

## Příloha A: Analytická část

### Úvod

Tento dokument vznikl a slouží jako referenční text pro návrhovou část projektu TB0500MV003 Vypracování certifikované metodiky pro publikování prostorových informací ve formě otevřených dat. Cílem analytických činností je podrobné zmapování stávající situace v oblasti publikování prostorových dat v otevřené podobě. Hlavní důraz je kladen na specifika prostorových dat, která vyžadují mnohdy odlišný přístup pro jejich lepší využití třetími stranami (komerční sféra a občané), především vzhledem k jejich objemnosti a využitelnosti, než je tomu u neprostorových dat.

V souladu s Přílohou B.1.1. nabídky na zpracování veřejné zakázky je text rozdělený na Analýzu stávající situace a Analýzu a definici uživatelských požadavků, přičemž obě hlavní kapitoly jsou dále detailně členěny.

Analytická část nepředstavuje kompletní a vyčerpávající popis struktury otevřených dat v České republice a ve světě, ale pouze výběr zajímavých příkladů dobré praxe, existujících systémů a technologií. Dokument také není uzavřený. Autoři se rozhodně nebrání doplněním ze stran odborné veřejnosti (z tohoto důvodu je dokument nad rámec povinností zpracovatele zakázky zveřejněný). Zároveň také předpokládají, že v průběhu tvorby návrhu metodiky a zpracování modelových ukázek bude docházet k postupné aktualizaci textu.

Poznámka vložená po oponentním řízení: *Analytická část metodiky není její přímou součástí, ale je připojena jako příloha. Byla určena především pro interní účely autorů. Původně neměla být publikována, ale měla sloužit jako podkladový materiál pro vlastní metodiku. O publikování bylo rozhodnuto až na základě kladných ohlasů a přímých výzev během kontrolních dnů. Analýza byla zaměřena především na příklady dobré praxe a vybraná řešení. Z tohoto důvodu nelze očekávat komplexní rešerši a její zhodnocení – tyto úkony by byly jednak nad rámec projektu a navíc by bylo vzhledem k cílům metodiky kontraproduktivní. Metodika měla být zaměřená na hledání cest, jak publikovat prostorová data jako otevřená, nikoli na to, proč je touto formou nepublikovat. To je důvod, proč byly do analýzy zařazeny zdánlivě vzdálené a s Českou republikou nesouvisející příklady jako Kolumbie nebo Tchaj-wan. Navíc v dnešním globalizovaném světě, jehož jsou otevřená prostorová data součástí, není vzdálenosti signifikantní.*

## **Analýza stávající situace**

Analýza stávající situace popisuje současný stav z hlediska publikace otevřených prostorových dat. Dále následuje popis aktuální situace publikování otevřených prostorových dat v České republice. Jednotlivé produkty, projekty, systémy apod. jsou v textu popsány (včetně odkazu na příslušné webové stránky) a pokud je to vhodné (například v případě tezaurů a jiných sémantických nástrojů), je připojena také SWOT analýza ukazující slabé a silné stránky, stejně jako příležitosti a hrozby.

## **Analýza “best practices” v jiných evropských zemích s využitím výsledků EU projektů**

Vzhledem k tomu, že příklady dobré praxe („best practices“) tvoří klíčovou část, a tudíž značně rozsáhlou část textu, bylo zvoleno pro snadnější orientaci členění na analýzu přístupu Evropské Unie a vybraných států k problematice prostorových otevřených dat, průzkum mezinárodních projektů zaměřených na výše uvedené téma a aktivity mezinárodních standardizačních organizací.

## **Politiky, EU direktivy, nadnárodní iniciativy**

### **INSPIRE**

Webové stránky INSPIRE (INfrastructure for SPatial InfoRmation in Europe) v České republice popisují tuto aktivitu jako „směrnice Evropské komise a Rady kladoucí za cíl vytvořit evropský legislativní rámec potřebný k vybudování evropské infrastruktury prostorových informací. Stanovuje obecná pravidla pro založení evropské infrastruktury prostorových dat zejména k podpoře environmentálních politik a politik, které životní prostředí ovlivňují. Hlavním cílem INSPIRE je poskytnout větší množství kvalitních a standardizovaných prostorových informací pro vytváření a uplatňování politik Společenství na všech úrovních členských států.“

Otevřená data sice nejsou v INSPIRE přímo zmíněna, ale je potřeba zdůraznit, že standardizace, integrace a harmonizace na poli prostorových dat je nutnou podmínkou pro otevírání takových dat, protože otevřenost nespočívá pouze v legislativních opatřeních, ale také v procesu, který data převede do srozumitelné a sdílitelné podoby.

V České republice je implementace INSPIRE popsána v dokumentu Strategie implementace INSPIRE. Tento dokument shrnuje technické a legislativní dokumenty týkající se INSPIRE. V dokumentu se sice přímo o otevřených datech nehovoří, ale je třeba si uvědomit, že otevřená prostorová data publikovaná povinnými poskytovateli budou muset splňovat požadavky INSPIRE a i požadavky kladené na otevřená data. Příkladem podobné aktivity kombinující INSPIRE datové modely a registry s pětihvězdičkovým klasifikačním systémem pro hodnocení otevřených dat je Open Transport Map vyvíjená v projektu OpenTransportNet (viz přehled projektů).

### **Copernicus**

Copernicus je klíčový evropský program, pro poskytování informací o životním prostředí a bezpečnosti. Copernicus je přímo financován Evropskou komisí. Data Sentinel jsou chápána jako informace veřejného sektoru (PSI), což znamená, že data jsou sbírána veřejným sektorem pro vlastní účely. Copernicus poskytuje příležitost pro vytvoření trhu pro služby DPZ a to i jako komerční služby i mimo

území EU. V souladu s datovou a informační politikou Evropské komise, program Copernicus poskytuje všem uživatelům volný, plný a otevřený přístup k údajům v oblasti životního prostředí. Uživatelé mohou získávat buď přímo originální data Sentinel nebo mohou získávat zpracované informace ze služeb Copernicus. Toto probíhá bez nutnosti registrace.

Kromě údajů z družic Sentinel mohou uživatelé Copernicus mít rovněž přístup za určitých podmínek k údajům z jiných družicových misí označované jako „podpůrné mise“. Podpůrné mise jsou rozděleny do pěti skupin, které odrážejí charakteristiky hlavních druhů misí. Většina údajů poskytnutých různými podpůrnými misemi jsou distribuovány prostřednictvím Evropské kosmické agentury (ESA). Mají podobu datových sad a jsou poskytovány uživatelům prostřednictvím přístupu k datovým službám. Obecně lze konstatovat, že služby Copernicus jsou poskytovány mezi jednou a třemi hvězdičkami, Hvězdičkový systém není pro rastrová data příliš vhodný. V současné době se předpokládá budování cloud řešení na principu Public Private Partnership.

## GEOSS

Skupina GEO (Group on Earth Observation) je světovou iniciativou, která si klade za cíl vybudovat globální systém, který v sobě bude propojovat systémy pozorování Země - GEOSS - Global Earth Observation System of Systems). Iniciativa GEO vznikla jako reakce na výzvu Světového summitu o udržitelném rozvoji v roce 2002 a států G8. V důsledku rostoucího potenciálu pozorování Země, který by měl sloužit jako nástroj pro podporu rozhodování, byla při společných diskuzích definována za klíčovou mezinárodní spolupráce. Skupina GEO proto funguje na bázi spolupráce mezi vládami jednotlivých zemí a mezinárodními institucemi. Tím GEO vytváří platformu pro uskutečňování nových projektů a koordinaci vynakládání investic a plánování strategií.

Hlavním principem GEOSS je, že sociální výhody plynoucí z pozorování Země mohou být plně dosaženy pouze prostřednictvím sdílení dat, informací, znalostí, produktů a služeb. GEO proto podporuje základní principy pro sdílení dat a rozšiřuje tyto trendy otevřených dat celosvětově. GEOSS prosazuje následující zásady sdílení údajů:

- Data, metadata a produkty budou sdíleny jako otevřená data bez omezení na opětovné použití, a to za podmínek registrace a opětovné publikování výstupů.
- Pokud mezinárodní nebo národní politiky nebo právní předpisy brání sdílení dat jako otevřená data, data by měla být dána k dispozici s minimálními omezeními na pokrytí nákladů na rozmnožování a šíření.
- Všechny sdílené datové produkty a metadata budou dány k dispozici s minimálním časovým zpožděním.

GEOSS prozatím neřeší publikování v souladu s hvězdičkovým systémem. Hlavní princip je budování GEOSS brockage, který bude on-fly harmonizovat data.

### „Best practices“ v jiných evropských zemích

Podle Global Open Data Index<sup>1</sup>, který publikuje organizace Open Knowledge, se mezi prvními pěti státy z hlediska úrovně publikování otevřených dat umístily Tchaj-wan, Velká Británie, Dánsko, Kolumbie a Finsko (v tomto pořadí). Česká republika obsadila jedenadvacáté místo. Global Open Data Index nabízí i pohled na prostorová data, budeme-li mezi ně počítat kategorie National Map, Location datasets a Land Ownership<sup>2</sup>.

Následující tabulky ukazují výše uvedená tři témata a jejich publikování v ČR a v pěti nejvýznamnějších státech z hlediska publikování otevřených dat. Všechny tabulky mají společnou legendu, kde zelená barva znamená „ano“, červená „ne“ a žlutá, že informace není dostupná nebo jasná. Údaje jsou publikovány přesně podle informací z Global Open Data Index, ačkoli autoři si jsou vědomi, že se v datech nachází nepřesnosti (například v příkladě možnosti nákupu státního mapového díla v ČR).

---

<sup>1</sup>Výsledky jsou z konce května 2016.

<sup>2</sup>Prostorový charakter má však větší množství témat datových sad jako například znečištění ovzduší nebo výsledky voleb.

## National map

	Tchaj-wan	Velká Británie	Dánsko	Kolumbie	Finsko	Česká republika
Otevřená licence						
Strojová čitelnost						
Zdarma						
Možnost nákupu						
Aktualizace						
Online						
Digitální						
Veřejné						
Existence dat						

## Location datasets

	Tchaj-wan	Velká Británie	Dánsko	Kolumbie	Finsko	Česká republika
Otevřená licence						
Strojová čitelnost						
Zdarma						
Možnost nákupu						
Aktualizace						
Online						
Digitální						
Veřejné						
Existence dat						

## Land ownership

	Tchaj-wan	Velká Británie	Dánsko	Kolumbie	Finsko	Česká republika
Otevřená licence						
Strojová čitelnost						
Zdarma						
Možnost nákupu						
Aktualizace						
Online						
Digitální						
Veřejné						
Existence dat						

Předchozí tabulky ukazují, že Česká republika z hlediska publikování otevřených prostorových dat nezaostává za nejlepšími státy. V případě publikování dat katastru dokonce většinu z nich (kromě Dánska a částečně Kolumbie) předstihuje.

Následující odstavce stručně popisují situaci v publikování prostorových otevřených v evropských státech z předchozího seznamu založeného na výsledcích Global Open Data Index.

### Velká Británie

URL: <https://www.ordnancesurvey.co.uk/opendatadownload/products.html>

Ordnance Survey jako hlavní producent prostorových dat na území Velké Británie nabízí obrovskou škálu otevřených produktů od státní mapy, přes různé typy hranic, až po geoteorii, poštovní směrovací čísla, adresy a rastrové velkoměřítkové produkty. Vektorové produkty jsou převážně k dispozici ve formátech ESRI Shapefile a GML.

### Dánsko

URL: <http://download.kortforsyningen.dk/> (data se nachází v různých sekcích stránek, které jsou však k dispozici pouze v dánštině); některá data poskytuje také dánská pošta.



Dánsko jako jediné ze zkoumaných zemí poskytuje všechny tři výše uvedené typy prostorových dat jako data otevřená. Konkrétně se jedná o Mapu Dánska v měřítku 1 : 200 000, včetně zgeneralizovaných vrstev menších měřítek a vybraných tematických vrstev (například Corine Land Cover). K dispozici je také náhled na katastrální mapu s možností stažení dat.

## **Finsko**

URL: <http://www.paikkatietoikkuna.fi/web/en/open-spatial-data>

Na webové stránce jsou k dispozici odkazy na otevřená data Geological Survey of Finland, Finnish Museum of Natural History, National Land Survey (Topographic Database, výškové modely, orthofoto snímky, laser scanning data), National Board of Antiquities, Environmental Administration, Helsinki Region Environmental Services Authority a další.

Výše uvedené webové stránky představují užitečný rozcestník s odkazy na vlastní data (ta jsou nabízena prostřednictvím různých technologií v různých formátech) i licenční podmínky.

## **Mezinárodní projekty**

### **ARE3NA**

URL: [http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-17action\\_en.htm](http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-17action_en.htm)

Hlavním cílem ARE3NA je stanovení referenční platformy obsahující softwarové komponenty a další materiály na podporu implementace směrnice INSPIRE. Mezi sekundární aktivity patří také podpora interoperability Digital Agenda for Europe a otevřených dat. Na webové stránce ARE3NA je možné nalézt důležité materiály z pohledu otevřených dat jako jsou webináře na téma RDF a perzistentní ID (ARE3NA's open webinars on RDF and PIDs for Location), dokumenty zaměřené na správu registrů nebo Study and guidelines on Geospatial Linked Data.

### **Citadel on the Move**

URL: <http://www.citadelonthemove.eu/>

Projekt Citadel on the Move byl zaměřený na využívání otevřených, převážně prostorových dat pro tvorbu inovativních mobilních aplikací. Z hlediska publikování otevřených dat jsou důležité dva aspekty projektu:

- Open Data Index - katalog otevřených dat publikovaných především jednotlivými městy zapojenými do projektu (z Česka se jedná o Plzeň a

Prahu). Data jsou k dispozici v různých formátech i datových modelech, proto představují ideální ukázkou mnoha podob otevřených dat.

- Nástroje a metodiky pro převod dat do podoby otevřených (také pro testování a publikování dat).

## **LOD2**

URL: <http://lod2.eu/>

LOD2 je jeden z dalších projektů EU zaměřených Linked Open Data. V rámci projektu bylo celkem vytvořeno 153 článků, včetně metodik (například Publishing and Consuming Linked Open Data with the LOD Statistical Workbench) a článků zaměřených na otevřená data (například Open Government Data Catalogs: Current Approaches and Quality Perspective, Towards an Open-Governmental Data Web nebo Linked Open Data Infrastructure for Public Sector Information: Example from Serbia). Z pohledu prostorových dat je důležitá vazba projektu na produkt LinkedGeoData, který představuje LOD variantu populární OpenStreetMap.

## **MELODIES**

URL: <http://www.melodiesproject.eu/>

Melodies je dalším z mnoha evropských projektů zaměřených na příležitosti vyplývající z používání otevřených dat. Partneři projektu se podílí na vývoji několika aplikací, které mohou být užitečné při publikování otevřených prostorových dat:

- Strabon - úložiště pro časoprostorová data ve formátu RDF.
- Silk - „link discovery framework“, využitelný při hledání vazeb v rámci linked data.
- Sextant - vizualizační nástroj pro práci s otevřenými propojenými prostorovými daty.
- Ontop Spatial - nástroj pro využití technik linked data pro data, která nejsou uložena v RDF.
- GeoTriples - nástroj umožňující transformaci prostorových dat (SHP, KML, GeoJSON) do RDF.

Z hlediska textový výstupů jsou pro účely tohoto úkolu zajímavé především dva dokumenty D11.1 Report on open data policies, progress and outlook (přístup k nástrojům i textům je z výše uvedené webové stránky projektu).

## **SDI4Apps**

URL: <http://sdi4apps.eu/>

V rámci projektu SDI4Apps vzniklo několik materiálů užitečných z hlediska publikování prostorových otevřených dat. Jedná se především o následující doku-

menty (dostupné z adresy <http://sdi4apps.eu/project-information/deliverables/>), které diskutují aspekty otevřených dat jako jsou sémantika nebo datové modely:

- 4.1.1 Harmonisation an Multilingual Tools - popis transformace dat, jejíž cílem jsou v některých případech otevřená data.
- 5.1 Initial Deployment and Methodology for Quality Assessment - dokument zabývající se kvalitou dat a způsobem jejího popisu a kontroly.
- 5.2.1 Semantic Annotation 1 - možnosti doplnění sémantických informací k existujícím datům.
- 6.1 Data Models and Platform Specifications for Single Pilots - popis datových struktur jednotlivých pilotních aplikací.

V rámci projektu také vznikají pilotní datové sady, které ve velké míře publikují nebo využívají otevřená data. Jako příklady takových otevřených datových produktů je možné uvést

- Smart Points of Interest (SPOI; [sdi4apps.eu/spoi](http://sdi4apps.eu/spoi)) - 27 miliónů tzv. bodů zájmu určených především pro cestovní ruch, ale také dopravu, státní správu nebo ochranu přírody, které jsou publikované jako tzv. pěti-hvězdičková linked open data pod licencí ODbL. Data využívají slovníky jako například RDFS, GeoSPARQL, FOAF, ISA Programme Location Vacabulary a další.
- Open Land Use (OLU; [sdi4apps.eu/open\\_land\\_use/](http://sdi4apps.eu/open_land_use/)) - evropská mapa vizualizující využití půdy v Evropě. V jednotlivých regionech jsou použity nejpodrobnější možné volně dostupné datové zdroje (například katastrální mapa v případě České republiky, nebo GRBgis pro Flandry). Datový model je navržený takovým způsobem, aby respektoval slovníky INSPIRE a byl transformovatelný do podoby tzv. pěti-hvězdičkových LOD.

## SmartOpenData

URL: <http://www.smartopendata.eu/>

Projekt byl zaměřený na vytvoření Linked Open Data infrastruktury pro veřejná a volně šiřitelná data z oblasti životního prostředí. Pro jednotlivé aspekty otevřených dat jsou k dispozici následující dokumenty <http://www.smartopendata.eu/public-deliverables>:

- Metadata - D3.1 Review of geographic resources metadata and related metadata standards
- Datové modely (na principu INSPIRE a LOD) - D3.2 SmartOpenData Initial Data Model, D3.4 The Final SmartOpenData Model
- Harmonizace dat - D3.3 Harmonization of data to SmartOpenData model. Initial iteration, D3.5 Harmonization of data to SmartOpenData model. Final iteration
- Sémantika - D4.1 Distributed Semantic Indexing Infrastructure

## OpenTransportNet

URL: <http://project.opentransportnet.eu/>

Projekt OpenTransportNet (OTN) je zaměřený především na oblast dopravy. Z hlediska otevřených prostorových dat jsou klíčové následující dokumenty a produkty:

- OTN Metadata Profile využívající pro prostorová data GeoDCAT-AP a INSPIRE metadata. Detailní informace, včetně způsobů implementace katalogizačních nástrojů CKAN a Mica, jsou dostupné v dokumentu D7.3 GI Innovation White Paper I. S metadaty souvisí také dokumenty D3.2 Geospatial Metadata Catalogue, D3.2 Geospatial Metadata Catalogue v2 (veškeré reporty projektu je možné stáhnout z adresy <http://project.opentransportnet.eu/otn/?q=deliverables>).
- V projektu je akcentována sémantika (D3.6 Semantic Report) a transformace dat (D4.4 Data Harmonisation and Integration, White Paper II: Data harmonisation, interoperability & licensing in Open Transport Net).
- V dokumentu White Paper II: Data harmonisation, interoperability & licensing in Open Transport Net jsou k dispozici informace ohledně licencování dat s důrazem na Open Data Licensing, včetně porovnání vhodných licencí.

## Aktivita mezinárodních organizací

### Aktivita JRC (Joint Research Centre)

URL: <http://inspire.ec.europa.eu/registry/>

Jak je uvedeno v sekci sémantické aspekty otevřených dat, číselníky, nomenklatury a další klasifikační systémy představují důležitý zdroj, který zajišťuje otevřenost dat především z hlediska významu přenášené informace a jeho sdílení. Z tohoto důvodu je klíčové publikování registrů INSPIRE, které zajišťují interoperabilitu mezi jednotlivými datovými sadami spadajícími pod směrnici INSPIRE, přičemž jejich použití je mnoho širší. Původně byly registry součástí pouze datových modelů specifikací INSPIRE a publikovány jako texty nebo součástí UML diagramů. Později byly jednotlivé codelisty a číselníky publikovány ve strukturovaných formátech. V současnosti jsou k dispozici ve velkém množství neproprietárních otevřených formátů jako jsou XML, CSV, JSON, Atom a samozřejmě RDF. Další obrovskou výhodou tohoto zdroje je kromě legislativního zakotvení také lokalizace do českého jazyka.

Kromě registrů JRC nabízí také open data katalog čítající k 31. květnu 2016 596 datových sad, z nichž většinu můžeme označit jako data prostorová.

### Aktivita OGC (Open Geospatial Consortium)

Podobně jako v případě W3C je stěžejním přínosem organizace OGC pro oblast otevřených prostorových dat především tvorba a údržba standardů. OGC je v této analýze samostatně zmiňováno především proto, že téměř veškerá jeho činnost je zacílena do domény prostorových dat a informací. Publikování prostorových dat s různým stupněm otevřenosti je spojeno s formáty GML a KML, s webovými službami jako například WMS, WFS, WCS nebo standard GeoSPARQL.

Na tomto místě je také zapotřebí podotknout, že kromě online dokumentů a zdrojů otevřených dat, jsou důležitá i osobní setkání a diskuzní fóra, kde dochází k rozvoji spolupráce mezi různými organizacemi a projekty. Příkladem takové aktivity, na které se podílelo OGC, W3C, Ordnance Survey, Google a projekt SmartOpenData byl workshop Linking Geospatial Data (<https://www.w3.org/2014/03/lgd/>).

### **Aktivity W3C (World Wide Web Consortium)**

W3C je klíčovým hráčem na poli otevřených dat. Vytváří a spravuje základní standardy, které mají spojitost otevřenými daty ve smyslu pětihvězdičkového klasifikačního systému. Konkrétně se jedná o XML další standardy založené na tomto metajazyku. Pro otevřená data jsou důležité především formáty RDF, RDFS, OWL, SPARQL nebo SKOS. W3C se také podílí na vývoji mnoha důležitých slovníků a ontologií (například vCard nebo DCAT).

Dalším důležitým zdrojem jsou články a textové dokumenty. Pro účely publikování otevřených dat je možné například uvést W3C Working Draft Publishing Open Government Data, W3C Working Group Note Best Practices for Publishing Linked Data nebo text W3C Data Activity: Building the Web of Data, který odkazuje na mnoho externích zdrojů (G8 Open Data Charter, PSI Directive nebo na aktivity W3C v oblasti eGovernmentu a sémantického webu. Také stránka LinkedData na wiki W3C (<https://www.w3.org/wiki/LinkedData>) obsahuje obrovské množství odkazů na články z oblasti Linked Open Data. Důležitým zdrojem z hlediska odborné terminologie může být i W3C Linked Data Glossary.

V oblasti otevřených dat působí řada dalších organizací a institucí, jejichž analýza je mimo rozsah tohoto projektu a dokumentu. Jedná se například o Open Data Institute, Open Data Foundation, Open Data for Development (OD4D) nebo standardizační organizace jako jsou ISO, IETF nebo ECMA. Podobně jako u jiných explicitně nezmíněných zdrojů vhodných pro publikování otevřených prostorových budou i výstupy výše uvedených subjektů zohledněny v návrhové části.

## **Analýza a příklady úspěšně fungujících systémů postavených na otevřených prostorových datech v ČR i v EU**

### **Portál [opendata.gov.cz](https://opendata.gov.cz) (a další aktivity MV ČR)**

Portál [opendata.gov.cz](https://opendata.gov.cz) je určen orgánům veřejné správy a popisuje doporučovaný postup otevírání dat a nabízí vzorové publikační plány, které doporučují (pro jednotlivé typy orgánů veřejné správy) konkrétní datové sady k publikaci ve formátu otevřených dat. Součástí portálu je Národní katalog otevřených dat (NKOD), který obsahuje řadu datových sad včetně prostorových dat (např. data RÚIAN).

SWOT analýza NKOD:

- Silné stránky: dostupná data o státní správě; portál obsahuje řadu užitečných instrukcí pro otevření dat.
- Slabé stránky: obtížné vyhledávání dat (např. chybí štítky, témata apod), data nejsou kategorizována, nelze vyhledávat podle datového formátu, stupně otevřenosti apod.
- Příležitosti: Jednotný bod pro přístup, možnost tvorby a distribuce aplikací na daty
- Hrozba: Nepřehlednost a komplikované vyhledávání na NKOD může vést ke ztrátě důvěry v myšlenku otevřených dat. Data mohou být uveřejněná, ale stále velice těžce nalezitelná.

### **Portál <https://data.gov.uk/>**

<https://data.gov.uk/> je projekt Velké Británie zaměřený na zajištění dostupnosti státních (neosobních) dat formou otevřených dat (open data). Datový portál [data.gov.uk](https://data.gov.uk/) funguje od roku 2009 a jeho prostřednictvím jsou publikována data většiny ministerstev a úřadů. Z hlediska prostorových dat je klíčovým poskytovatelem organizace Ordnance Survey.

Specifika prostorových dat:

Hlavní dodavatel prostorových dat je Ordnance Survey (národní mapovací agentura).

Na portále lze nalézt následující datové sady v různých formátech (v závorce je uveden počet hvězdiček vyjadřující stupeň otevřenosti):

- WMS – 4099 položek (uváděno jako \*\*\*)
- WCS – 3961 položek (uváděno jako \*\*\*)
- WFS – 1355 položek (uváděno jako \*\*\*)
- SHP – 301 položek (uváděno jako \*\*\*)
- CSV

Portál dále obsahuje 146 položek RDF (většinou bez prostorové informace).

SWOT analýza portálu:

- Silné stránky: přehlednost, velké množství dat, utříděnost pojmů (API, WMS, RDF, SPARQL)
- Slabé stránky: chybí prostorová data v RDF, špatné prostorové vyhledávání
- Příležitosti: Jednotný bod pro přístup, možnost tvorby a distribuce aplikací na daty
- Hrozba: zneužití dat

## **Analýza technické a sémantické stránky publikace otevřených prostorových dat**

### **Technická stránka publikace otevřených prostorových dat**

Pro publikaci otevřených dat lze využít několika softwarových nástrojů, které umožňují správu dat i metadat a zároveň poskytují uživatelské funkce a vyhledávání. V této části se zaměříme na nástroje používané ve výše zmíněných projektech a nástroje využívané v úspěšně fungujících systémech.

### **Obecné platformy**

#### **CKAN**

Comprehensive Knowledge Archive Network (CKAN) je software určený ke správě a uložení a distribuci otevřených dat. Software je vydán pod open source licenci (GNU GPL v3.0.). CKAN je vyvíjen Open Knowledge Foundation.

Po technické se CKAN skládá ze serverové části v jazyce Python a webové klientské části v jazyce HTML a Javascript. Databázové úložiště je realizováno pomocí databáze PostgreSQL.

CKAN je obecným nástrojem pro správu jakýchkoliv otevřených dat, včetně prostorových, avšak standardy spojené s prostorovými jsou dostupné pouze ve formě pluginů a rozšíření. V současnosti je dostupný plugin pro DCAT a plugin pro import metadat ISO 19139. Platforma dále umožňuje uložení dat v databázi PostGIS.

Systém CKAN je široce rozšířený a je nasazený např. na portálu <https://data.gov.uk/>.

### **Prostorová data**

#### **Micka**

Micka je software vyvíjen firmou HS-RS s.r.o. určený pro správu metadat prostorových dat a budování prostorové infrastruktury (SDI). Tento systém obsahuje nástroje pro editaci a správu metadat prostorových dat, webových služeb a dalších zdrojů (dokumenty, webové stránky apod.), jejich vyhledávání na internetu zobrazení v mapě nebo stáhnutí na lokální počítač.

Micka je kompatibilní s normami závaznými pro budování Evropské prostorové infrastruktury (INSPIRE), proto je připravena na propojení s dalšími uzly připravované sítě katalogů metadat (testováno proti pilotnímu Evropskému geoportálu).

Hlavní vlastnosti platformy Micka:

- Podpora ISO 19115 / 19119 / 19139 (metadata prostorových dat a služeb)
- Podpora ISO 15836 (metadata Dublin Core)



- Podpora OGC CSW 2.0.2 (katalogová služba) včetně transakcí
- Podpora metadatového profilu INSPIRE
- Podpora INSPIRE stahovacích služeb pomocí ATOM
- Podpora GeoDCAT-AP, GeoRSS, GeoJSON, OAI-PMH, MARC21
- Harvesting z OGC služeb (WMS, WFS, ...), CSW katalogů, souborů ISO 19139,...
- Vytváření uživatelských profilů metadat
- Webové prostředí pro editaci metadat
- Vícejazyčné prostředí
- Kontextová nápověda
- Importy metadat z ESRI ArcCatalog, ISO 19139, CKAN
- Mapový prohlížeč pro vytváření mapových kompozic na základě webových služeb
- Správa katalogu objektů (feature katalogů) pro popis jednotlivých objektů datové sady a jejich atributů (ISO 19110)
- Podpora tématických a prostorových slovníků (thesaurus, gazetteer)
- Práce s thesauri (GEMET, ...), registrem INSPIRE apod.
- Zobrazení změn pomocí RSS kanálu
- Úprava vzhledu a funkčnosti pomocí uživatelských šablon
- Hromadná kontrola kompatibility záznamů se standardem INSPIRE

#### GeoNetwork

GeoNetwork je aplikace určená pro katalogizaci a management prostorových dat. Umožňuje editaci metadat a vyhledávací funkcionalitu, včetně mapového prohlížeče. Z hlediska metadatových standardů jsou podporovány specifikace ISO19115/119/110 a Dublin Core.

GeoNetwork je k dispozici pod OpenSource licencí GPL v2 license. GeoNetwork umožňuje harvestování metadat z řady zdrojů zahrnující:

- OGC-CSW 2.0.2 ISO Profile,
- Web Accessible Folders,
- ESRI GeoPortal,
- Další GeoNetwork katalogy.

#### Sémantická stránka publikace otevřených prostorových dat

\*Jak uvádí Mylopoulos (2004) „Tři největší problémy při používání geografických dat jsou – Sémantika, sémantika a sémantika“.\*

Definice otevřených dat vyžaduje dostupnost dat, používání standardů s volně dostupnou specifikací a užívání dat s minimem omezení a při vynaložení minima možných nákladů. Tyto požadavky velice úzce souvisí se sémantickou stránkou prostorových dat, protože právě doplnění prvků explicitní sémantiky je spojené se standardizací a širší využitelností dat, včetně snižování nákladů (v tomto konkrétním případě se jedná především o náklady souvisejících s délkou zpracování dat a odstraňováním chyb vzniklých díky nedostatkům plynoucím z využívání

implicitní a neformalizované sémantiky). Aby bylo možné co nejefektivněji využívat data i aplikace, je nezbytné pochopit význam datových sad v maximální možné míře. K tomu slouží nejen studium oborů, které data produkují, ale také využívání významových (sémantických) informací.

Termín „sémantika“ je v tomto textu užíván ve smyslu informačních technologií (nikoli jako pojem z lingvistiky nebo logiky) – „mapování objektu reálného světa a modelu reprezentujícího skutečný objekt, přičemž tento model může být reprezentován v informačním systému například jako objekt nebo položka v databázi. Toto mapování zajišťuje sémantiku modelu prostřednictvím popisu nebo identifikace významu prvků...“ (Sheth, 1997).

Jak již bylo dříve uvedeno pojem „sémantická stránka publikace otevřených prostorových dat“ je spojený především s explicitní sémantikou ve formě propojení dat na standardizované definice, popisy a vysvětlení (s uvedením zdroje), publikováním metadat a kritické používání existujících katalogů dat, slovníků, číselníků, taxonomií, nomenklatur a ontologií. To je důležité především v případě takových prostorových dat, která obsahují odborné a expertní informace, ale jsou přitom používána širokou veřejností nebo odborníky z jiných oborů, což je přesně případ většiny prostorových dat.

Zdroje formálních explicitních sémantických informací (slovníky a tezaury) jsou uvedeny v další části této analytické zprávy (Existující slovníky a tezaury užívané v souvislosti s otevřenými prostorovými daty).

Již dříve bylo uvedeno, že doplňování explicitních sémantických informací by mělo být kritické. To znamená ověřování správnosti (aktuálnosti z hlediska aktuálního stavu poznání, korektnosti překladu, vazby na legislativní dokumenty a národní i mezinárodní normy apod.), uvádění zdrojů a komunikace s autory nebo správci jednotlivých zdrojů sémantických informací.

Výhody používání sémantiky shrnuje následující seznam:

- Obecně zvyšuje kvalitu a použitelnost prostorových dat.
- Urychlení a zefektivnění procesu pochopení významu dat a informací.
- Omezení omylů při zpracování a vizualizaci dat.
- Možnost automatizovaného zpracování informací, včetně harmonizace, mapování a integrace.
- Umožňuje jednoznačnou komunikaci na různých úrovních.

Používání sémantických informací výrazně urychlí proces poznávání dat a díky tomu bude možné zamezit nejrozličnějším nesrovnalostem při interpretaci a zpracování dat (konflikty mezi datovými sadami, omyly při využívání vizualizačních prostředků apod.). Příkladem nepochopení významu a smyslu dat může být například zpracování dat kvalitativního charakteru pomocí kartografických vyjadřovacích prostředků umožňujících znázornění kvantity. Obecně je možné doplnění a používání sémantických informací chápat jako snížení nejistoty při jakémkoli nakládání s daty, včetně prostorových dat. Zároveň může docházet ke zvýšení kvality rozhodovacích procesů závislých na datech obsahujících

sémantické informace.

Sémantická informace v existujících systémech je dostupná pouze v omezeném množství a kvalitě. V zásadě se jedná o čtyři stávající způsoby připojení sémantické informace k datům (viz seznam níže), kdy maximálně v poslední případě lze říci, že se jedná o způsob vycházející z moderních principů otevřených propojených dat.

1. Metadata.
2. Datové modely.
3. Vazba na slovníky, které nepublikují položky ve formě LOD.
4. V případě dat kompatibilních s INSPIRE vazba na INSPIRE registry.

## **Analýza existujících slovníků a tezurů užívané v souvislosti s otevřenými prostorovými daty**

V této části jsou publikovány SWOT analýzy jednotlivých produktů (slovníků a tezurů), které by mohly sloužit jako zdroj sémantických informací pro publikování otevřených prostorových dat. Slovníky a tezaury jsou seřazeny v abecedním seznamu. Tento seznam není zcela vyčerpávající (chybí například tezaury z oblasti knihovnictví jako je ECLAS (European Commission Libraries Catalogue), ale důležité slovníky a tezaury pro prostorová data jsou zmíněné beze zbytku. Příklady publikování Linked data (resp. využívání a sdílení jednotlivých slovníků a tezurů) jsou k dispozici také v článcích Čerba & Jedlička (2014) a Čerba & Jedlička (2016).

### **AGROVOC**

URL: <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>

AGROVOC je slovník vyvíjený organizací FAO (Agricultural Information Management Standards) určený převážně pro účely zemědělství, lesnictví nebo životního prostředí. Vzhledem k tomu, že výše uvedené jsou velice úzce svázány s prostorovými daty a informací, představuje AGROVOC velice cenný zdroj sémantických informací pro otevřená (i jiná) data.

Silné stránky:

- Obrovské množství informací (32 000 konceptů)
- Multilingualita (23 jazyků), včetně češtiny
- Technologie (RDF, SKOS SPARQL)
- Dlouhodobý vývoj (od roku 1980)
- Intenzivní vývoj

Slabé stránky:

- Není zaměřeno na prostorová data a informace
- Občas nepřesné překlady do českého jazyka (například surveying - snímání)
- Poskytované nemusí být kompletní

Příležitosti:

- Dobrá komunikace s autory slovníku
- Linked data vazby na externí zdroje
- Mnoho odborných publikací zaměřených na AGROVOC

Hrozby:

- —

### **DBpedia**

URL: <http://wiki.dbpedia.org/>

Informace z projektu Wikipedia transformované do podoby datových záznamů. DBpedia tvoří hlavní komponentu sémantického webu.

Silné stránky:

- Obrovské množství informací (38,3 miliónů pojmů)
- Multilingualita (127 jazyků), včetně češtiny
- Technologie (RDF, SPARQL)
- GNU GPL licence

Slabé stránky:

- Není zaměřeno na prostorová data a informace
- Na vývoji se nepodílí žádná významná mezinárodní standardizační organizace (pouze univerzity a soukromá společnost)

Příležitosti:

- Možnost ovlivnit obsah
- Linked data vazby na externí zdroje

Hrozby:

- Informace nemusí být spolehlivé (editace dobrovolníky v rámci projektu Wikipedia)
- Konkurence projektu Wikidata

## **EuroVoc**

URL: <http://eurovoc.europa.eu/>

Mnohojazyčný tezaurus Evropské unie, který využívají pro překlady nejen orgány EU, ale parlamenty členských zemí a jejich regionů jako i ostatní správní orgány. Z tohoto důvodu je zařazení tohoto sémantického zdroje důležité především z hlediska legislativních aspektů otevřených dat.

Silné stránky:

- Záštitu EU (Úřad pro publikace EU)
- Multilingualita (23 jazyků EU + makedonština, srbština a albánština)
- Standard SKOS
- Otevřená licence

Slabé stránky:

- Poměrně malé množství termínů (ve srovnání s ostatními produkty)
- Velmi široký záběr (není zaměřený na prostorová data a informace)
- Na vývoji se nepodílí žádná významná mezinárodní standardizační organizace z oblasti prostorových dat

Příležitosti:

- Vazba na legislativu EU i národních států

- Linked data vazby na externí zdroje (GEMET, ECLAS, INSPIRE, AGROVOC, UNBIS)

Hrozby:

- Omezení na území EU z hlediska pokrytí a z velké části i z hlediska vazeb na externí zdroje.

### **GEMET (GEneral Multilingual Environmental Thesaurus)**

URL: <https://www.eionet.europa.eu/gemet/>

Tezaurus zaměřený především na oblast životního prostředí, který vzniká pod hlavičkou EEA (European Environmental Agency). Hesla jsou uvedena ve 36 jazycích, které kromě evropských jazyků pokrývají i čínštinu nebo americkou angličtinu.

Silné stránky:

- Záštitu EEA
- Multilingualita (36 jazyků)
- Tradice (od roku 2001)
- Často se vyskytující explicitní popisy
- Úzké zaměření tezauru
- Standard SKOS
- Licence CC BY

Slabé stránky:

- Úzké zaměření (není zaměřený čistě na prostorová data a informace)

Příležitosti:

- Vazba na legislativu INSPIRE (vyhledávání podle datových témat INSPIRE)
- Metadata INSPIRE vyžadují citace témat z tohoto slovníku
- Linked data vazby na externí zdroje (EuroVoc, AGROVOC)

Hrozby:

- Omezení na území EU
- Konkurence EuroVoc

### **NAL Thesaurus**

URL: <http://agclass.nal.usda.gov/>

Tezaurus a slovník Národní zemědělské knihovny (National Agricultural Library) v USA je velice úzce provázaný s tezaurem AGROVOC (především díky společnému tématu). Počet termínů zaměřených na prostorová data je poměrně úzký. To samé platí i pro vazby typu Linked data (na externí datové zdroje).

Silné stránky:

- Standard SKOS

Slabé stránky:

- Úzké zaměření na zemědělství a ochranu přírody a z toho vyplývající malý počet termínů týkajících se prostorových dat a informací.

Příležitosti:

- Vazba na prostředí v USA

Hrozby:

- Produkt obsahuje některá lokální specifika
- NAL nabízí velmi omezený počet vazeb na externí zdroje sémantických informací
- Pouze anglická a španělská verze

### **Ordnance Survey**

URL: <http://data.ordnancesurvey.co.uk/ontology>

Britská Ordnance Survey poskytuje sémantické informace ve formě ontologií, které jsou zaměřené na poštovní kódy, geometrie, prostorové vazby, geografické objekty (gazetteer) a administrativní jednotky.

Silné stránky:

- Standardy OWL, SPARQL
- Zaměření na doménu prostorových dat a informací
- Správu a rozvoj zajišťuje respektovaná organizace

Slabé stránky:

- Konkrétní instance se týkají Velké Británie

Příležitosti:

- Produkty Ordnance Survey v současnosti představují příklady dobré praxe v oblasti prostorových dat

Hrozby:

- Produkt obsahuje některá lokální specifika
- Pouze anglická verze

### **Terminologický slovník zeměměřictví a katastru nemovitostí**

URL: <http://www.vugtk.cz/slovník/>

Terminologický slovník zeměměřictví a katastru nemovitostí je zpracováván Terminologickou komisí Českého úřadu zeměměřického a katastrálního v jedenácti

zvolených oborech. Slovník obsahuje české termíny (popř. včetně synonym, oddělených od hlavního termínu čárkou), české výklady termínů a cizojazyčné ekvivalenty.

Silné stránky:

- Úzké zaměření na prostorová data a informace
- Multilingualita
- Explicitní sémantika (popisy termínů se zdroji)
- Na vývoji se podílí skupina erudovaných odborníků

Slabé stránky:

- Není strojově čitelný
- Není oficiálně zmíněna licence

Příležitosti:

- Vazba na mezinárodní normy

Hrozby:

- Bez vazeb na ostatní zdroje poskytující sémantické informace
- Zaměření pouze na resort ČÚZK

## **US National Map**

URL: <http://cegis.usgs.gov/ontology.html>

Center of Excellence for Geospatial Information Science (jako organizace v rámci USGS - U.S. Geological Survey) poskytuje ontologie a slovníky, kterou jsou součástí U.S. National Map. Tyto produkty jednak nabízí prvky U.S. National Map, včetně geografických jmen, a jednak slovníky obsahují základní typy prvků z oblastí jako jsou zastavěná území, povrchové vody nebo terénní prvky.

Silné stránky:

- Úzké zaměření na prostorová data a informace
- Provázaný systém slovníků, ontologií a gazetteerů
- Na vývoji se podílí skupina erudovaných odborníků

Slabé stránky:

- Omezení na oblast USA

Příležitosti:

- Rozsáhlé publikační aktivity

Hrozby:

- Zdroj může obsahovat lokální specifika
- Pouze v angličtině



## Wikidata

URL: <https://www.wikidata.org>

Formalizovaná verze databáze Wikipedia. Wikidata jsou rozsáhlou volně dostupnou webovou znalostní bází, která má sloužit jako podklad pro ostatní projekty nadace Wikimedia Foundation a umožnit jednoduchý přístup ke strojově čitelným datům.

Silné stránky:

- Obrovské množství informací (17,5 miliónů pojmů)
- Multilingualita
- Technologie (RDF, SPARQL)
- Creative Commons CC0 License
- Vývoj a správa jsou garantovány nadací s dlouhou působností v oblasti znalostních technologií

Slabé stránky:

- Specifický datový model
- Není zaměřeno na prostorová data a informace

Příležitosti:

- Možnost ovlivnit obsah
- Linked data vazby na externí zdroje

Hrozby:

- Informace nemusí být spolehlivé (editace dobrovolníky v rámci Wikipedia)
- Konkurence projektu DBpedia, který má delší tradici

## WordNet

URL: <https://wordnet.princeton.edu/>

WordNet je lexikální databáze pro anglický jazyk vyvíjená od roku 1985 týmem v laboratoři kognitivních věd na Princetonské univerzitě.

Silné stránky:

- Velké množství informací (17,5 miliónů pojmů)
- Tradice

Slabé stránky:

- Z hlediska struktury a zaměření databáze WordNet není vhodná pro sémantické rozšíření prostorových dat a informací
- Licence

Příležitosti:

-Reference v případech anglických výrazů

Hrozby:

- Pouze v angličtině
- Specifický formát a datový model
- Slabá provázanost ve smyslu linked dat

## Vymezení hlavních aktérů v oblasti publikace prostorových dat na národní a regionální úrovni

V průběhu průzkumu aktérů v oblasti publikace prostorových dat na národní a regionální úrovni byly zpracovány informace o 135 organizacích na území ČR. Při jejich výběru jsme se opírali o dlouholeté zkušenosti s tvůrci a poskytovateli prostorových dat u nás. Celkový počet 135 organizací lze rozdělit podle jejich typu takto:

- Vybrané ústřední orgány státní správy (5)
  - Ministerstva - práce a sociálních věcí, zemědělství, zdravotnictví, ČÚZK (včetně ZÚ), Český telekomunikační úřad
- Vybrané organizační složky státu (23)
  - Např.: Policie ČR, Armáda ČR (včetně Geografické služby Armády ČR), HZS, ÚHÚL, AOPK, krajské hygienické stanice, veterinární správa ...
- Vybrané příspěvkové organizace (9)
  - Např.: Národní památkový ústav, Česká geologická služba, ČHMÚ, správy národních parků, Ředitelství silnic a dálnic (ŘSD)
- Vybrané veřejné výzkumné instituce (3)
  - Výzkumný ústav meliorací a ochrany půd v.v.i. (VÚMOP), Centrum dopravního výzkumu v.v.i., Výzkumný ústav vodohospodářský v.v.i.
- Vyšší územní samosprávné celky - kraje (13)
- Vybrané základní územní samosprávné celky - obce (64)
  - Vybrány všechny obce s vyšším počtem obyvatel než 20000 - žádná z menších obcí není uvažována.
- Vybrané státní podniky (7)
  - Podniky správy jednotlivých povodí, LČR, Česká Pošta
- Vybrané velké obchodní společnosti (11)
  - Např: České dráhy, velcí správci inženýrských sítí (např. ČEZ, RWE), velcí dopravci (např. České dráhy, Dopravní podnik hl.m. Prahy) poskytovatelé telekomunikačních služeb (O2, T-Mobile, Vodafone)

Průzkum poskytování prostorových dat (včetně otevřených prostorových dat) u těchto aktérů byl prováděn výhradně z jejich aktuálních webů (květen 2016). Při tomto průzkumu byla pozornost zaměřena na tyto skutečnosti související s poskytováním prostorových dat:

1. Název poskytovatele, web poskytovatele
2. Uvádí poskytovatel na svém webu, že publikuje otevřená data?
  - adresa těchto stránek
  - jaká data jsou nabízena prostřednictvím otevřených dat
  - jaké formáty jsou nabízeny pro tato data
  - jaká licence
3. Uvádí poskytovatel na svém webu, že nabízí geodata, (geodata pro stažení)?
  - adresa těchto stránek

- jaká data jsou nabízena
  - jaké formáty jsou nabízeny pro tato data
  - jaká licence
4. Další doplňkové informace související
- jsou-li data těchto poskytovatelů publikována dále i prostřednictvím mapových serverů
  - jaké OWS služby (WMS, WFS služby) dále tyto poskytovatelé nabízejí

Z celkového počtu 135 vybraných organizací bylo zjištěno, že 11 těchto poskytovatelů nabízí svoje data jako otevřená data. Příklady, jakým způsobem probíhá poskytování těchto dat u těchto aktérů, je detailněji uvedeno v následující kapitole. Příklady organizací v České republice publikujících již nyní otevřená prostorová data. Zde jsou detailněji probrány i formáty takto uveřejňovaných dat, licence pro tato data a další aspekty.

Poskytovatelé nabízející jiná svoje prostorová data ke stažení (NE otevřená data) postupují různým způsobem:

- umožňují volné stažení dat
- data jsou dostupná prostřednictvím obchodním modulů na svých portálech (většinou jde o data zpoplatněná)
- data jsou dostupná prostřednictvím různých automatizovaných výdejních modulů pro stahování vyžadující většinou přihlášení uživatele
- data jsou nabízena pouze prostřednictvím osobního jednání a navazující smlouvy o využívání dat
- data jsou dostupná prostřednictvím webových služeb (WFS)

Formáty těchto poskytovatelů jsou velice různé, nejčastěji jde o proprietární formáty využívaných v oboru geoinformačních technologií (nejčastěji formát ESRI Shapefile, DGN formát, AutoCAD DXF), zcela výjimečně je volen GML formát. Velice často jsou prostorová data publikována pouze jako PDF soubory - zejména se jedná o územní plány zveřejňované prostřednictvím územních samosprávních celků. Řadu prostorových dat lze rovněž stahovat jen jako obrazové soubory (JPG, PNG ...).

Licence pro využití těchto dat jsou též velice různé - využití je nejčastěji vázáno kontakt na správce dat, příp. na smluvní ujednání mezi poskytovatelem dat a zájemcem o data. Velice často též jakákoliv informace o licenčním ujednání chybí nebo je obtížně zjistitelná.

Ve většině případů absentují rovněž jakákoliv metadata o datech, která lze od těchto poskytovatelů získat.

## **Příklady organizací v České republice publikujících již nyní otevřená prostorová data**

### **ČÚZK**

Web poskytovatele: <http://www.cuzk.cz/>

Portál otevřených dat: [http://geoportal.cuzk.cz/\(S\(m3js0244wv0olp3hsyliwwov\)\)/Default.aspx?mode=TextMeta&side=dSady\\_RUIAN&metadataID=CZ-CUZK-RUIAN\\_CR&menu=331](http://geoportal.cuzk.cz/(S(m3js0244wv0olp3hsyliwwov))/Default.aspx?mode=TextMeta&side=dSady_RUIAN&metadataID=CZ-CUZK-RUIAN_CR&menu=331)

Druh organizace: Ústřední orgán státní správy

Popis:

Popis může být citován pouze ze stránek poskytovatele (RÚIAN): „Registr územní identifikace, adres a nemovitostí (RÚIAN) je jedním ze čtyř základních registrů, které tvoří systém základních registrů veřejné správy ČR. RÚIAN obsahuje údaje o územních prvcích, účelových územních prvcích, územně evidenčních jednotkách, adresách na území ČR a jejich vzájemných vazbách. RÚIAN obsahuje referenční údaje a je jediným referenčním zdrojem informací o adresách (všechny ostatní informační systémy veřejné správy zobrazují data o adresách odkazem do RÚIAN). Editačními agendovými informačními systémy pro RÚIAN jsou Informační systém katastru nemovitostí (ISKN) a Informační systém územní identifikace (ISÚI). Data RÚIAN editují zákonem definovaní editoři (stavební úřady, obce, ČÚZK, ČSÚ). Jednou z forem poskytování dat RÚIAN je jejich předávání ve formě souborů obsahujících data RÚIAN nebo ISÚI ve výměnném formátu RÚIAN (VFR). Soubory VFR jsou ve formátu XML, konkrétně se jedná o formát GML 3.2.1. Série datových sad RÚIAN - ČR obsahuje prvky správního a územního členění od úrovně státu až po úroveň základních sídelních jednotek (ZSJ), a to ve formě základní sady, která obsahuje prvky stát až ZSJ včetně definičních bodů a popisů vlajek a znaků obcí a MOMC, nebo kompletní datové sady, která navíc obsahuje vektorová data ve formě generalizovaných nebo originálních hranic (polygonů). Historická data jsou dostupná v základní datové sadě bez grafických údajů, tj. definičních bodů a polygonů prvků (od 1.7. 2012).“

Specifikace prostorových dat: Data jsou publikována ve formátu GML prostřednictvím Veřejného dálkového přístupu ČÚZK.

Licence: Data jsou volně dostupná za jakýmkoliv účelem

SWOT analýza portálu:

- Silné stránky: Význam otevření těchto podrobných základních dat je obrovský. Komplexnost dat. Existující podrobná standardizovaná metadata.
- Slabé stránky: Nejsou v tomto okamžiku zásadní. Aktualizační dávky RÚIAN dostupné pouze v jednom souboru, vztaženou na celou republiku.
- Příležitosti: Data jsou každodenně běžně používána pro širokou škálu tvůrců aplikací a i pro širší veřejnost.
- Hrozba: Hrozba zneužití zatím není specifikována.

**Institut plánování a rozvoje hl. m. Prahy**

Web poskytovatele: <https://www.praha.eu/>

Portál otevřených dat:

- <http://www.geoportalpraha.cz/cs/.opendata#.VzLHQ4SLT>
- <http://www.geoportalpraha.cz/cs/mapy-mc-ke-stazeni#.VzLg6oSLT>
- <http://www.iprpraha.cz/clanek/1346/graficka-cast-ke-stazeni>
- <http://www.iprpraha.cz/vykresyup>

Druh organizace: Příspěvková organizace, jejímž zřizovatelem je hl. město Praha

Popis:

Ke stažení je rozsáhlé množství prostorových datových sad rozdělených do následujících skupin: 3D model, Digitální technická mapa Prahy, Doprava, Geologie, Hluk, Kvalita životního prostředí, Mapové podklady, Občanská vybavenost, Ochrana přírody a krajiny, Ortofoto, Ovzduší, Platný územní plán, Socioekonomická data, Technická infrastruktura - vodní hospodářství, Urbanismus a krajina .

Specifikace prostorových dat:

Podle charakteru dat – rastrová anebo vektorová – data jsou dostupná ke stažení v různých formátech (GML, TIFF, JSON, Esri Shapefile - SHP atd., DXF Autocad, standard Atom). Mezi dostupnými formáty jsou jak formáty s tříhvězdičkovou klasifikací, tak další běžné 2-hvězdičkové formáty využívané v geoinformatice. Z pohledu široké škály využitelných formátů nevznikají tedy ze strany soudobých uživatelů těchto otevřených prostorových dat IPR žádné technologické bariéry. Licence: Data jsou ke stažení a užívání dle licenci: Uveďte původ - zachovejte licenci 4.0 - mezinárodní

SWOT analýza portálu:

- Silné stránky: Velmi velký výběr rozmanitých geoprostorových dat, velký výběr formátů, podrobná standardizovaná metadata.
- Slabé stránky: Licence uveďte původ – zachovejte licenci asi není nejlepší volba pro data, publikována městem. Licence – (jen) uveďte původ je vhodnější pro publikaci městských dat.
- Příležitosti: Příležitost pro vývojáře tvorby velké řady aplikací nad otevřenými daty hlavního města Prahy.
- Hrozba: Hrozba zneužití není identifikovaná.

## **Plzeňský kraj**

Web poskytovatele: <http://www.plzensky-kraj.cz/>

Portál otevřených dat: <http://geoportal.plzensky-kraj.cz/gs/volne-dostupne-datove-sady/>

Typ organizace: Vyšší územní samosprávné celky – kraje

Popis:

Součástí geoportálu Plzeňského kraje je k dispozici méně obsáhlá stránka věnovaná geoprostorovým datům ke stažení. Ke stažení jsou nabízeny 4 prostorové datové sady: Záplavová území, USES ZUR, přírodní parky, ArcČR 500. První tři datové sady jsou publikované přímo Plzeňským krajem. Poslední datová sada je vytvořena obchodní společností Arcdata .

Specifikace otevřených prostorových dat: Otevřená data jsou poskytována ve formátu ESRI Shapefile (SHP) a metadata jsou publikována prostřednictvím systému METIS. Data hodnotíme jako 2-hvězdičková.

Licence: Na popisovaném portálu je uvedeno, že všechna data jsou uveřejněna v licenci CC BY - uveďte původ

SWOT analýza portálu:

- Silné stránky: Nabízená data mají vyplněná metadata a jsou poskytované pod otevřenou licenci ve formátu, který je snadno zpracovatelný.
- Slabé stránky: Menší počet nabízených datových sad.
- Příležitosti: Otevření většího množství datových sad.
- Hrozba: Žádné hrozby nejsou identifikované. V případě nabízených dat se nejedná o osobně citlivá údaje.

## **Město Ostrava**

Web poskytovatele: <https://www.ostrava.cz/cs>

Portál otevřených dat: <http://mapy.ostrava.cz/opendata/opendata.php>

Druh organizace: Základní územní samosprávné celky - obce

Popis:

Na geoportálu statutárního města Ostrava jsou nabízené ke stažení kolem 30 různě zaměřených prostorových datových sad. Například: 3D modely budov, adresní body, uliční síť, zastávky MHD, památné stromy a další sady. Řada dat jsou úplná, zčásti neúplná - vzorky dat (například 3D modely budov). U většiny datových sad se však jedná o úplná data. Všechna nabízená data jsou prostorová.

Specifikace otevřených prostorových dat:

Data jsou poskytována v populárních vektorových formátech (GeoJSON, GML, a též proprietární formáty AutoCAD DXF a Esri Shapefile - SHP). Jedna se o 3-hvězdičková data. Data jsou dostupná v souřadnicových systémech WGS-84 a rovněž S-JTSK.

Licence: Data jsou ke stažení a k užívání podle licence: Uveďte původ - zachovejte licenci 4.0 - Mezinárodní (<http://creativecommons.org/licenses/by-sa/4.0/deed.cs>)

SWOT analýza portálu:

- Silné stránky: Velký výběr snadno zpracovatelných geoprostorových dat, výběr formátů, metadata.
- Slabé stránky: Není explicitně uvedeno, kde se jedna o úplnou datovou sadu a kde pouze o neúplný vzorek dat. Licence uveďte původ – zachovejte licenci asi není nejlepší volba pro data, publikovaná městem. Licence – (jen) uveďte původ je vhodnější pro publikaci zmiňovaných dat.
- Příležitosti: Příležitostí může být propagování široké škály otevřených dat pro tvorbu moderních mobilních aplikací.
- Hrozba: Žádné hrozby nejsou identifikované. Vesměs se nejedná o osobně citlivé údaje.

## Kraj Vysočina

Web poskytovatele: <http://www.kr-vysocina.cz/>

Portál otevřených dat: <http://opendata.kr-vysocina.cz/index.php?akce=sady>

Druh organizace: Vyšší územní samosprávné celky – kraje

Popis: Prostřednictvím Geoportálu Kraje Vysočina jsou nabízeny ke stažení následující otevřené datové sady: Správní hranice, Metropolitní síť, Data životního prostředí, Památky, Zastávky autobusové dopravy.

Specifikace prostorových dat: Data jsou poskytována ve formátu KML a v tomto ohledu odpovídají 3-hvězdičkovým datům.

Licence: Není specifikována žádná přesná licence. Každý případ se řeší zvlášť, a k použití a šíření dat (výstupů) je třeba kontaktovat příslušného správce dat.

SWOT analýza portálu:

- Silné stránky: Nabízená prostorová data jsou ve formátech, které jsou snadno zpracovatelné.
- Slabé stránky: Není k dispozici příliš velký výběr geoprostorových dat (jen 5 datových sad), nejasná licence, nejsou k dispozici standardizovaná metadata.
- Příležitosti: Příležitostí je otevření většího množství geoprostorových dat.
- Hrozba: S výjimkou dat metropolitní sítě (kde je otevření dat snad i diskutabilní) se nejedná o důvěrná data nebo osobně citlivá data.

## Město Děčín

Web poskytovatele: <http://www.mmdecin.cz/>

Portál otevřených dat: [http://www.mmdecin.cz/dokumenty/cat\\_view/238-otevrena-data](http://www.mmdecin.cz/dokumenty/cat_view/238-otevrena-data)

Druh organizace: Základní územní samosprávné celky - obce

Popis:



Po podpisu memoranda o spolupráci při prosazování veřejných dat s Fondem Otakara Motejla pokračuje město Děčín v aktivní publikaci otevřených dat. Data jsou publikována na stránkách: [http://www.mmdecin.cz/dokumenty/cat\\_view/238-otevrena-data](http://www.mmdecin.cz/dokumenty/cat_view/238-otevrena-data). Jedna ze sekcí - Grafická a mapová data (GIS) se orientuje vysloveně na prostorová data. Ke stažení jsou následující datové sady: Části obce, Majetek města Děčín, Veřejné osvětlení, Dopravní značky, Komunikace, Volební okrsky, Povodně, Školské obvody, Sběrné nádoby a sběrné dvory.

Specifika prostorových dat: Data jsou poskytována ve formátu ESRI Shapefile - SHP, což odpovídá 2-hvězdičkovým datům.

Licence: Všechny slovně popsání podmínky odpovídají licenci CC BY - Uveďte původ („Data zveřejněná na webových stránkách statutárního města Děčín a označená jako otevřená data smí uživatel:šířit – kopírovat, distribuovat a sdělovat veřejnosti, využít a citovat, využívat komerčně a to při dodržení podmínky, že jako zdroj dat musí být uvedeno statutární město Děčín.“).

SWOT analýza portálu:

- Silné stránky: Středně velký výběr strojově čitelných otevřených prostorových dat.
- Slabé stránky: Zdá se, že od roku podepsání memorandu o spolupráci s Fondem Otakara Motejla počet publikovaných geoprostorových dat postupně klesá.
- Příležitosti: Příležitostí může být propagování nabízených otevřených dat pro tvorbu moderních mobilních aplikací.
- Hrozba: Žádné hrozby nejsou identifikované. Nejedná se o osobně citlivá údaje.

## Město Bohumín

Web poskytovatele: <https://www.mesto-bohumin.cz/>

Portál otevřených dat: <https://www.mesto-bohumin.cz/cz/radnice/otevrena-data/>

Druh organizace: Základní územní samosprávné celky - obce

Popis: Ke stažení je nabízen velký výběr otevřených dat sad, často se ale nejedná o typická prostorová data (Příklady dat: seznam bytových techniků, událostí ve městě, licitace bytů, plánované odstávky, seznam ulic města atd., technická mapa).

Specifikace prostorových dat: Data jsou většinou poskytována v proprietárním XML formátu. Odpovídají 2-3-hvězdičkovým datům. U technické mapy není licence uvedena.

Licence: Licence odpovídá licenci typu - CC BY - Uveďte původ (uživatelé dat musí uvést zdroj). U technické mapy není licence stanovena.

SWOT analýza portálu:

- Silné stránky: Větší výběr dat informačního charakteru o městě.
- Slabé stránky: Data jsou většinou proprietárním (nestandardizovaným) XML formátu (například [https://www.mesto-bohumin.cz/export/databaze\\_firem.xml.php](https://www.mesto-bohumin.cz/export/databaze_firem.xml.php)), chybějící metadata.
- Příležitosti: Příležitosti pro vývojáře vytváření mobilních aplikací informačního charakteru o městě (kde najít nejbližší restauraci, nejbližší služby, sportoviště atd.)
- Hrozba: Žádné hrozby nejsou identifikované. Nejedná se vesměs o osobně citlivé údaje.

### **Povodí Labe**

Web poskytovatele: <http://www.pla.cz/>

Portál otevřených dat: - [http://www.pla.cz/planet/webportal/internet/cs/obsah/download\\_520.html](http://www.pla.cz/planet/webportal/internet/cs/obsah/download_520.html) - <http://www.pla.cz/planet/projects/uap/>

Druh organizace: Státní podnik

Popis: Ke stažení jsou různorodá data v rámci vymezené působnosti tohoto správce povodí (například záplavové zóny, mapy ohrožení, osy toku v povodí Labe a další zajímavá data).

Specifikace prostorových dat: Data jsou poskytována ve formátu ESRI Shapefile SHP. Odpovídají 2-hvězdičkovým datům.

Licence: V podstatě odpovídá licenci typu CC BY - Uvedte původ, uživatele dat musí uvést zdroj. Konkrétně v podmínkách využití je uvedeno: “V případě tisku si zpracovatel vyhrazuje uvedení copyrightu na vytištěném listu takto: (c) Zpracováno s použitím dat Povodí Labe”.

SWOT analýza portálu:

- Silné stránky: Velký výběr prostorových datových sad o povodí Labe.
- Slabé stránky: Jen formát SHP, nestandardizovaná metadata nabízených dat.
- Příležitosti: Publikování i některých živých dat z on-line senzorů organizace.
- Hrozba: Hrozba zneužití dat není identifikována.

### **Český telekomunikační úřad**

Web poskytovatele: <http://www.ctu.cz/>

Portál otevřených dat: <http://data.ctu.cz/dataset>

Druh organizace: Ústřední orgán státní správy

Popis: Ke stažení jsou datové sady (přibližně 100 datových sad) úzce souvisejících s provozem telekomunikačními sítěmi. Mezi datovými sadami jsou některé sady které

jsou vztaženy k administrativním celkům , například, k ZSJ (Pokrytí oblastí ČR infrastrukturou NGA – ZSJ), a také datové sady které jsou lokalizovány prostřednictvím souřadnic poloh zařízení (Stanice bezdrátového místního informačního systému (BMIS)). Nicméně většina dat nemá bezprostřední prostorový charakter.

Specifikace prostorových dat: Data jsou ke stažení většinou ve formátu CSV. Některé datové sady jsou poskytovány přes API rozhraní, které vrací data ve formátu JSON (viz například <http://data.ctu.cz/node/38/api>).

Licence: V podstatě odpovídá licenci typu CC BY - Uvedte původ , uživatele dat musí uvést zdroj . Hlavní podmínkou uvedení zdroje, jinak data mohou být použité k jakémukoliv účelu (i ke komerčnímu). Podrobné podmínky jsou publikovány zde: <http://data.ctu.cz/podminky-uziti-dat>.

SWOT analýza portálu:

- Silné stránky: Velmi velký výběr specializovaných dat daného oboru (tele+rádio komunikace). Existence API rozhraní pro publikování některých datových sad. Jedna datová sada je publikována v RDF formátu a to Data o rozpočtu ČTÚ (<http://data.ctu.cz/dataset/rozpocet>) .
- Slabé stránky: Málo dat s prostorovou složkou.
- Příležitosti: Příležitost pro vytváření zajímavých tematických aplikací o dostupnosti radiotelekomunikačních služeb.
- Hrozba: Hrozba zneužití dat není identifikována.

### **Agentura ochrany přírody a krajiny (AOPK)**

Web poskytovatele: <http://www.ochranaprirody.cz/>

Portál otevřených dat: - [http://portal.nature.cz/publik\\_syst/ctihtmlpage.php?what=6181&nabidka=rozbalitModul&modulID=467](http://portal.nature.cz/publik_syst/ctihtmlpage.php?what=6181&nabidka=rozbalitModul&modulID=467) - [http://portal.nature.cz/publik\\_syst/ctihtmlpage.php?from=mapa&what=6142](http://portal.nature.cz/publik_syst/ctihtmlpage.php?from=mapa&what=6142)

Druh organizace: Organizační složka státu

Popis:

Prostřednictvím portálu AOPK je publikována řada prostorových otevřených dat formou WFS služeb. Příkladem jsou Chráněná území, Mezinárodně významné části přírody a územní působnost úmluv, Průchodnost krajiny pro velké savce, Soustava NATURA 2000, Památné stromy, Údaje o území poskytované AOPK ČR pro tvorbu územně analytických podkladů a další. Větší datových sad je publikován formou WMS služeb. Prostorová data (WFS, WMS) jsou společně publikována prostřednictvím ArcGIS REST služeb.

Specifikace prostorových dat: Data, která jsou publikována WFS službou jsou standardně dostupná ve formátu GML (3-hvězdičková data), WMS služby jsou 1-hvězdičková data.

Licence: Licence odpovídá licenci typu - CC BY - Uvedte původ, uživatele dat musí uvést zdroj.

SWOT analýza portálu:

- Silné stránky: Velké množství dat související s činností organizace
- Slabé stránky: Menší množství vektorových dat. Převažují většinou WMS služby a obrazové formáty.
- Příležitosti: Užitečná data pro tvůrce aplikací pro tvůrce aplikací orientujících se na přírodu České republiky.
- Hrozba: Hrozba zneužití není identifikovaná.

### **Česká správa sociálního zabezpečení**

Web poskytovatele: <http://www.cssz.cz/cz>

Portál otevřených dat: <https://data.cssz.cz/>

Druh organizace: Organizační složka státu

Popis:

Pro stažení je k dispozici celá řada otevřených dat organizace (Výše důchodu, Nejčastější příčiny vzniku invalidity atd.). Většina těchto otevřených dat nemá přímou prostorovou složku. Řada dat je prostorově lokalizovatelná prostřednictvím identifikátorů okresů nebo krajů obsažených v datech.

Specifikace prostorových dat: Data jsou publikována ve formátech CSV a RDF. Můžeme je tedy považovat za 4-5 hvězdičková data. Jsou propojené s jinými daty například: <http://ruian.linked.opendata.cz>.

Licence: Licence v podstatě odpovídá licenci typu - CC BY - Uvedte původ, uživatele dat musí uvést zdroj. Podrobné licenční podmínky jsou dostupné zde: <https://data.cssz.cz/web/otevrena-data/podminky-uziti>.

SWOT analýza portálu:

- Silné stránky: Velké množství statistických dat o problematice sociálního zabezpečení v České republice (i když se nejedná o typická prostorová data), která jsou dokonce dostupná jako 4 až 5-hvězdičková data.
- Slabé stránky: Málo dat s prostorovou složkou. Data pouze do úrovně okresu mohou připadat v některých případech jako dost hrubá.
- Příležitosti: Data mohou být zajímavá pro tvůrce aplikací orientovaných na různé statistické ukazatele.
- Hrozba: Hrozba zneužití není identifikovaná.

## **Analýza současných systémů a praktik publikace prostorových dat veřejné správy**

V současné době existuje řada subjektů veřejné správy, jež v současné době poskytují otevřená data včetně prostorových. Konkrétní příklady jsou popsány v části Vymezení hlavních aktérů v oblasti publikace prostorových dat na národní a regionální úrovni.

Z hlediska prostorových dat se lze setkat především s těmito datovými formáty:

- GML
- Shapefile
- KML
- DXF
- GeoJSON
- CSV
- V některých případech je poskytována prohlížecká služba WMS. V některých případech (např. Česká správa sociálního zabezpečení) jsou k dispozici jednoduché vizualizace dat pomocí mapy.

Data jsou ve většině případů k dispozici v těchto souřadnicových referenčních systémech:

- S-JTSK / Křovákovo zobrazení (EPSG 5514)
- WGS84 (EPSG 4326)

Licence:

- CC - BY s požadavkem na uvedení odkazu na původce dat

Pro úplnost uvedme další možnosti poskytování dat tak, jak to specifikují relevantní standardy případně zaběhlé postupy v oblasti GIS.

Souřadnicové systémy specifikované INSPIRE nebo používané v ČR:

- S-JTSK / Krovak East North - EPSG:5514
- S-JTSK (Ferro) / Krovak East North - EPSG:5221
- ETRS89 - EPSG:4258
- ETRS89 / LCC Europe - EPSG :3034
- ETRS89 / LAEA Europe - EPSG:3035
- ETRS89 / UTM zone 33N (N-E) - EPSG:3045
- ETRS89 / UTM zone 34N (N-E) - EPSG:3046
- WGS 84 - EPSG:4326
- WGS 84 / UTM zone 33N - EPSG:32633
- WGS 84 / UTM zone 34N - EPSG:32634
- WGS 84 / Pseudo-Mercator - EPSG: 3857 (někdy označováno jako Google Mercator s kódem 900913)
- Pulkovo 1942(83) / Gauss-Kruger zone 3- EPSG:3835
- Pulkovo 1942(83) / Gauss-Kruger zone 4 - EPSG:3836

## **Ochrana osobních údajů dat určená pro zveřejnění**

V některých oblastech nelze k otevřenému zveřejnění dat přistoupit, resp. je nutné při jejich zveřejnění dbát příslušné legislativy, a to především v případě dat chráněných zákonem o ochraně osobních údajů. V ČR je problematika řešena zákonem č. 101/2000 Sb., o ochraně osobních údajů, který definuje a dále upravuje a specifikuje možnosti zpracování a zveřejnění osobních údajů.

Zákon definuje osobní údaj takto: “Osobním údajem je jakákoliv informace týkající se určené nebo určitelné fyzické osoby, k níž se osobní údaje vztahují. Tato se považuje za určenou nebo určitelnou, jestliže lze fyzickou osobu přímo či nepřímo identifikovat zejména na základě čísla, kódu nebo jednoho či více prvků, specifických pro její fyzickou, fyziologickou, psychickou, ekonomickou, kulturní nebo sociální identitu”.

Dále je definován citlivý údaj jako: “Citlivým údajem osobní údaj vypovídající o národnostním, rasovém nebo etnickém původu, politických postojích, členství v odborových organizacích, náboženství a filozofickém přesvědčení, odsouzení za trestný čin, zdravotním stavu a sexuálním životě subjektu údajů a genetický údaj subjektu údajů; citlivým údajem je také biometrický údaj, který umožňuje přímou identifikaci nebo autentizaci subjektu údajů,”

## **Analýza a definice uživatelských požadavků**

Ve vybraných organizacích, které mají praktické zkušenosti s publikováním otevřených dat, proběhly osobní konzultace. Organizace jsme vybrali podle zkušeností s publikováním otevřených prostorových dat. Šlo tyto subjekty:

- Institut plánování a rozvoje hl. m. Prahy (Mgr. Jiří Čtyroký)
- Český úřad zeměměřický a katastrální ( Ing. Petr Souček, Ph.D)
- Český statistický úřad (Ing. Eduard Durník)
- Agentura ochrany přírody a krajiny (Mgr. Jan Zárybnický)

Výsledky pohovorů byly následně rozděleny podle jednotlivých témat.

### **Ekonomické a legislativní aspekty**

- Data jsou publikována zejména na základě INSPIRE, někdy české legislativy
- Řada dat publikována z vlastní iniciativy
- Důsledkem může být snížení administrativy (např. kolem zák. 106 - IPR apod.), což může vést ke snížení nákladů
- S otevřením dat se snížil příjem ČÚZK z prodeje katastrálních dat cca o 50 mil. Kč/rok.



## Licence

- Většina subjektů se snaží použít licenci Creative Commons ver. 3.0 nebo 4.0 (CC)
- Limitujícím faktorem je neexistence českého překladu CC 4.0
- Ne všechna data lze publikovat pod licencí CC, neboť existují specifické licenční požadavky (např. u dat publikovaných veřejně na základě zákona není stanoven požadavek uvádět původce apod.)
- Data spravovaná zejména soukromými subjekty nelze publikovat jako otevřená nebo data se smíšeným obsahem, Je nutné velice pečlivě dbát na tato pravidla.
- Je třeba velice pečlivě dbát na to, aby nebyla publikována data podléhající utajení nebo dalším speciálním požadavkům (agregace osobních dat ČSÚ apod.)

## Formáty

- Data jsou proto publikována ve formátech daných datovými specifikacemi INSPIRE (GML - ČÚZK),
- Způsob zveřejňování jako stahovací či prohlížečské služby INSPIRE (WFS, WMS - AOPK, Atom, WFS, WMS - ČÚZK)
- Kromě toho jsou data zveřejňována v některých proprietárních formátech z historických důvodů pro praktičnost užití koncovým uživatelem (VFK, SHP, XLS)
- Objevují se další otevřené formáty: GeoJSON, DXF, Geopackage, CSV, XML (IPR, ČÚZK)
- Z uvedeného vyplývá, že data jsou většinou na úrovni 3 hvězdiček. Není obecný požadavek publikovat data v RDF, i když některé pokusy probíhají (ČÚZK)
- Limitem je nedostatek uživatelského software, který by umožnil efektivně data RDF zpracovat na straně koncového uživatele
- Model RDF není příliš vhodný pro prostorová data s polygonovou nebo liniovou geometrií. V aplikacích bude nutno uvažovat o jakémsi smíšeném modelu sémantického webu s klasickým analytickým GIS přístupem (to se již odráží ve specifikaci např. GeoSPAQL)

## Metadata a katalogy

Aktuální situace:

- V současnosti jsou pro prostorová používány metadatové standardy vyžadované směrnicí INSPIRE (ISO 19115/19119/19139)
- Metadata jsou uložena v lokálních metadatových katalozích jednotlivých organizací a harvestována na Národní katalog INSPIRE (<http://geoportal.gov.cz>). odsud pak na Evropský geoportál INSPIRE. Katalogy jsou založeny na katalogové službě OGC CSW 2.0.2 - ISO-AP 1.0 v souladu se směrnicí INSPIRE
- Paralelně probíhá harvesting na NKOD (ČÚZK) pomocí protokolu CKAN API
- Národní katalog INSPIRE umožňuje export metadat ve formátu DCAT-AP / GeoDCAT-AP RDF-XML, takže je umožněn import do katalogů podporujících tento formát
- NKOD bude mít implementován import DCAT-AP v příštím roce (MV)

Potřeby:

- Pro správnou interpretaci dat třeba poskytovat pečlivou dokumentaci aplikačního schematu (struktury dat)
- V DCAT-AP jsou problémy s interpretací serií a vazeb mezi jednotlivými datovými sadami

## **Thesaury, ontologie**

Aktuální situace:

- Často je používáno třídění podle GEMET. Ovšem tento slovník je vytvořen pro životní prostředí a pro další obory je nedostatečný. Navíc obsahuje řadu chyb a nekonzistencí v třídění.
- IPR - používá vlastní tematické slovník (cca 20 termínů)
- ČÚZK - zadáno vytvoření thesauru pro Zeměměřický slovník

Potřeby:

- Využít či definovat nějaký obecný slovník pro veřejnou správu ČR
- Definovat ontologie pro téma landuse, územní plánování (IPR)

## Cesta k Linked Open Data

Pro snadnější implementaci principů propojených otevřených dat byly detekovány tyto potřeby a překážky:

Definovat sémantické vazby mezi těmito entitami

- Je třeba definovat systém URI pro základní registry (RÚIAN, Osoby...)
- Vytvořit webové služby z těchto registrů podle principů LOD
- Překážkou vytvoření systému URI pro právnické osoby je neexistence IČO u některých subjektů.

Další detekované problémy

- Interpretace složených geometrií - způsob dekompozice složených objektů, definice vazeb, logiky
- Časové aspekty - kdy vzniká nový objekt, kdy jde o modifikaci stávajícího, jak kódovat různé verze objektů v čase, jak se na ně dotazovat, jak kódovat vazby mezi nimi
- Která data publikovat na NKOD - v budoucnosti budou publikovány vybrané datové sady ČÚZK na NKOD. Je to správné? Nebo publikovat jen agregované datové sady?

## Seznam literatury

Čerba, O., & Jedlička, K. (2014). Geomatic Concepts in Agriculture Thesauri. Papers in Economics and Informatics.

Čerba, O., & Jedlička, K. (2016, v tisku). Linked Forest. Open GeoSciences.

Mylopoulos, J. (2004). Data Semantics Revisited: Databases and the Semantic Web. In Proceedings of the DASFAA (Vol. 4, pp. 17-19).

Sheth, A. (1997). Panel: Data semantics: what, where and how?. In Database Applications Semantics (pp. 601-610). Springer US.