

ROZŠIŘOVÁNÍ A VEDENÍ ČÍSELNÍKŮ NAD RÁMEC SMĚRNICE INSPIRE

Michal MED¹

¹ Sekce centrální databáze, Český úřad zeměměřický a katastrální, Pod Sídlištěm 1800/9, 182 00, Praha 8,
Česká republika
michal.med@cuzk.cz

Abstrakt

Od roku 2015 vznikly v rámci implementace INSPIRE na ČÚZK dvě datové sady rozšiřující modely datových sad definované v Datových specifikacích a ve schématech XSD. S tím je spojena tvorba nejen datových modelů, ale i nových číselníkových hodnot, případně celých číselníků. Na adrese <http://inspire.ec.europa.eu/codelist/> existuje aplikace spravovaná JRC obsahující číselníky a jejich hodnoty. Některé z těchto číselníků jsou rozšiřitelné, jiné nikoliv. Nerozšiřitelné číselníky se týkají hodnot atributů striktně vyžadovaných v datových modelech INSPIRE. Rozšiřitelné číselníky jsou obecně ty, které obsahují přidanou hodnotu k datům, pomáhají je lépe popsat nebo porovnat na základě vlastností, které se mezi jednotlivými objekty často opakují. Pro rozšíření datové sady Parcely (CPX) byla vytvořena řada číselníků popisující nové typy prostorových prvků a popisující rozšířené vlastnosti stávajících. Mezi nimi jmenujme například typy hranice parcel, značky druhu pozemku nebo typy geodetických bodů. Číselníky a jejich hodnoty jsou vedeny v databázi. Pro základní zobrazování číselníků a jejich hodnot vznikla na ČÚZK webová aplikace. Po vzoru aplikace na webu JRC zobrazuje obsah databáze v jazyce HTML, ale umožňuje i export dat ve strojově čitelných formátech JSON a XML. Během posledních let jsem se setkal s řadou dotazů, jak číselníky rozšiřovat. V článku se na to pokusím částečně odpovědět. Rozšiřování číselníků je poměrně jednoduché, je potřeba pouze dodržet strukturu a zajistit místo, kde hodnoty publikovat. To je obtížnější část úkolu. Máme možnost centralizace číselníků, ať už na resortní, národní nebo mezinárodní úrovni, nebo lze číselníky spravovat samostatně -- decentralizovaně. Obě možnosti mají výhody i nevýhody.

Abstract

Two datasets extending existing data models defined in Data Specification documents and XML schemas were implemented since 2015. That involved extension of existing and creation of new codelist values and also brand new codelists. At the <http://inspire.ec.europa.eu/codelist/> there is we application managed by JRC containing codelists and their values. Some of those are extensible, some are not. Non-extensible codelists contain strictly required codelist values. Extensible codelists are meant to identify or describe attributes or properties or enhance the value of the data. While extending Parcels (CPX), a set of new codelists was created. New codelists describe attributes of new feature types, or extended attributes of current feature types, e.g. boundary type, land type symbol, geodetic point field type etc. Codelists and their values are stored in database. For the basic veiw of codelists, web application was created. According to the JRC web app, basic view are in HTML, but data may be exported to machine readable formats XML or JSON. I have encountered many questions on extending codelists during last few years. The article answers the question partially -- extending itself is rather easy, it only requires to keep the structure and then publish it somewhere. Access point for the service is the real problem here, considering either centralized or local solution. Both options have pros and cons.

Klíčová slova: INSPIRE, Parcely, Budovy, číselník, rozšíření

Keywords: INSPIRE, Parcels, Buidlings , codelist, extension

KAPITOLA 1 - ÚVOD

Od začátku roku 2016 probíhá rozšiřování datové sady Parcely z důvodu zvýšení kvality jejích dat a atraktivity pro uživatele doplněním nových typů prostorových prvků a přidáním nových atributů prvků stávajících. Důvodem pro rozšiřování byla absence prostorových prvků (features), nezbytných pro národní

uživatelé katastru nemovitostí. Mezi typy těchto prvků patří například věcná břemena, geodetická bodová pole nebo další prvky mapy. Kromě toho rozšíření obsahuje i parcely z analogové mapy ve formě definičních bodů. Stávající typy prostorových prvků jsou také rozšířeny o nové atributy, například katastrální hranice je rozšířena o typ katastrální hranice, nebo katastrální parcela o vazbu na budovy, mapovou značku druhu pozemku, kódy hilucs atd. Řada těchto atributů bude obsahovat dokola stejné hodnoty, nebo vyžadovat seznam, přípustných hodnot. Pro tento účel se jeví jako vhodné využití číselníků.

KAPITOLA 2 – POZADÍ IMPLEMENTACE

Podkapitola 2.1 – Číselníky v INSPIRE

Problematika číselníků je do INSPIRE zabudována již od jeho počátku. Kromě toho, že datové modely pro téměř všechna témata počítají s jejich využitím, počítalo JRC od počátku s tím, že nedokáže dát dohromady všechny číselníkové hodnoty, které budou chtít mapovací agentury členských států využít. Proto jsou některé číselníky vedeny jako tzv. rozšiřitelné. Na adrese <http://inspire.ec.europa.eu/codelist> je k dispozici přehled všech číselníků, informací o nich a o jejich hodnotách. K číselníkům i k jejich hodnotám lze přistupovat přímo za pomoci tzv. trvalých identifikátorů (persistent identifiers) – principu používaného v linked data. Identifikátor ve formě IRI (např. <http://inspire.ec.europa.eu/codelist/CadastralZoningLevelValue> pro Úrovně územního členění katastrálních území) jednoznačně identifikuje číselník a umožňuje přístup k jeho reprezentacím v různých jazycích, ať už ve smyslu jazyků užívaných lidmi (čeština, angličtina), nebo jazyků užívaných počítačem (XML, JSON, RDF/XML, CSV, ATOM).

Podkapitola 2.2 – Centralizace číselníků

Seznam číselníků užívaných v INSPIRE je centralizovaný na výše uvedené adrese <http://inspire.ec.europa.eu/codelist>. K tomu, aby byl číselník vytvořen na stejném místě, by bylo potřeba, aby ho tam umístil někdo z JRC, kdo má k umístění přístup. Tím by se ovšem z národního rozšíření stala součást oficiálního INSPIRE číselníku. Číselníky je proto potřeba vytvářet na národní úrovni. Otázkou zůstává, jestli je vhodnější certifikovat národní autoritu a centralizovat registr číselníků u ní, nebo zda je vhodnější vytvářet číselníky samostatně. Oba přístupy mají výhody i nevýhody.

Výhodou je snadné vyhledání všech národních číselníků na jednom místě. Nevýhodou je nutnost spoléhat na jednu autoritu, která může snadno selhat. Na jednoduchém příkladu lze ukázat výhody i nevýhody obou přístupů. Dvě organizace spravující podobná data se rozhodnou pro jejich národní rozšíření a v souvislosti s tím vytváří nový číselník. Jedna organizace tedy vytvoří číselník a umístí ho do centralizovaného národního registru. V okamžiku, kdy se druhá strana rozhodne vytvořit vlastní číselník, snadno zjistí, že pro danou problematiku již číselník existuje. Pokud jejím účelům bude vyhovovat, je vše v pořádku. Předpokládejme však, že vyhovovat nebude. Pak musí druhá organizace vytvořit nový číselník a umístit ho do registru vedle stávajícího číselníku, což může působit zmatečně, nebo ho autorita nemusí vůbec přijmout, nebo se musí domluvit s první organizací a číselník náležitě upravit. Pak ale může dojít k inkonzistenci v již používaných datech. Pokud by číselník první organizace vystavila ve vlastním registru, druhá organizace by se o něm nemusela dozvědět a tím pádem by možná dělala práci, kterou již někdo udělal. Na druhou stranu je zde zřejmé, že číselník byl vytvořen danou organizací za daným účelem a slouží tedy především k tomu. Navíc nic nebrání tomu, aby na něj odkazovaly i další organizace. Nevýhodou pak je to, že tvůrce a správce číselníku není autoritou a může svoji práci kdykoliv změnit. Tím by zásadně ovlivnil všechny, kteří číselník používají. Střední cestou může být uchování číselníků v místních registrech s centralizovanou evidencí.

KAPITOLA 3 – TVORBA REGISTRU

Systém registru číselníků na ČÚZK vychází vizuálně a funkčně z výše zmiňovaného registru číselníků JRC. Stojí na několika pilířích. Prvním z nich jsou data. Jejich struktura byla navržena a upravována průběžně podle toho, co se ukázalo jako potřebné v číselnicích evidovat. Data samotná jsou spravována v databázi Oracle a aktualizována jsou přímo v ní pomocí příkazů INSERT, UPDATE a DELETE. Nad daty databáze je postavena webová aplikace napsaná v jazyce Java sloužící k vizualizaci a procházení číselníků a jejich

hodnot. Kromě vizualizace v jazyce HTML umožňuje systém distribuci dat ve formátech XML a JSON. Do budoucna je plánována ještě podpora linked datových formátů jako je RDF/XML nebo Turtle. Nejjednodušším výdejním formátem je pak CSV, o jehož implementaci se také uvažuje.

Podkapitola 3.1 – Datový model a uložení dat v databázi

Model byl zvláště navržen pro číselníky a pro jejich obsah. Každá číselníková hodnota patří pouze do jednoho číselníku. V publikační databázi vznikly dvě nové tabulky, PUB_CODELISTS a PUB_CODELIST_VALUES a jejich struktura je zřetelná z obrázku 1.

<<table>> PUB_CODELISTS	<<table>> PUB_CODELIST_VALUES
+IDENT: VARCHAR2 +URI: VARCHAR2 +TITLE_CZ: VARCHAR2 +TITLE_EN: VARCHAR2 +DESCRIPTION_CZ: VARCHAR2 +DESCRIPTION_EN: VARCHAR2 +VALID_FROM: DATE +VALID_TO: DATE	+IDENT: VARCHAR2 +URI: VARCHAR2 +TITLE_CZ: VARCHAR2 +TITLE_EN: VARCHAR2 +DESCRIPTION_CZ: VARCHAR2 +DESCRIPTION_EN: VARCHAR2 +MAPPING_CODESPACE: VARCHAR2 +MAPPING_VALUE: NUMBER +VALID_FROM: DATE +VALID_TO: DATE +CODELIST: VARCHAR2 +HILUCS_CODE: NUMBER +SYMBOL_LOCATION: VARCHAR2 +FEATURE_TYPE: VARCHAR2

Obr. 1. Struktura relací v publikační databázi obsahující informace o číselnících a jejich hodnotách

Atributy popisující název, popis a časový cyklus objektů jsou zřejmé. Zastavme se u hodnoty URI. Ta obsahuje trvalý identifikátor číselníku nebo jeho hodnoty v absolutní cestě. Plně tedy stačí k identifikaci prvku. Číselníkové hodnoty obsahují údaje potřebné k mapování hodnoty ke konkrétním prvkům. Jedná se o dvojici atributů MAPPING_CODESPACE a MAPPING_VALUE, kde první obsahuje název atributu u mapovaného prvku v publikační databázi, a druhý jeho hodnotu. Například číselníková hodnota <http://services.cuzk.cz/registry/codelist/BoundaryTypeValue/StateBoundary> označující státní hranici bez značky se mapuje všem prvkům, které mají v atributu TYPPOD_KOD (MAPPING_CODESPACE) hodnotu 20100 (MAPPING_VALUE). Konečně atribut CODELIST definuje, kterému číselníku hodnota náleží. Poslední tři uvedené hodnoty jsou nepovinné a objevují se jen ve specifických případech: HILUCS_CODE propojuje hodnotu z číselníku s hodnotou HILUCS pro využití území a druh pozemku, SYMBOL_LOCATION obsahuje odkaz na mapovou značku, kterou má být daný prvek zobrazován v mapě a FEATURE_TYPE slouží v kontextu rozšíření tématu Parcely k podrobnějšímu určení typu jiného prvku mapy.

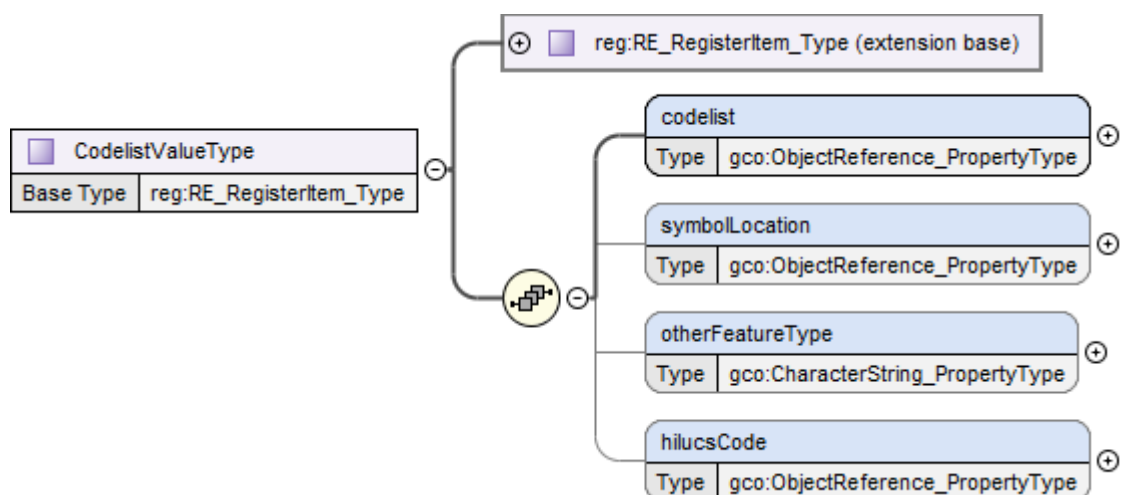
Podkapitola 3.2 – Aplikace

Aplikace sloužící pro zobrazování a procházení dat je dostupná na adrese <http://services.cuzk.cz/registry/codelist>. Při pohledu na úvodní stránku je zřejmá inspirace registrem JRC. Aplikace byla napsána v jazyce Java a umožňuje vyhledávání v číselnících a jejich hodnotách a jejich zobrazení v různých jazycích (k dispozici jsou čeština a angličtina) a v různých formátech (XML a JSON), jejichž struktura je popsána v další kapitole.

Podkapitola 3.1 – Výstupní data

Výstupní data jsou publikována (kromě HTML podoby) ve formátech XML a JSON. Pro formát XML byl jako základ využit profil ISO 19135. Pro tento profil existuje XML schéma. Pro publikaci číselníkových hodnot je však poměrně nepoužitelné, protože neumožňuje snadné přidání atributů, se kterými ISO 19135 nepočítá.

Z tohoto důvodu bylo pro číselníkové hodnoty vytvořeno rozšíření profilu vycházející z ISO 19135, které ho rozšiřuje o prvky `codelist`, `symbolLocation`, `otherFeatureType` a `hilucsCode`, jak je vidět na obrázku 2.



Obr. 2. Rozšíření schématu XML pro ISO 19135

Pro číselníky je reprezentace s využitím XSD pro ISO 19135 dostatečná. Informace navíc se vyskytují pouze v číselníkových hodnotách, na které je z číselníku pouze odkazováno (pomocí technologie XLINK). Rozšíření je dostupné na <http://services.cuzk.cz/xsd/inspire/cuzkCodelistValue/1.0/CuzkCodelistValue.xsd>.

Na rozdíl od XML nemá JSON žádný dokument, který by definoval strukturu dat (jako XSD v případě XML). To je nevýhoda pro uživatele, kteří tak musí spoléhat na to, že poskytovatel strukturu dat nijak nezmění. Jinak je formát JSON poměrně populární, protože zápis pomocí JSON je v porovnání s XML mnohem úspornější. Z důvodu absence schématu pro JSON vychází struktura dat v tomto formátu ze struktury XML, která je mírně zjednodušená. Ukázka dat ve struktuře JSON je na obrázku 3.

```

{
  "CodelistValue": {
    "itemIdentifier": "14",
    "name": "ArableGround",
    "status": "valid",
    "definition": "Orná půda",
    "description": "Orná půda je užívána v zemědělství k pěstování plodin.",
    "codelistValueUri": { "uri": "http://services.cuzk.cz/registry/codelist/LandTypeValue/ArableGround" },
    "codelist": { "uri": "http://services.cuzk.cz/registry/codelist/LandTypeValue" },
    "symbolLocation": { "uri": "http://services.cuzk.cz/registry/codelist/BoundaryTypeValue/StateBoundary/StateBoundary.svg" },
    "otherFeatureType": "Building",
    "hilucsCode": { "uri": "http://inspire.ec.europa.eu/codelist/HILUCSValue/1_1_1_CommercialAgriculturalProduction" }
  }
}

```

Obr. 3. Ukázka dat ve formátu JSON

KAPITOLA 4 – ZÁVĚR

Rozšiřování starých a vytváření nových číselníků se v kontextu implementace směrnice INSPIRE a především při jejím rozšiřování jeví jako nutný krok. Pro vedení vlastního registru číselníku je nutné mít server, na kterém registr poběží a navrhnout vhodný systém pro uchování a publikování dat, v porovnání s implementací dat či služeb INSPIRE se však nejedná o nic složitého. Je potřeba server, databáze a jednoduchá webová aplikace.

Do budoucna uvažujeme o rozšíření poskytovaných formátů o linked datové formáty (RDF/XML a/nebo turtle), případně CSV.

LITERATURA

Poláček, Jiří (2012) Implementing INSPIRE for Czech Cadastre of Real Estates. Geoinformatics FCE CTU, Praha, vol. 8, strany 9-16.

Van Loenen B., Grothe M. (2014) INSPIRE Empowers Re-Use of Public Sector Information. International journal of Spatial Data Infrastructures Research, vol. 9, strany 59-85.

ISO 19135-1:2015 (2015) Geographic information – Procedures for item registration. ISO/TC 211. Švýcarsko

Nařízení komise (EU) č. 1089/2010 (2010) Implementační pravidla pro interoperabilitu prostorových data a služeb. Dostupné z: <http://eur-lex.europa.eu/legal-content/CS/TXT/?uri=CELEX%3A32010R1089>