátoru objektu (bodu),
- **vydavatel** (dc:publisher) – vydavatel dat (SPOI – <http://sdi4apps.eu/spoi>),
- **titul, název** (dc:title) – nejdůležitější název bodu zájmu (většinou kopie názvu bodu), který se uvádí pro možnou spolupráci s metadatovými portály a vyhledávači,
- **licence** (dc:rights) – licence zájmového bodu v databázi SPOI (ODbL – <http://opendatacommons.org/licenses/odbl/1.0/>).

Dalšími užitečnými (vhodnými) informacemi – atributy pro body zájmu do databáze SPOI jsou:

- **popis** (rdfs:comment) – další informace o bodu, které ho blíže specifikují, nastíní vlastnosti, historii či jiné vhodné a zajímavé údaje,
- **adresa** (locn:fullAddress) – standardní zápis adresy ve formátu ulice, číslo popisné, poštovní směrovací číslo, město, část města,
- **další kontaktní informace** – uvádí se, pokud je daný bod má (obchod, restaurace, instituce, úřady,...):
 - **e-mail** (foaf:mbox),
 - **telefon** (foaf:phone),
 - **otevírací doba** (poi:openingHours),
 - **webová stránka** (foaf:homepage) – odkaz na webovou stránku,
- **fotografie** – nemá svůj zvláštní atribut – odkaz na fotografii se vkládá do vazby rdfs:seeAlso (více v následující kapitole),
- **odkazy na daný bod** (skos:exactMatch, owl:sameAs v nynější verzi datového modelu) – webové stránky odkazující na stejný bod,
- **odkazy na informace o bodu** (rdfs:seeAlso) – webové stránky poskytující další informace,
- **vazby (odkazy) na jiné body** (gn:nearby; geos:sfContains, geos:sfCrosses, org:unitOf, org:hasUnit, gn:locatedIn a další vazby navržené Kiráľem (2017) – viz kapitola 1.2.4.) – vazby (odkazy) týkající se vkládaného bodu,
- **autor** (dcterms:creator) – jméno toho, kdo body zaevidoval (nasbíral, získal),
- **originální licence** (dcterms:license) – odkaz na originální (původní) licenci (v případě, že se nejedná o body sbírané přímo v terénu, tj. nepřebírané).

V případě, že se jedná o aktualizaci či opravu bodu, resp. jeho informací (údaje a odkazy), je vhodné doplnit také tyto atributy:

- **datum změny** (dcterms:modified) – datum poslední editace či aktualizace informací (hodnot) u bodu,
- **editor** (dcterms:contributor) – jméno toho, kdo přispěl (doplnil informace – hodnoty) k bodu – provedl editaci či aktualizaci (více v kapitole 1.2.4. nebo v kapitole následující).

Co se týče adresy, tak je velmi vhodné ji vyplňovat a to z důvodu, že v mapovém klientu nejsou zobrazovány souřadnice. To znamená, že, pokud není vyplněna adresa, uživatel přímo v mapovém klientu nezjistí přesnou adresu či polohu bodu (pouze vizuálně z mapy nebo přes odkaz poi_id bodu v Advanced view). Jiná metoda, jak zjistit souřadnice bodu, je poté již jen skrze SPARQL dotaz, jež neovládá každý, či nevhodným kontaktováním správce.

2.2. Evidence dat

Sbíraná či získávaná data (body zájmu) spolu s informacemi je nutné zanést (vložit) do databáze SPOI. Vkládání bodů přímo uživatelem do databáze SPOI (pomocí SPARQL příkazu INSERT nebo skrze mapového klienta) není povoleno (jsou omezena práva uživatelů jakožto veřejnosti pro vklad a jiné nakládání s body mimo vyhledávání, tj. dotazování nad databází) z důvodu možného zneužití (např. vymazání, nesprávný přesun bodu). Jediným možným způsobem zanesení nových bodů do databáze SPOI je poskytnutí souboru s evidovanými (nasbíranými či získanými) daty správci. Kontrola přijatých dat probíhá tak, že se kontrolují povinné prvky (dané atributy – viz předchozí kapitola) a poté již pouze určité záležitosti a jen v některých formátech. Z hlediska syntaxe (skladby) kontroluje správce pouze data vedená ve formátech, které lze běžně kontrolovat (automatická kontrola) pomocí dostupných nástrojů (parserů¹³⁵) – např. RDF, XML,... Data vedená v souboru CSV jsou kontrolována z hlediska správnosti vložení oddělovacího znaku, tj. zda oddělovací znak – středník (;) není použit i v textu u popisu hodnot jednotlivých atributů, jelikož by se poté atributy a k nim patřící informace (hodnoty) rozdělily a přiřadily špatně.

Poskytnutí souboru s evidovanými (nasbíranými či získanými) daty správci¹³⁶ lze učinit:

- zasláním e-mailem,
- sdílením online,
- stažením skrze WFS.

Zaslání a sdílení

Použitelnými formáty, ve kterých budou data uchováвана a poskytována správci, jsou formáty strukturované. Jelikož se ovšem nepředpokládá znalost programovacích jazyků u široké veřejnosti a uživatelů, je nejvhodnějším způsobem uchovávat data v tabulkách (XLS a CSV). Tento formát je dostupný a navíc je v něm možná jednoduchá dodatečná editace již vložených dat (úprava atributů a jejich hodnot).

Konkrétním použitelným kancelářským programem je např. Office Excel¹³⁷ od společnosti Microsoft, který je velmi rozšířený a navíc jej ovládá velké množství lidí (mnoho z nich má tento program na svých počítačích). Excel je offline program, což znamená, že pro jeho

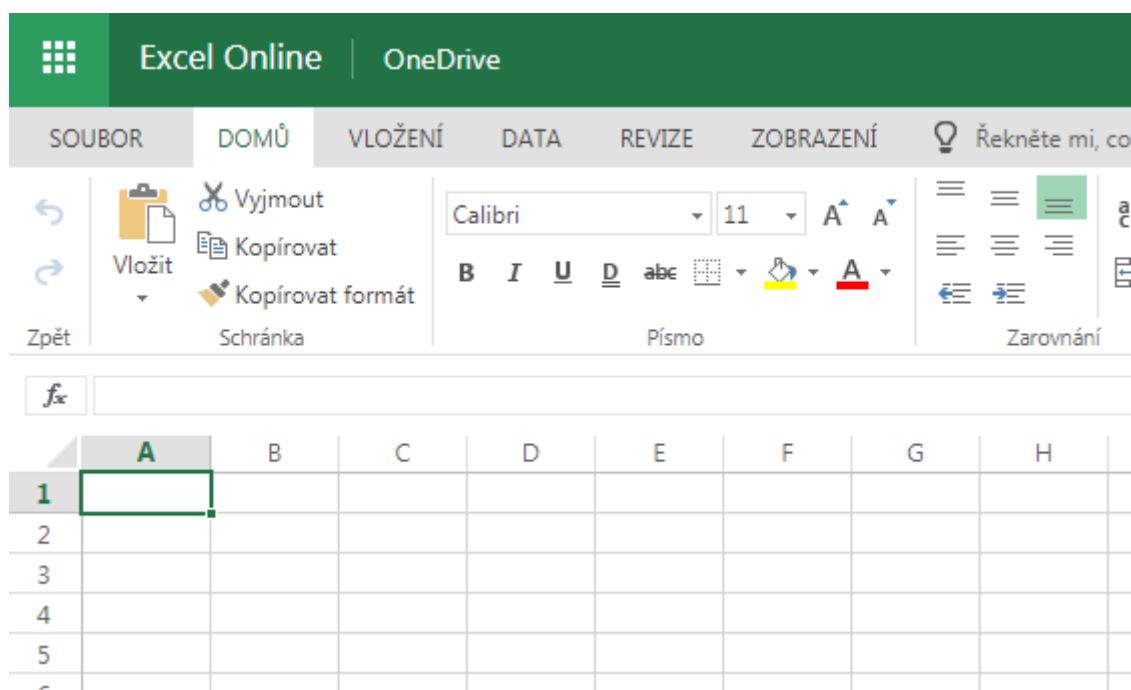
¹³⁵ <http://www.kosek.cz/clanky/swn-xml/ar02s30.html>

¹³⁶ Veškeré kontaktní informace na správce jsou dostupné na webové stránce projektu SPOI – <http://sdi4apps.eu/spoi/>.

¹³⁷ <https://products.office.com/cs-cz/excel>

použití se nepotřebuje připojení k Internetu. Podobným způsobem lze využít i jiné tabulkové offline alternativy, kterými jsou např. Apache OpenOffice¹³⁸ nebo LibreOffice.org¹³⁹.

Co se týče online aplikací pro evidování dat (bodů), tak lze využít např. také Microsoft Excel Online¹⁴⁰ (Obr. 15), ke kterému je ale potřeba vytvořit účet (zaregistrovat se). Soubor může být buď vytvořen zcela nový anebo nahrán z počítače, poté jej lze uložit na úložiště OneDrive¹⁴¹ také od společnosti Microsoft anebo stáhnout ve formátu XLSX či OSD (OSD 2017). Tato online aplikace má stejné vlastnosti a možnosti jako desktopová verze, navíc ji lze sdílet s dalšími uživateli (mohou obsah také upravovat) a je možné v ní pracovat po nastavení také v offline režimu, kdy se po opětovném připojení k Internetu vše automaticky uloží na OneDrive. Microsoft Excel lze taktéž stáhnout na mobilní zařízení jako mobilní aplikaci, která umožňuje uložit tabulku do zařízení jako XLSX nebo ji exportovat ve formátu ODS či PDF. Obě aplikace (mobilní i online) jsou napojeny na úložiště OneDrive, což znamená, že uložené soubory lze otevřít kdekoliv.



Obr. 15: Snímek prázdného listu v aplikaci Microsoft Excel Online
Zdroj: Microsoft Excel Online

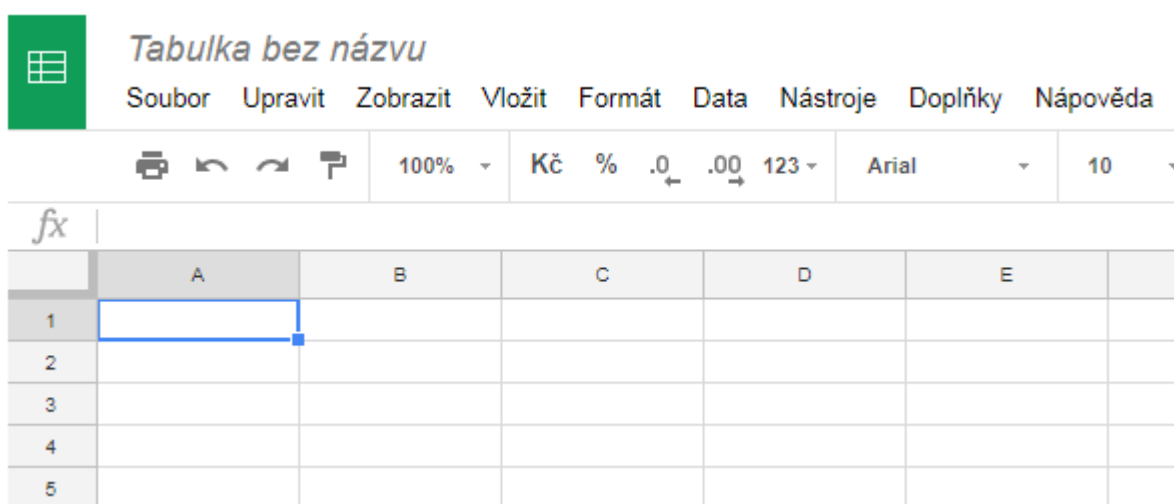
¹³⁸ <https://www.openoffice.cz/>

¹³⁹ <https://cs.libreoffice.org/>

¹⁴⁰ <https://office.live.com/start/Excel.aspx?omkt=cs-CZ>

¹⁴¹ <https://onedrive.live.com/about/cs-cz/>

Pro sdílení a sepisování dat (bodů) více lidmi je vhodným online prostředkem taktéž aplikace Tabulky¹⁴² (Obr. 16) využívající úložiště Google Disk¹⁴³ (Drive) obojí od společnosti Google, Inc.¹⁴⁴ Tato aplikace umožňuje uživateli vytvořit novou tabulku anebo nahrát (na úložiště) tabulku již vytvořenou s daty (body) na desktopovém počítači a dále s ní pracovat (vkládat další body či editovat ty stávající) nebo ji sdílet s dalšími uživateli. Pro tuto práci je nutné vlastnit účet (být zaregistrován, tj. mít e-mail) u společnosti Google, Inc. Aplikace Tabulky je dostupná online (je nutné připojení na Internet), ale lze s ní po nastavení pracovat taktéž v offline režimu a po opětovném připojení k Internetu vše zpětně synchronizovat, navíc existuje také jako mobilní aplikace. Jeden soubor je tedy možné mít otevřen na mobilním zařízení a zároveň online na Internetu na počítači (díky napojení na úložiště).



Obr. 16: Snímek prázdného listu v aplikaci Tabulky Google
Zdroj: Tabulky Google

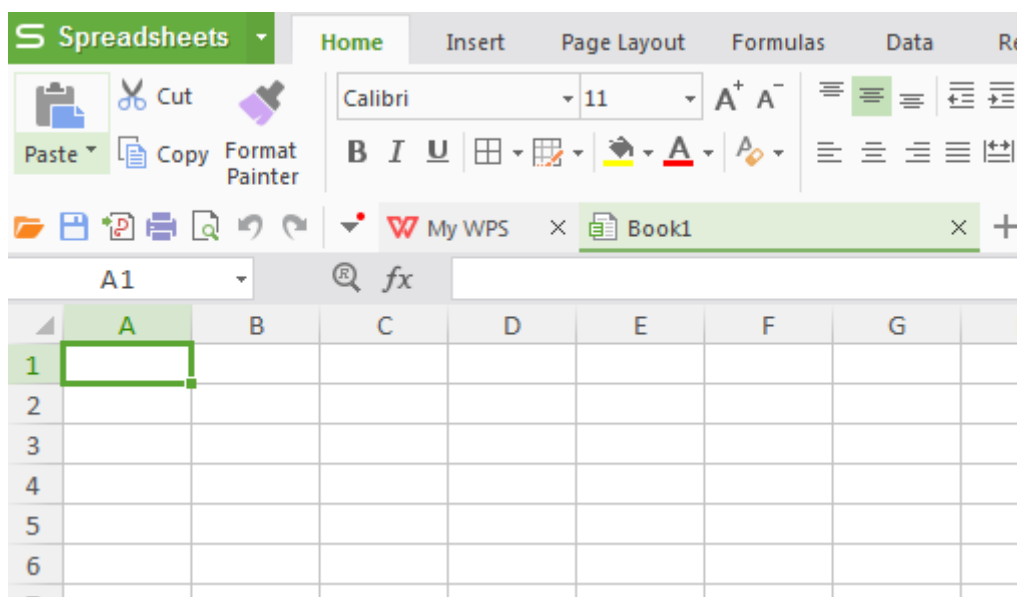
Nespornou výhodou obou těchto aplikací (Microsoft Excel Online a Google Tabulky) je možnost přístupu (po přihlášení) odkudkoliv (z jakéhokoliv zařízení s přístupem k Internetu) a také možnost, jak již bylo zmíněno výše, pracovat offline a editovat, komentovat či prohlížet soubor vícero lidmi (možno i zároveň), pokud mají odkaz na daný soubor. Právě z těchto důvodů jsou tyto aplikace nejvhodnějším způsobem pro evidenci sbíraných a získávaných dat.

¹⁴² <https://docs.google.com/spreadsheets/>

¹⁴³ <https://drive.google.com/>

¹⁴⁴ https://www.google.cz/intl/cs_cz/

Jako další mobilní aplikaci pro evidenci lze využít (mimo zmíněné aplikace Microsoft Excel Online a Google Tabulky) WPS Office¹⁴⁵ (Presentation, Spreadsheets [Obr. 17] a Writer). Tato aplikace nabízí uložení dat ve formátu XLS nebo XLSX a export do PDF či obrázku. Ukládat a exportovat (nahrávat) je možné na různá úložiště. Aplikaci lze stáhnout také jako desktopovou (zdarma nebo rozšířenou za poplatek). Po zaregistrování mohou sdílet mobilní i desktopová aplikace stejné úložiště poskytované přímo kancelářskou aplikací WPS Office.



Obr. 17: Snímek prázdného listu v aplikaci WPS Office – Spreadsheets
Zdroj: WPS Office – Spreadsheets

Stáhnutí skrze WFS

Stáhnutí je další způsob, jak lze data publikovat, tj. dostat ke správci, – jedná se o zpřístupnění dat prostřednictvím služby WFS (OGC: Web Feature Service). WFS je zkratka pro Web Feature Service, což je webová stahovací služba, která slouží k publikování vektorových dat. Principem je nahrání souborů s daty na server, z kterého si je jiný uživatel (správce) bude moci stáhnout a dále používat (nahrát do databáze SPOI).

Příkladem serveru, poskytujícím tuto službu, tj. kde lze sesbíraná či získaná data publikovat tímto způsobem, je GeoServer.org¹⁴⁶. Tento server je určen pro sdílení geoprostorových dat, je open source a jeho spouštěcí komponenty, nadstavbové prvky a dokumentaci lze stáhnout na stejnojmenných webových stránkách¹⁴⁷. Na serveru¹⁴⁸ lze po instalaci potřebných komponentů a následném přihlášení vytvořit nový pracovní prostor (místo) neboli

¹⁴⁵ <https://www.wps.com/>

¹⁴⁶ <http://geoserver.org/>

¹⁴⁷ <http://geoserver.org/release/stable/>

¹⁴⁸ <http://localhost:8080/geoserver/web/>

nadřazenou složku pro práci (Workspace), kde je nutné zaškrtnout WFS u nabídky služeb (Services), nové úložiště (Store), které je mimo jiné možné propojit s určitým adresářem v počítači (volba při vytváření úložiště Vector Data Sources – Directory of spatial files) pomocí nastavení cesty do tohoto adresáře a poté do něj přidávat jednotlivé soubory (s daty) ve formátu SHP, které mají být sdíleny. Všechny tyto soubory v adresáři se poté zobrazí na serveru skrze přidání nové vrstvy (Layer), kde uživatel vybere konkrétní název „Workspace:Store“, kde jsou jeho data uložena, a může je publikovat a tím zpřístupnit ostatním (Obr. 18). V nastavení publikování je nutné vyplnit prostorové ohraničení dat (minimální a maximální hodnoty souřadnic – Bouding boxes), jež lze vypočítat přímo z nich (kliknutí na Compute from data a Compute from native bounds) a navíc je možné v záložce publikování (Publishing) nastavit styl vrstvy (Default Style), která je přidávána. Zpřístupněná data je poté možno zobrazit (stáhnout) pomocí služby WFS skrze GIS program (např. QGIS – zobrazená stažená data (svět a rovnoběžky s poledníky) uložená na GeoServer.org skrze webovou službu WFS v programu QGIS jsou na Obr. 19), kam je zadávána příslušná URL adresa (http://localhost:8080/geoserver/<nazev_workspace>/wfs), ze které jsou data (vrstvy) stahována. Více informací o WFS a nastavení na serveru GeoServer.org se nachází v uživatelském manuálu¹⁴⁹.

New Layer

Add a new layer

Add layer from

You can create a new feature type by manually configuring the attribute names and types. [Create new feature type...](#)

Here is a list of resources contained in the store 'svet'. Click on the layer you wish to configure

<< < 1 > >> Results 0 to 0 (out of 0 items)

Published	Layer name	Action
	lat_lon	Publish
	svet	Publish

<< < 1 > >> Results 0 to 0 (out of 0 items)

Obr. 18: Publikování vrstev na serveru GeoServer.org

Zdroj: vlastní

¹⁴⁹ <http://docs.geoserver.org/2.12.2/user/services/wfs/index.html>



Obr. 19: Stažená data z webové služby WFS z GeoServer.org v programu QGIS
Zdroj: vlastní

Vložení dat do databáze SPOI

V případě již správně evidovaných bodů (ověřených) ve vlastním zdroji (např. v Tabulce Google, XLS [Office Excel], CSV,...) je dalším krokem vložení nasbíraných dat přímo do databáze SPOI, jež provádí správce, který obdrží soubor s body (zaslaný či sdílený) nebo si jej stáhne. Před samotným vkladem do databáze SPOI provede správce harmonizaci dat a také jejich kontrolu (viz výše). Pro jednodušší kontrolu povinných atributů je vhodné, aby poskytované soubory (ve formátu tabulky [CSV a XLS], ale také konvertované soubory např. SHP) měly vždy jednotnou strukturu, tj. pořadí jednotlivých atributů. Na začátku musí být povinné prvky (Obr. 20), které vyplňuje „sběrač“ (název, souřadnice [délka, šířka], typ objektu, umístění [stát, kraj, okres, obec], zdroj a datum vložení), a po nich následují další užitečné (vhodné) informace (popis, adresa, kontaktní informace [e-mail, telefon, webová stránka, otevírací hodiny], fotografie, odkazy na daný bod, na další informace o něm či na jiné související body, autor, originální licence, datum poslední aktualizace a jméno toho, kdo ji provedl v případě editace či aktualizace dat – editor).

	A	B	C	D	E	F	G	H	I	J	K
1	název	zeměpisná délka	zeměpisná šířka	typ objektu	stát	kraj	okres	obec	zdroj	datum	autor
2											
3											
4											
5											

Obr. 20: Hlavička tabulky pro evidenci POI bodů – povinné atributy
Zdroj: vlastní

Pokud by měl jeden atribut více hodnot a jednalo by se o vazbu (odkaz na jiný bod či jiný zdroj informací), je příhodné, aby se všechny nacházely za sebou, resp. v po sobě jdoucích sloupcích (např. odkaz na umístění – stát, kraj, okres, obec atd.). Pokud by se jednalo o atribut, který nese určitý údaj (textový řetězec, např. telefonní číslo, e-mail atd.), mohou být tyto údaje zapsány do jednoho atributu, resp. sloupce, pomocí oddělovacího znaku svislé čáry s mezerami (|). Tento znak svislé čáry se standardně v textu nepoužívá, proto, pokud bude v dané hodnotě atributu použit, bude jasné, že se jedná o další informaci, tj. že atribut má 2, popř. více údajů.

Vstupy, tj. poskytované soubory s daty by měly mít stejně označovány atributy, aby byly jednotlivě správně přepisovány spolu s odpovídajícími hodnotami (údaji a odkazy) skriptem do databáze. Pro lepší přehlednost a eliminaci možných chyb při doplňování se označují jednotlivé atributy českými ekvivalenty ke standardním názvům atributů (dle dokumentu, standardu či datové sady, z které vychází – viz kapitola 1.2.3. a 1.2.4.). V případě dat zasílaných či sdílených ve formátu CSV či XLS bere skript první řádek jako jména atributů po sloupcích a ostatní řádky jako jednotlivé hodnoty k nim, přičemž jeden řádek reprezentuje jeden bod zájmu. To znamená, že struktura souborů musí být zachována – hlavička (první řádek) jako nositel atributů po sloupcích a ostatní řádky jako reprezentanti jednotlivých bodů zájmu a po sloupcích jejich jednotlivé údaje a odkazy (hodnoty).

Dodržení formátu a struktury jednotlivých hodnot atributů u bodů zájmu je v některých případech zásadní. Zde jsou uvedena pravidla pro doplňování informací o bodech.

- Podstatné je, aby zadávané souřadnice bodu byly geografická délka a šířka v souřadnicovém systému WGS 84 v desetinném úhlovém formátu, přičemž oddělovačem desetinného místa je vždy tečka (.). Délka a šířka se vyplňují každá zvlášť do svého atributu, resp. sloupce a to z důvodu eliminace chyby během evidence.
- Důležitým pravidlem je nepoužívat v textu středník (;) a svislou čáru (|) – viz výše.
- Dodržovat se také musí zápis data zaevidování či editace a aktualizace bodu, kterým je 'RRRR-MM-DD' (např. 2018-01-31).
- Vhodné je dodržovat jeden formát telefonního čísla, kdy je důležité, aby číslo určitým způsobem zahrnovalo také mezinárodní předvolbu (+420), jelikož databáze SPOI obsahuje body z celého světa. Formát čísla by měl být '+420 xxx xxx xxx', '00420 xxx xxx xxx' anebo '420 xxx xxx xxx' a je možné jej zapsat také bez mezer.

- Je také doporučeno, aby název bodu nebyl zvláště dlouhý a to z důvodu, aby při zobrazování v mapové aplikaci nezabral příliš mnoho místa z obrazovky (nežádoucí např. v mobilním zařízení) nebo se případně ukázal celý či v rozumné velikosti.

V případě sběru v terénu a používání mobilní aplikace pro sběr souřadnic a dalších informací o bodech je nutné data vyexportovat v určitém formátu, který musí mít opět stejnou konečnou strukturu (pořadí a názvy atributů) – to znamená, že vyexportovaný soubor je nutné upravit tak, aby odpovídal standardnímu výstupu. V případě exportu dat z aplikace v souboru CSV lze data převést do tabulky (XLS) a dále je upravovat, v případě exportu dat v jiném formátu se musí použít konvertor (viz předchozí kapitola) a transformovat soubor do formátu jiného (nejlépe opět do tabulky) a také jej poté upravit do standardní podoby.

Pokud uživatel k bodům pořídí nové fotografie (své vlastní), je nutné tyto fotografie uložit na Internet a poté do atributu (k vazbě určené pro fotografie – `rdfs:seeAlso`) vložit odkaz přímo na nově uloženou fotografii na Internetu.

2.3. Možnosti kontroly a aktualizace

2.3.1. Kontrola

Nejenom sběr a doplňování databáze SPOI novými body zájmu je důležité. Zásadní je také udržovat data (údaje a odkazy) aktuální a správné. Zastaralé informace (např. kontaktní údaje, otevírací doba, nesprávná adresa z důvodu přemístění prodejny či provozovny atd.) jsou nežádoucí a zbytečně matoucí. Také konzistentnost dat z hlediska propojení (odkaz na sousedící bod zájmu, podřízený či nadřazený objekt atd.) je nutné udržovat. V případě nesprávně propojených dat způsobených neaktualizací (nezanesení a neopravení) nové skutečnosti v datech, by potom vznikaly nejasnosti – např. přesun pobočky banky a neaktualizace nové adresy by způsobila to, že nadřazený zájmový bod (hlavní organizace či sídlo) by odkazoval na svůj podřízený bod (pobočku), který ale již neexistuje, resp. existuje na jiné adrese.

Kontrola dat může být prováděna pomocí:

- mapového klienta,
- správce,
- SPARQL dotazu.

V mapovém klientu lze kliknutím na bod zájmu (a dále Advanced view) zjistit zaevidované informace o něm a tím také případné chyby a nejasnosti. Nelze ovšem zjistit další vazby vůči ostatním bodům – pouze vizuálně z mapy. Kontrola dat ze strany správce je možná, ale jelikož se stará o celou databázi SPOI, zabývá se mnoha dalšími věcmi s ní souvisejícími a bodů je velké množství, není v jeho moci kontrolovat aktuálnost dat. Další a opět také nejpříhodnější možností, jak může uživatel zjistit údaje a odkazy týkající se bodu či bodů dle jeho požadavku, je pomocí SPARQL dotazu (více v kapitole 2.1.1.).

Příklady dotazů na databázi SPOI

Zelenou barvou jsou části, které lze změnit dle uživatelského zájmu a tím dostat nový dotaz.

1. Body na území České Republiky, které byly přidány do databáze před 1. 1. 2018 včetně:

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
PREFIX dcterms: <http://purl.org/dc/terms/1.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT ?bod ?navez ?datum ?souradnice ?adresa
WHERE
{
  ?bod geos:sfWithin <http://dbpedia.org/resource/Czech_Republic> .
  ?bod geos:asWKT ?souradnice .
  ?bod rdfs:label ?navez .
  ?bod locn:fullAddress ?adresa .
  ?bod dcterms:created ?datum .
  FILTER (?datum <= xsd:dateTime('2018-01-01'))
}
ORDER BY DESC 3
LIMIT 1000
```

2. Všechny dostupné informace z databáze o bodu, který má v adrese slovo 'Černohorská':

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
SELECT DISTINCT ?bod ?atribut ?hodnota
WHERE
{
  {
    {SELECT ?bod
     WHERE
     {
       ?bod geos:sfWithin <http://dbpedia.org/resource/Czech_Republic> .
       ?bod locn:fullAddress ?adresa .
       FILTER regex(?adresa, 'Černohorská', 'i')
     }}
    ?bod ?atribut ?hodnota
  }
}
ORDER BY ASC 1
```

3. 1000 bodů na území České Republiky, které byly naposledy pozměněny:

Dotaz je využitelný v případě využívání atributu dcterms:modified v databázi SPOI a lze v něm pozměnit počet bodů nebo vybrat body, které byly pozměněny nejdříve (řazení ASC).

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX geos: <http://www.opengis.net/ont/geosparql#>
PREFIX dcterms: <http://purl.org/dc/terms/1.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT ?bod ?nazev ?datum ?souradnice ?adresa
WHERE
{
  ?bod geos:sfWithin <http://dbpedia.org/resource/Czech_Republic> .
  ?bod geos:asWKT ?souradnice .
  ?bod rdfs:label ?nazev .
  ?bod locn:fullAddress ?adresa .
  ?bod dcterms:modified ?datum
}
ORDER BY DESC 3
LIMIT 1000
```

2.3.2. Aktualizace

V případě zjištění nesrovnalostí či chybných údajů, lze data aktualizovat a to tím, že se:

- provede editace v mapovém klientu,
- vytvoří a pošle změnový soubor:
 - s tabulkou pouze měněných hodnot atributů u bodu,
 - s novou tabulkou obsahující kompletní informace o bodu.

Zjištěné chyby nebo chybějící údaje a informace může uživatel ihned opravit, resp. aktualizovat a doplnit v mapovém klientu. Způsob přidávání fotografií k již evidovaným bodům je pomocí odkazů na ně a také vložením přímo fotografie, jako obrázku, k danému bodu. Opravené nebo nově vložené informace a fotografie (jakýmkoliv způsobem) se ukládají do vedlejší změnové databáze a jsou s primární databází SPOI propojeny (vložené do ní) jednou za čas (až je ve vedlejší databázi velké množství aktualizovaných či nových informací). To znamená, že uživatelem vložené a uložené informace nelze hned vidět – nejsou ihned po uložení u bodu evidované. V mapovém klientu ovšem nelze editovat souřadnice zájmového bodu, tj. s ním hýbat – přesunovat jej na jiné místo.

Nahlášení chyby, resp. změněných (opravených či aktualizovaných) dat lze učinit také pomocí zaslání e-mailu správci. Tento e-mail by měl obsahovat změnový soubor, kterým je buď tabulka pouze se změněnými hodnotami atributů u bodu, nebo tabulka s kompletními informacemi o bodu zájmu. V tabulce s měněnými hodnotami atributů (Obr. 21) budou uvedeny body, kterých se změna týká pomocí jedinečného identifikátoru bodu pro jeho jednoznačnou identifikaci (dc:identifier – lze zjistit pomocí mapového klienta v Advanced view [také poi_id bodu] či SPARQL dotazem). Dále bude uveden konkrétní atribut, který je chybný či neúplný a má být změněn (nadepsáno jako „původní hodnoty“), a nová (aktuální) hodnota (odkaz či údaj), která tu původní nahradí (nadepsáno jako „nové hodnoty“). Pokud je aktualizovaných hodnot více, uvádí se nejprve všechny nesprávné a poté za ně ve stejném pořadí hodnoty nové. V případě pouhého doplnění informací (ne jejich oprava) se použije mapový klient anebo druhý způsob nahlášení. Druhá možnost (vytvoření celé nové (aktualizované) tabulky s kompletními informacemi o bodu – viz kapitola 2.2.) znamená vytvoření tabulky s body, kterých se změna týká, se všemi atributy, z nichž některé jejich hodnoty (údaje či odkazy) jsou změněny (jsou jiné než ty doposud vedené). V obou těchto souborech by měla být uvedena také informace o tom, kdo (dcterms:contributor – odkaz na oficiální profil) a kdy (dcterms:modified – datum ve tvaru ‘RRRR-MM-DD’) tuto

skutečnost zaevidoval, tj. opravil, aktualizoval. Takový soubor (možno převedený do jiné strukturované formy) se poté zašle správci, který jej zpracuje a zaeviduje do dat v databázi SPOI, tj. dle něj aktualizuje a doplní údaje a odkazy. Tento soubor se opět kontroluje jako soubor s nově evidovanými body (viz kapitola 2.2.).

	A	B	C	D	E	F	G
1		původní hodnoty		nové hodnoty			
2	identifikátor	e-mail	telefon	e-mail	telefon	datum změny	editor
3	http://www.sdi	xy@xy.cz	+420 xxx	yx@yx.cz	+420 yyy	2018-05-01	http://www
4							
5							

Obr. 21: Tabulka s měněnými hodnotami

Zdroj: vlastní

Pomocí SPARQL příkazů nelze data editovat (aktualizovat) přímo v databázi (stejně tak jako přidávat [vytvářet] body nové – viz kapitola 2.2.). Jelikož ale pro aktualizaci a editaci bodů v databázi SPOI (jakýmkoliv jiným způsobem) není potřeba žádné registrace či ověření, tj. editovat body zájmu může kdokoliv, je tedy na každém uživateli, jak svědomitě a správně bude body editovat a databázi doplňovat.

3. Ověřování metodiky

3.1. Způsob ověřování metodiky

Metodika byla ověřována sběrem a získáváním bodů podle pokynů v ní napsaných. Sběr a získávání bodů prováděly nezávislé osoby a to z důvodu, že je metodika určená pro laiky a pro lidi, kteří o projektu a doplňování bodů zájmu do databáze Smart Points of Interest neví všechny informace. Testování mnou, jakožto autorem metodiky, by nebylo objektivní.

Osoby, které prováděly ověřování, tj. sbíraly a získávaly data do databáze SPOI, následně poskytovaly zpětnou vazbu, která se týkala metodiky jak z hlediska porozumění, použitelnosti a formální stránky, tak také z hlediska průběhu sběru a získávání dat. Zpětná vazba byla získávána ve formě krátkého dotazníku od osob, které metodiku testovaly.

Krátký dotazník se skládal z otázek:

- 1) Je metodika pochopitelná? Ano / Ne – Blíže specifikujte.
- 2) Lze podle metodiky snadno postupovat během sběru a získávání dat?
Ano / Ne – Blíže specifikujte.
- 3) Byla Vám metodika během sběru, získávání a následného evidování dat užitečná?
Ano / Ne – Blíže specifikujte.
- 4) Je v metodice nějaká část, která by chtěla více popsat pro lepší srozumitelnost?
Ano / Ne – Blíže specifikujte.
- 5) Jaké další informace byste v metodice ještě uvítali, resp. co v ní chybí nebo není dostatečně popsáno či vysvětlené?
- 6) Jaké jsou klady metodiky?
- 7) Jaké jsou zápory metodiky?
- 8) Jaký je podle Vás největší přínos metodiky?
- 9) Kterou metodu sběru či získávání dat jste si vybrali a proč?
- 10) Popište krátce postřehy během sběru a získávání dat dle metodiky.

3.2. Výsledky dotazníku

Odpovědi na dotazník byly zaslány pouze 6 respondenty, přičemž jich bylo obesláno 56. Dále bylo obesláno 19 informačních centrem po celé České republice, ale nebyly obdrženy žádné relevantní odpovědi. Ověření metodiky 6 respondenty není dokonalé, ale stále je to určitá forma zpětné vazby. Její hodnocení bylo velmi rozličné – 2 dotázaní uváděli, že metodika je složitá a nesrozumitelná a bez přečtení teorie ji nelze pochopit, na druhé straně byly 4 názory, že metodika je pochopitelná, lze podle ní přesně postupovat během sběru a získávání dat do databáze SPOI a její podstata je v práci dostatečně vysvětlena.

4 dotazovaní uváděli, že by uvítali, kdyby byly jednotlivé odborné záležitosti (např. jazyk SPARQL, transformace dat či vytěžování skriptem) více popsány a vysvětleny. Cílem této práce ovšem není všechny odborné záležitosti vysvětlit ani naučit, jak je používat. Není to ani v jejích rozsahových možnostech, protože tím by se také obsah práce podstatně rozšířil. Pokud by uživatel zajímal, jak tyto věci fungují a jak se je naučit používat, jsou v práci uváděny odkazy na standardy, dokumenty a stránky zabývající se danou problematikou. Na Internetu je navíc mnoho dalších dostupných zdrojů.

Obecně lze říci, že při čtení metodiky je třeba věnovat pozornost tomu, jak vypadá databáze SPOI a jednotlivým souvisejícím a navazujícím oblastem, aby byla správně pochopena. Je doporučeno přečíst si také teorii v 1. kapitole, aby se uživatel dovedl lépe orientovat v dané problematice.

Díky propojení sběru dat s turistikou si většina dotazovaných (5) vybrala metodu ručního sběru dat v terénu – spojení příjemného s užitečným. Další v pořadí se umístila metoda sběru dat s využitím dat z Internetu, která je veřejnosti také blízká, jelikož v dnešní době tráví lidé na Internetu mnoho času.

Mezi hlavní postřehy během sběru a získávání dat dle metodiky patří to, že při ručním sběru dat v terénu se může uživatel setkat s tím, že na některých místech nejsou dostatečné informace pro nepovinné, ale užitečné atributy (kontakt, otevírací doba atd.), tudíž je zapotřebí využít i ruční metodu sběru dat na Internetu. Při ručním sběru dat v terénu se hodí vytisknout si s sebou připravené prázdné tabulky a data vepisovat do nich, popř. u chytrých zařízení je mít připravené a vyplňovat tabulku během sběru dat v terénu přímo v telefonu či tabletu. Jako jediná překážka použití metody sběru dat v terénu je to, že ne na každém místě

je dostatečná viditelnost z hlediska signálu družic, a tak změřené GPS souřadnice mohou být na některých místech méně přesné.

3.2.1. Silné a slabé stránky metodiky

Jelikož hodnocení metodiky a celkové práce bylo velmi různorodé, jsou zde uvedeny spíše konkrétní klady a zápory dle dotazovaných respondentů.

Klady

- Kladně hodnocené je uvedení příkladů, odkazů na doplňující informace a obrázků pro snazší představu, za přehlednost pak obzvláště Obr. 3: Diagram postupu doplňování databáze SPOI.
- Za pozitivní respondenti považovali objasnění možností, kterými je možné data sbírat, jak pro laickou veřejnost, tak pro osoby, které mají znalosti dostačující k tomu, aby mohly využít složitější metody pro sběr dat (vytěžování pomocí skriptu a transformace).
- Jako nejkladnější z celé metodiky byla hodnocena část Data z terénu a Doplňování atributů v kapitole 2.1.2. Sběr a získávání dat, kde je popsán sběr dat z terénu a do detailu popsán, na co se uživatel musí zaměřit a jaké informace musí získat o daném místě (bodě), jež by chtěl přidat do databáze SPOI.
- Jeden z dotazovaných také zmínil, že metodika popisuje celý proces od začátku do konce a je tedy komplexní. Pokud někdo bude chtít přispět do databáze SPOI, bude nyní vědět, čím začít a jak postupovat. Díky tomu by do SPOI mohlo přispívat více lidí.
- Dále někteří respondenti viděli velký přínos v tématu práce a to z důvodu, že i „obyčejný člověk“ může přispět svým konáním k tomu, aby se další lidé mohli dozvědět něco o zajímavých místech České republiky.

Zápory

- Úplní laici by uvítali podrobnější popsání zkratk, odborných termínů či jednotlivých aplikací, programů a informačních systémů.
- Respondenti také uváděli, že by bylo vhodné pro lepší pochopení uvedené příklady SPARQL dotazů na databázi SPOI více vysvětlit a také uvést ještě více příkladů vkládání bodů do databáze SPOI.
- Záporům této metodiky podle některých dotázaných je to, že při jejím čtení (konkrétně například u SPARQL dotazů či transformace dat) se v ní může neodborná veřejnost lehce ztratit – je v ní mnoho různých možností a uživateli může být nejasné, co si má vybrat.

- Pro člověka, který neumí programovat a neumí používat metodu SPARQL dotazu, je podle pár respondentů velmi složité a časově náročné ověřit, zda už bod v databázi SPOI je nebo není.
- Jeden z dotázaných poukázal na to, že v metodice není jasně řečeno, co je využitelné pro naprostého laika a že je metodika příliš dlouhá (jednotlivé možnosti jsou zbytečně moc popisovány).

3.3. Možnosti vylepšení

Zde jsou uvedena možná vylepšení metodiky plynoucí z odpovědí od dotázaných respondentů.

- Podrobně popsat a vysvětlit jednotlivé použité příklady SPARQL dotazů na databázi SPOI, což by pomohlo laické veřejnosti v tom, aby lépe pochopila, jak funguje databáze a dotazování nad ní (ověřování).
- Uvést ještě více příkladů vkládání bodů do databáze SPOI, aby byla metodika názornější a uživatel měl větší výběr možností. Tyto možnosti více popsat, aby bylo jasné, co která obnáší.
- Přehledně sepsat, které metody sběru a získávání dat do databáze SPOI jsou vhodné pro širokou veřejnost a které jsou odbornějšího rázu.
- Naopak práci a popis jednotlivých kroků a možností napsat jasně, věcně a stručně a nevysvětlovat vše zbytečně rozsáhle.

S ohledem na výše uvedené možnosti vylepšení a celkové hodnocení z předchozích kapitol, lze konstatovat, že cíl práce (seznámit s problematikou databáze Smart Points of Interest, doplňování bodů zájmu Points of Interest a navrhnout metodiku pro doplňování těchto bodů do databáze SPOI) byl splněn.

Závěr

Tato diplomová práce se zabývá databází Smart Points of Interest, jejím rozšiřováním a obohacováním body zájmu Points of Interest a návrhem metodiky pro doplňování těchto bodů do databáze SPOI.

Cíle práce stanovené v úvodu byly naplněny. Záměrem práce bylo seznámit čtenáře a budoucí uživatele s problematikou týkající se databáze Smart Points of Interest, poukázat na otázku jejího obohacování body zájmu a především vytvořit metodiku, podle které by se uživatelé řídili při doplňování dat do této databáze.

V teoretické části byly probrány obecné informace o databázi SPOI a projektu a také konkrétní odborné záležitosti týkající se důležitých záležitostí ohledně databáze SPOI. Byly probrány jednotlivé vazby a atributy datového modelu SPOI, navržené atributy rozšířeného datového modelu SPOI dle Kiráľa (2017) a taktéž navrženy 3 nové atributy, které by bylo vhodné zakomponovat do budoucího datového modelu. Také byly popsány dílčí kroky harmonizace dat vkládaných do databáze SPOI.

V praktické části, tedy metodice, bylo navrženo, jak by měl uživatel postupovat a byly nabídnuty a popsány způsoby provedení jednotlivých kroků, z nichž si může každý „sběrač“ vybrat, co je pro něj nejvhodnější a jak bude během své práce postupovat. Bylo zde popsáno mnoho alternativních možností a postupů pro ověřování, sběr a získávání, evidenci, kontrolu a aktualizaci dat vkládaných do databáze SPOI (sběr dat v terénu nebo na Internetu, pomocí transformace nebo ručně atd.). Také byla nabídnuta řada konkrétních programů a aplikací pro realizaci daných kroků (evidence dat offline či online, způsoby transformace formátů, poskytování dat správci atd.). Výsledkem metodiky je ulehčení práce uživatelům, jakožto „sběračům“ a také správci, který body do databáze SPOI vkládá. Především jsou pak díky používání metodiky a dodržování pokynů v ní evidovaná data uživateli zadávána jednotně a systematicky.

V poslední kapitole je metodika a celá diplomová práce hodnocena z hlediska použitelnosti a pochopení z řad veřejnosti. Kvůli nízké účasti mezi dotázanými není ověření statisticky průkazné, nicméně zasláné odpovědi jsou stále určitou zpětnou vazbou. Z odpovědí tázaných respondentů v dotazníku plynou velmi rozdílné názory. Většina dotazovaných hodnotila metodiku jako srozumitelnou a využitelnou v praxi, někteří ale naopak za velmi složitou

s nedostatečně vysvětlenými termíny a metodami. Objevily se také názory, že práce je velmi rozsáhlá s příliš podrobným vysvětlováním a popisem. Souhrnně dle odpovědí respondentů lze tedy říci, že by bylo vhodné jednotlivé možnosti, kroky a termíny v metodice více popsat a vysvětlit, ale zároveň tím práci příliš neprodlužovat, ba naopak ji zjednodušit a zkrátit. Na základě zpětné vazby byly některé termíny více vysvětleny a přidán krátký popis databáze Smart Points of Interest také na začátek metodiky, aby byla pochopitelnější a čtenář měl základní pojem o daných záležitostech v případě, že si nepřečte teoretickou část. Rozsah práce, bohužel, neumožňuje přeanalyzovat celou situaci, nicméně má diplomová práce spolu se zpětnou vazbou může být podkladem pro další bádání, rozvoj a vylepšení metod a záležitostí okolo databáze Smart Points of Interest.

Seznam použité literatury

[MS-XLS]: Excel Binary File Format (.xls) Structure (2017). v20171212. 12. 12. 2017. *Microsoft Corporation*. Dostupné z: [http://interoperability.blob.core.windows.net/files/MS-XLS/\[MS-XLS\].pdf](http://interoperability.blob.core.windows.net/files/MS-XLS/[MS-XLS].pdf)

[MS-XLSX]: Excel (.xlsx) Extensions to the Office Open XML SpreadsheetML File Format (2017). v20171212. 12. 12. 2017. *Microsoft Corporation*. Dostupné z: [http://interoperability.blob.core.windows.net/files/MS-XLSX/\[MS-XLSX\].pdf](http://interoperability.blob.core.windows.net/files/MS-XLSX/[MS-XLSX].pdf)

ANSI/NISO Z39.85-2012: The Dublin Core Metadata Element Set (2012). *ANSI: NISO* [online]. 20. 2. 2012 [cit. 2018-02-19]. Dostupné z: https://groups.niso.org/apps/group_public/download.php/10258/Z39-85-2012_dublin_core.pdf

AUER, Sören, Jens LEHMANN a Sebastian HELLMANN (2009). *Linkedgeodata: Adding a spatial dimension to the web of data* (pp. 731-746) [online]. [cit. 2018-02-19]. Springer Berlin Heidelberg. Dostupné z: https://link.springer.com/content/pdf/10.1007%2F978-3-642-04930-9_46.pdf

BAO, Jie, Elisa F. KENDALL, Deborah L. MCGUINNESS, Peter F. PATEL-SCHNEIDER, Li DING a Ankesh KHANDELWAL, ed. (2012). *OWL 2 Web Ontology Language Quick Reference Guide (Second Edition): W3C Recommendation 11 December 2012*. *World Wide Web Consortium* [online]. © 2012 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/TR/2012/REC-owl2-quick-reference-20121211/>

BAUER, Florian a KALTENBÖCK, Martin (2011). *Linked open data: The essentials* [online]. [cit. 2018-02-19]. Edition mono/monochrom, Vienna. Dostupné z: <https://www.reeep.org/LOD-the-Essentials.pdf>

BIZER, Christian, Tom HEATH a Tim BERNERS-LEE (2009). *Linked data-the story so far* [online]. [cit. 2018-02-19]. *International journal on semantic web and information systems*, 5(3), 1-22. Dostupné z: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

BIZER, Christian, Tom HEATH, Kingsley IDEHEN a Tim BERNERS-LEE (2008). *Linked data on the web (LDOW2008)* [online]. [cit. 2018-02-19]. In *Proceedings of the 17th*

international conference on World Wide Web (pp. 1265-1266). ACM. Dostupné z: <https://pdfs.semanticscholar.org/7472/17c301fe06309cd1088fbd53520fd1e38894.pdf>

BRAY, Tim (2015). RFC 7493: The I-JSON Message Format. *Internet Engineering Task Force (IETF): ISSN: 2070-1721* [online]. March 2015 [cit. 2018-02-19]. Dostupné z: <https://tools.ietf.org/html/rfc7493>

BRAY, Tim, Jean PAOLI, C. M. SPERBERG-MCQUEEN, Eve MALER, François YERGEAU a John COWAN (2006). Extensible Markup Language (XML) 1.1 (Second Edition): W3C Recommendation 16 August 2006, edited in place 29 September 2006. *World Wide Web Consortium: Extensible Markup Language (XML) 1.1 (Second Edition)* [online]. 16. 8. 2006, 29. 10. 2016 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/TR/2006/REC-xml11-20060816/>

BRICKLEY, Dan a GUHA, R. V. (2014). RDF Schema 1.1: W3C Recommendation 25 February 2014. *World Wide Web Consortium* [online]. © 2004-2014 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>

BUTLER, Howard, Martin DALY, Allan DOYLE, Sean GILLIES, Stefan HAGEN a Tim SCHAUB (2016). RFC 7946: The GeoJSON Format. *Internet Engineering Task Force (IETF): ISSN: 2070-1721* [online]. August 2016 [cit. 2018-02-19]. Dostupné z: <https://tools.ietf.org/html/rfc7946>

CYGANIAK, Richard, David WOOD a Markus LANTHALER (2014). RDF 1.1 Concepts and Abstract Syntax: W3C Recommendation 25 February 2014. *World Wide Web Consortium* [online]. © 2004-2014 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

ČERBA, Otakar (2017). Smart Points of Interest: Data Model. *SDI4APPS* [online]. 3. 4. 2017 [cit. 2018-02-19]. Dostupné z: http://sdi4apps.eu/spoi/doc/SPOI_data_model.pdf

ČERBA, Otakar a Tomáš MILDORF (2016). Smart Points of Interest: Big, Linked and Harmonized Spatial Data. *The GMU Geocrowdsourcing Testbed* [online]. [cit. 2018-02-19]. Dostupné z: http://geo.gmu.edu/AutoCarto2016/Cerba_and_Mildorf.pdf

ČERBA, Otakar, Michal KEPKA, Jan MACURA a Jáchym KELLAR (2017). Smart Points of Interest. *GIS Ostrava 2017: Geoinformatika v pohybu* [online]. Ostrava, 22. – 24. 3. 2017

[cit. 2018-02-19]. Dostupné z: http://gisak.vsb.cz/GIS_Ostrava/GIS_Ova_2017/sbornik/papers/gis2017589c453c426d5.pdf

Česká Republika. Zákon č. 121/2000 Sb. ze dne 7. dubna 2000 o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). [online] Sbírka zákonů České republiky. 2000. [cit. 2018-02-19]. ISSN 1211-1244. Dostupné také z: <http://www.psp.cz/sqw/sbirka.sqw?r=2000&cz=121>

DBpedia Maintainers and Contributors (2008). About: <http://dbpedia.org/ontology/>. *DBpedia* [online]. 17. 11. 2008, 16. 1. 2017 [cit. 2018-02-19]. Dostupné z: <http://dbpedia.org/ontology/>

DCMI Usage Board (2012a). DCMI: DCMI Metadata Terms. *Dublin Core Metadata Initiative* [online]. 14. 6. 2012, 11. 1. 2016 [cit. 2018-02-19]. Dostupné z: <http://dublincore.org/documents/dcmi-terms/>

DCMI Usage Board (2012b). DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description. *Dublin Core Metadata Initiative* [online]. 14. 6. 2012, 5. 11. 2013 [cit. 2018-02-19]. Dostupné z: <http://dublincore.org/documents/dces/>

ESRI Shapefile Technical Description (1998). *ESRI: An ESRI White Paper* [online]. July 1998 [cit. 2018-02-19]. Dostupné z: <https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

FAULKNER, Steve, Arron EICHOLZ, Travis LEITHEAD, Alex DANILO a Sangwhan MOON (2017). HTML 5.2: W3C Recommendation, 14 December 2017. *World Wide Web Consortium* [online]. © 2017 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/TR/html/>

FOSTER, Dan. GPX 1.1 Schema Documentation. *Topografix: GPX: the GPS Exchange Format* [online]. [cit. 2018-02-19]. Dostupné z: <http://www.topografix.com/GPX/1/1/>

GOODCHILD, Michael F. (1992). Geographical data modeling. *Computers & Geosciences*, 18(4), 401-408.

HARRIS, Steve a Andy SEABORNE (2013). SPARQL 1.1 Query Language: W3C Recommendation 21 March 2013. *World Wide Web Consortium* [online]. © 2013 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>

HUSÁRIKOVÁ, Světlana (2006). *ETL procesy* [online]. Praha, 2006. Bakalářská práce. Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra informačních technologií. [cit. 2018-02-19]. Dostupné z: <https://vskp.vse.cz/eid/638>

ISO 15836-1:2017(en): Information and documentation — The Dublin Core Metadata Element Set — Part 1: Core elements (2017). *International Organization for Standardization: ISO 15836-1:2017(en)* [online]. May 2017 [cit. 2018-02-19]. Dostupné z: <https://www.iso.org/obp/ui/#iso:std:iso:15836:-1:ed-1:v1:en>

ISO 32000-1:2008: Document management -- Portable document format -- Part 1: PDF 1.7 (2008). *International Organization for Standardization: ISO 32000-1:2008* [online]. 2008 [cit. 2018-02-19]. Dostupné z: <https://www.iso.org/standard/51502.html>

ISO/IEC 9075-1:2011(E): Information technology — Database languages — SQL — Part 1: Framework (SQL/Framework) (2011). *INTERNATIONAL STANDARD: ISO/IEC* [online]. 15. 12. 2011 [cit. 2018-02-19]. Dostupné z: http://standards.iso.org/ittf/PubliclyAvailableStandards/c053681_ISO_IEC_9075-1_2011.zip

JANECKA, Karel, Otakar Cerba, Karel Jedlicka a Jan Jezek (2013). Towards Interoperability Of Spatial Planning Data: 5-Steps Harmonization Framework. *International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management*, 1, 1005.

KARASOVÁ, Lucie (2018). Nejvhodnější aplikace pro získání souřadnic zájmových bodů pro databázi Smart Points of Interest. *Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra geomatiky: GeoWeb: Semestrální práce* [online]. 5. 1. 2018. [cit. 2018-02-19]. Dostupné z: <https://docs.google.com/document/d/18bGMThW120kqOEx5J3i3zoyXJnXkAc0ZliZj59sE6wY/edit>

KHANDELWAL, Ankesh, Libby MILLER a Dan BRICKLEY, et al. (2014). FOAF Vocabulary Specification 0.99: Namespace Document 14 January 2014 - Paddington Edition. *FOAF project: xmlns.com* [online]. © 2000-2014 [cit. 2018-02-19]. Dostupné z: <http://xmlns.com/foaf/spec/20140114.html>

KIRÁL, Martin (2017). *Objektové vazby pro prostorová Linked Data* [online]. Plzeň, 2017. Diplomová práce. *Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra*

matematiky. [cit. 2018-02-19]. Dostupné z: [https://otik.zcu.cz/bitstream/11025/26816/1/Diplo
mova_Prace.pdf](https://otik.zcu.cz/bitstream/11025/26816/1/Diplo
mova_Prace.pdf)

KUHN, Werner, Tomi KAUPPINEN a Krzysztof JANOWICZ (2014). Linked data-A paradigm shift for geographic information science [online]. [cit. 2018-02-19]. In *Geographic Information Science* (pp. 173-186). Springer International Publishing. Dostupné z: <http://geog.ucsb.edu/~jano/GIScience2014.pdf>

KUNZE, John A. a Thomas BAKER (2017) .RFC 5013: The Dublin Core Metadata Element Set. *IETF: RFC* [online]. August 2017 [cit. 2018-02-19]. Dostupné z: <https://tools.ietf.org/html/rfc5013>

LONGLEY, Paul A., Michael F. GOODCHILD, David J. MAGUIRE a David W. RHIND (2001). *Geographic information system and Science* [online]. [cit. 2018-02-19]. England: John Wiley & Sons, Ltd. 2. edice. Dostupné z: http://scholar.cu.edu.eg/?q=ahmedabdelhalim/files/geographic_information_systems_and_science.pdf

MILES, Alistair a BECHHOFFER, Sean (2009). SKOS Simple Knowledge Organization System Reference: W3C Recommendation 18 August 2009. *World Wide Web Consortium* [online]. © 2009 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>

OGC: Geography Markup Language. *Open Geospatial Consortium: Extensible Markup Language* [online]. © 2018 [cit. 2018-02-19]. Dostupné z: <http://www.opengeospatial.org/standards/gml>

OGC: KML. *Open Geospatial Consortium: KML* [online]. © 2018 [cit. 2018-02-19]. Dostupné z: <http://www.opengeospatial.org/standards/kml>

OGC: Web Feature Service. *Open Geospatial Consortium: Web Feature Service* [online]. © 2018 [cit. 2018-02-19]. Dostupné z: <http://www.opengeospatial.org/standards/wfs>

OSD: Introduction to Open Software Description (Internet Explorer) (2017). *Microsoft Corporation: Microsoft Docs* [online]. 15. 8. 2017 [cit. 2018-02-19]. Dostupné z: [https://docs.microsoft.com/en-us/previous-versions/windows/internet-explorer/ie-developer/platform-apis/aa740951\(v=vs.85\)](https://docs.microsoft.com/en-us/previous-versions/windows/internet-explorer/ie-developer/platform-apis/aa740951(v=vs.85))

PEREGO, Andrea a Michael LUTZ (2015). ISA Programme Location Core Vocabulary: Second version in w3.org/ns space. *World Wide Web Consortium* [online]. © 2012-2015, 23. 3. 2015 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/ns/locn>

PERRY, Matthew a John HERRING (2012) GeoSPARQL: A Geographic Query Language for RDF Data. *Open Geospatial Consortium: Implementation Standard* [online]. © 2012 [cit. 2018-02-19]. Dostupné z: <http://www.opengeospatial.org/standards/geosparql>

REYNOLDS, Dave (2014). The Organization Ontology: W3C Recommendation 16 January 2014. *World Wide Web Consortium* [online]. © 2012-2014 [cit. 2018-02-19]. Dostupné z: <https://www.w3.org/TR/2014/REC-vocab-org-20140116/>

SHAFRANOVICH, Yakov (2005). RFC 4180: Common Format and MIME Type for Comma-Separated Values (CSV) Files. *Internet Engineering Task Force (IETF)* [online]. October 2005 [cit. 2018-02-19]. Dostupné z: <https://tools.ietf.org/html/rfc4180>

SHEKHAR, Shashi, Mark COYLE, Brajesh GOYAL, Duen-Ren LIU a Shyamsundar SARKAR (1997). Data models in geographic information systems [online]. [cit. 2018-02-19]. *Communications of the ACM*, 40(4), 103-111. Dostupné z: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.154.6834&rep=rep1&type=pdf>

VATANT, Bernard a WICK, Marc (2012). The GeoNames Ontology. *Geo Semantic Web: GeoNames Ontology* [online]. 2012 November [cit. 2018-02-19]. Dostupné z: <http://www.geonames.org/ontology/documentation.html>

Obsah přiloženého CD

Text diplomové práce

- DP Karasová Lucie.pdf

Obrázky

Veškeré obrázky v digitální podobě číslované dle této práce.

- 1) Obr. 1 – Datový model SPOI k 3. 4. 2017 (graficky upravený).png
- 2) Obr. 2 – Rozšířený datový model SPOI dle Ing. Martina Kiráľa.png
- 3) Obr. 3 – Diagram postupu doplňování databáze SPOI.svg
- 4) Obr. 4 – Snímek mapového okna s vybraným bodem zájmu (Lokál Pod Divadlem).png
- 5) Obr. 5 – Snímek hlavní stránky Smart tourism guide.png
- 6) Obr. 6 – Snímek Smart tourism guide po vyhledání POI bodů v městě Plzni.png
- 7) Obr. 7 – Snímek dotazovacího okna Virtuoso SPARQL Query Editor.png
- 8) Obr. 8 – Snímek výstupu dotazu ve Virtuoso SPARQL Query Editor – tabulka.png
- 9) Obr. 9 – Snímek z konvertoru Convertcsv.com (CSV – GeoJSON).png
- 10) Obr. 10 – Vytvoření nové vrstvy ze souboru CSV v programu QGIS.png
- 11) Obr. 11 – Konverze souboru CSV v programu CSV Converter.png
- 12) Obr. 12 – ETL proces.svg
- 13) Obr. 13 – Snímek vizualizovaných dat v editoru Mapshaper.png
- 14) Obr. 14 – Snímek informací zobrazených v editoru Mapshaper.png
- 15) Obr. 15 – Snímek prázdného listu v aplikaci Microsoft Excel Online.png
- 16) Obr. 16 – Snímek prázdného listu v aplikaci Tabulky Google.png
- 17) Obr. 17 – Snímek prázdného listu v aplikaci WPS Office – Spreadsheets.png
- 18) Obr. 18 – Publikování vrstev na serveru GeoServer.org.png
- 19) Obr. 19 – Stažená data z webové služby WFS z GeoServer.org v programu QGIS.png
- 20) Obr. 20 – Hlavička tabulky pro evidenci POI bodů – povinné atributy.png
- 21) Obr. 21 – Tabulka s měněnými hodnotami.png