Nástroj pro katalogizaci mapových děl

Jiří Kozel   
Masarykova univerzita, Přírodovědecká fakulta, Geografický ústav

v2014-08-15

Nástroj umožňuje katalogizátorovi jednoduše předvyplnit metadata katalogizovaného mapového listu podle předem připravené šablony a zkopírovat tato metadata pro další použití, např. pro systém Aleph.

Nástroj je dostupný na stránce <http://mapseries.georeferencer.org/jk/mapseries.html>.

Zdrojový kód: <https://github.com/klokan/mapseries> (přístup přiděluje Klokan).

# Úvod

## Slovník pojmů

**Mapové dílo** je souhrn mapových listů, které pokrývají souvisle území, jehož zobrazení v daném měřítku není možné na jedné mapě. Mapové dílo má jednotný klad mapových listů, systematické označení mapových listů, jednotné mapové značky, jednotné kartografické zobrazení a zpravidla jednotné měřítko. [TSZKN] V anglicky psané literature se používá termín “**map series**”.

**Mapový list** je samostatná jednotka mapového díla vyhotoveného obvykle v souvislém kladu listů [TSZKN] V anglicky psané literature se používá termín “**map sheet**”.

**Klad mapových listů** je způsob rozdělení a uspořádání souvislého mapového díla na jednotlivé mapové listy. [TSZKN] V anglicky psané literatuře pravděpodobně neexistuje ustálené synonymum.

## Postup přidání mapového díla do nástroje

Přidání nového mapového díla vyžaduje několik kroků:

1. Příprava šablony
2. Příprava kladu mapových listů
3. Publikace kladu pomocí GeoServeru
4. Zaregistrování mapového díla do nástroje

Časově nejnáročnější jsou první dva kroky.

# Příprava šablony

**Šablona** je textový soubor, který obsahuje nekompletní metadatový záznam jednoho mapového listu daného mapového díla. Nekompletní proto, že některé jeho části jsou nahrazeny **proměnnými** (v matematickém či programátorském slova smyslu), které reprezentují některé vlastnosti mapových listů. Tyto proměnné jsou v průběhu samotné katalogizace postupně nahrazovány hodnotami, čímž vznikne kompletní záznam určený ke zkopírování.

## Proměnné a vlastnosti

**Proměnná** reprezentuje proměnlivou, variantní nebo vícenásobnou vlastnost mapových listů daného mapového díla.

### Proměnlivá vlastnost

Proměnlivá vlastnost mapových listů je taková, která není shodná pro všechny listy mapového díla, ale je shodná pro exempláře jednoho mapového listu.

Příkladem proměnlivé vlasnosti může být označení mapového listu, ale také třeba jeho název nebo pokrytá oblast.

### Variantní vlastnost

Variantní vlastnost mapových listů je taková, která není shodná pro různé exempláře alespoň jednoho mapového listu daného mapového díla.

Příkladem variantní vlastnosti je jazyk dokumentu u 3VM, který se může lišit exemplář od exempláře (některé exempláře jsou vytištěny v češtině, jiné v němčině nebo třeba v maďarštině).

Dalším příkladem variantní vlastnosti je název listu 3VM, který je na některých examplářích v české podobě a na jiných v německé (např. Brno vs. Brünn). Na rozdíl od jazyku dokumentu jsou varianty různé list od listu.

### Vícenásobná vlastnost

Vícenásobná vlastnost mapových listů je taková, která může pro jeden exemplář mapového listu nabývat více hodnot.

Příkladem vícenásobné vlastnosti je název pokryté oblasti mapového listu, kdy jeden mapový list může pokrývat i více oblastí (např. list 36°50° 3VM pokrývá oblast Ostrava i oblast Opava).

## Jak určit proměnné

Pravděpodobně nejsnazším způsobem, jak začít určovat proměnné, je vzít si několik hotových metadatových záznamů mapových listů daného díla a srovnat je. Místa, kde se od sebe jednotlivé záznamy liší, jsou kandidáty na proměnné. Následně je třeba případ od případu posoudit, o jakou vlastnost se jedná.

## Závislost hodnot proměnných

Často se stává, že hodnota jedné vlastnosti přímo závisí na hodnotě jiné vlastnosti. Například na vydavatelské korporaci závisí místo vydání i bibliografický předchůdce. Vytipováním těchto závislostí a jejich následným zápisem do šablony lze velmi urychlit proces katalogizace. Katalogizátorovi následně stačí určit pouze vydavatelskou korporaci a místo vydání i bibliografický předchůdce jsou doplněny automaticky.

Speciálním případem závislých proměnných jsou situace, kdy si hodnoty dvou (nebo dokonce více) proměnných vždy odpovídají. Příkladem může být rok vydání, který se objevuje v metadatovém záznamu na několika místech.

Nástroj umožňuje zapisovat jednoduché závislosti, kdy hodnota jedné proměnné je přímo závislá na hodnotě jiné proměnné. Na jedné proměnné může být závislé více proměnných, ale jedna proměnná může být závislá na maximálně jedné jiné proměnné. Proměnná, na jejíž hodnotě je závislá hodnota jiné proměnné, nesmí být sama závislá. Počet možných hodnot obou proměnných musí být v rámci každého mapového listu shodný. Ani jedna z proměnných nesmí být vícenásobná.

## Jak vytvořit šablonu

Nejjednodušším způsobem je zkopírovat hotový metadatový záznam nějakého mapového listu daného díla a uližit jej do samostatného textového souboru. Následně je třeba místa v záznamu, která obsahují hodnoty výše vytipovaných proměnných nahradit zápisem těchto proměnných.

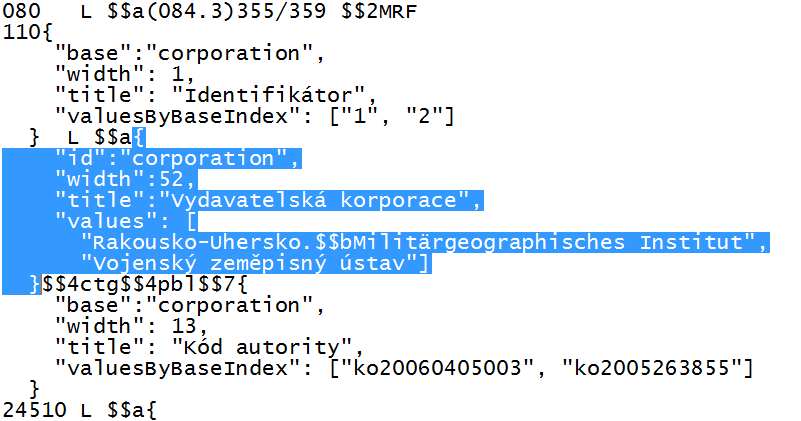
Příklad existující šablony 3VM je k vidění zde:

<http://mapseries.georeferencer.org/jk/templates/ah-ms3-200.txt>

## Syntaxe proměnných v šabloně

Každá proměnná je v metadatovém záznamu reprezentována objektem zapsaným ve **striktní** [JSON](http://www.json.org/) notaci. Každý takový objekt musí mít několik atributů, které specifikují, jakým způsobem bude při katalogizaci nahrazena daná proměnná konkrétní hodnotou.

Na následujícím obrázku je modře vyznačená jedna proměnná s atributy id, width, title a values. Před ní je na obrázku zapsaná jiná proměnná s atributy base, width, title a valuesByBaseIndex. Na konci posledního řádku je vidět začátek další proměnné (složená závorka), jejíž atributy již vidět nejsou.



Je důležité si uvědomit, že všechen text, který je vně objektů proměnných, včetně mezer, tabulátorů a konců řádků, se objeví ve výsledném metadatovém záznamu. Pouze místo objektů proměnných se objeví vyplěné hodnoty. Obrázek nahoře by se po nahrazení proměnných „smrskl“ na tři řádky začínající čísly 080, 110 a 24510 (v případě, že by žádná z vyplněných hodnot neobsahovala znak konce řádku).

Na následujícím obrázku je vidět šablona tak, jak vypadá, když ji vyplňuje katalogizátor. Zvýrazněná proměnná z předchozího obrázku je reprezentovaná rozbaleným seznamem hodnot. Žlutě jsou podbarvena pole proměnných, která jsou na vydavatelské korporaci závislá.



### Atributy proměnných

|  |  |  |
| --- | --- | --- |
| Atribut | Typ | Popis |
| base | Text | Přítomnost tohoto atributu indikuje, že se jedná o závislou proměnnou a naopak nepřítomnost tohoto atributu indikuje, že se jedná o nezávislou proměnnou. Hodnotou samotného atributu je identifikátor jiné proměnné, na jejíž hodnotě je závislá hodnota této proměnné. |
| enabled | Boolean | Určuje, zda katalogizátor může měnit hodnotu nezávislé proměnné. Výchozí hodnota je true. Tento atribut má smysl přenastavit na false pouze v případě nevariantních nezávislých proměnných, které mají jedinou hodnotu.  Hodnoty závislých proměnných katalogizátor měnit nemůže. |
| Id | Text | Identifikátor proměnné. V šabloně nesmí existovat dvě proměnné se stejným identifikátorem. Atribut není povinný, ale je nutné jej uvést u proměnné, na níž je závislá jiná proměnná, aby se na ni dalo odkazovat atributem base. |
| multipleValues | Boolean | Indikuje, zda je hodnota vícenásobná. Výchozí hodnota je false. Atribut může být true pouze u proměnných, které nejsou závislé a zároveň na nich žádná proměnná nezávisí. |
| title | Text | Název pole. Objeví se v nástroji při najetí myši na pole dané proměnné a v případě variantní proměnné i jako nápověda přímo v poli. |
| width | Číslo | Přibližná šířka pole, které se objeví v nástroji na místě proměnné. Použitou jednotkou jsou znaky textu. |
| value | Hodnota / Odkaz | Hodnota nebo odkaz na hodnotu(-y) dané proměnné. Pozor na to, že odkaz může reprezentovat i více hodnot. Způsob zápisu viz dále.  Tento atribut nesmí být uveden u závislých proměnných. |
| values | pole Hodnot / Odkazů | Pole hodnot nebo odkazů na hodnoty dané proměnné. V jednom poli mohou být jak hodnoty, tak odkazy. Pozor na to, že odkaz může reprezentovat i více hodnot. Způsob zápisu viz dále.  Tento atribut nesmí být uveden u závislých proměnných. |
| valuesByBaseIndex | pole Hodnot / Odkazů | Vyplňuje se pouze u závislých proměnných, a to pouze u těch, jejichž hodnota není rovna hodnotě proměnné, na které závisí. U závislých proměnných je automaticky vybrána hodnota, která má takovou pozici (index) v rámci možných hodnot, jakou má zvolená hodnota proměnné, na které závisí, v rámci svých možných hodnot.  Jedná se o pole hodnot nebo odkazů na hodnoty dané proměnné. V jednom poli mohou být jak hodnoty, tak odkazy. Pozor na to, že odkaz může reprezentovat i více hodnot. Způsob zápisu viz dále. |
| valueSeparator | Text | Oddělovač hodnot vícenásobné proměnné. Má smysl jej uvádět pouze pokud je multipleValues true. Konec řádku lze zapsat jako „\n” (zpětné lomítko a malé en). |

### Hodnoty a Odkazy na hodnoty

Ve třech atributech objektu proměnné – value, values a valuesByBaseIndex – se vyskytuje zápis hodnot nebo odkazů na hodnoty.

**Hodnota** je vždy reprezentována libovolným textovým řetězcem (a naopak – pokud se na daném místě objeví textový řetězec, je tento považovaný za hodnotu). **Hodnota** se objeví nezměněná v poli formuláře nebo jako jedna z možných variant.

**Odkaz na hodnotu(-y)** je vždy reprezentován objektem zapsaným ve validní [JSON](http://www.json.org/) notaci. Může mít dva atributy – attr a formatFunction – které jsou popsány dále. V určitých případech může odkaz reprezentovat i více hodnot.

Odkaz na hodnotu si lze představit jako převodní funkci, která má dva vstupní parametry

* **stringValue**: textový řetězec
* **sheet**: objekt typu OpenLayers.Feature.Vector

a jeden výstup, kterým je hodnota, která se objeví nezměněná v poli formuláře nebo jako jedna z možných variant.

Pokud je uveden atribut **attr**, dosadí se za první parametr – stringValue – hodnota z příslušného sloupce tabulky. Pokud atribut attr uveden nebyl a jedná se o závislou proměnnou, dosadí se za první parametr hodnota proměnné, na níž závisí tato proměnná. Za druhý argument – sheet – se vždy dosadí objekt, který představuje mapový list včetně geometrie.

Pokud je uveden atribut **formatFunction**, aplikuje se daná formátovací funkce na oba argumenty a její návratová hodnota se objeví v poli formuláře nebo jako jedna z možných variant. Pokud atribut formatFunction uveden nebyl, objeví se nezměněná hodnota parametru stringValue v poli formuláře nebo jako jedna z možných variant.

Jak již bylo řečeno, může odkaz reprezentovat i více hodnot. To nastane v případě, že hodnota ve sloupci tabulky určeném atributem attr obsahuje znak „|“, který slouží jako oddělovač hodnot. V tom případě se výše popsaný postup aplikuje na každou hodnotu zvlášť a výsledkem je stejný počet hodnot, jaký je v daném sloupci uložen.

V hodnotách, které jsou specifikovány uvnitř šablony, je lepší se znaku „|“ zcela vyhnout.

#### Atributy odkazů na hodnoty

|  |  |  |
| --- | --- | --- |
| Atribut | Typ | Popis |
| attr | Text | Název sloupce tabulky kladu mapových listů, ve kterém je uložena hodnota (nebo více hodnot), která vstupuje do výše popsané převodní funkce.  Více hodnot je možné do sloupce uložit s pomocí oddělovače „|“. Toho lze s výhodou využít v případě variantních nebo vícenásobných proměnných. |
| formatFunction | Text | Název formátovací funkce, která je aplikována na dva argumenty – stringValue a sheet (viz výše). V nástroji existují dvě připravené formátovací funkce:   * ms:marc21\_0341\_bbox * ms:marc21\_255\_bbox\_czech   Tyto formátovací funkce vrací prostorový rozsah daného mapového listu pro příslušná pole metadatového záznamu MARC21.  Kromě toho je možné si pro každou mapovou sérii v konfiguračním souboru nadefinovat vlastní formátovací funkci (viz dále) a na ni se zde odkázat. |

## Jak zapsat jednotlivé typy proměnných

### Proměnlivá vlastnost

Hodnoty proměnlivé vlastnosti je ideální zapsat a následně získat z tabulky nebo z geometrie kladu mapových listů. V šabloně se na takovou hodnotu lze odkázat odkazem s nastaveným atributem attr a/nebo formatFunction. Příkladem ze šablony 3VM budiž pole 043, 0341 nebo 255.

### Variantní vlastnost

U variantních vlastností je situace složitější.

Pokud pro všechny exempláře všech mapových listů existuje omezené množství možných hodnot (například jazyk dokumentu nebo vydavatelská korporace), je nejjednodušším způsobem zápis možných hodnot přímo do šablony. Příkladem ze šablony 3VM jsou pole 0410 nebo 110$$a.

Pokud pro všechny exempláře jednoho mapového listu existuje velké množství možných hodnot a tyto hodnoty se navíc vyskytují i u většiny jiných listů (například rok vydání nebo číslo vydání), je nejjednodušším způsobem možné hodnoty neuvádět a nechat na katalogizátorovi jejich explicitní doplnění. Příkladem ze šablony 3VM jsou pole 008.1 nebo 300.

Pokud pro exempláře alespoň jednoho listu existuje malé množství možných hodnot a tyto hodnoty se u různých mapových listů neopakují (například název listu 3VM), je ideální zapsat hodnoty do jednoho sloupce tabulky kladu mapových listů a oddělit je znakem „|“. Příkladem ze šablony 3VM je pole 24510$$a.

Speciálním případem předchozího odstavce je potom situace, kdy sice existuje více možných hodnot, ale jedna z druhé je strojově odvoditelná. Příkladěm je označení mapového listu 3VM, které se vyskytuje ve dvou variantách se stupni a bez stupňů (34°49° vs. 3449). V tomto případě stačí do tabulky uložit pouze jednu hodnotu a druhou hodnotu potom získat formátovací funkcí. Příkladem ze šablony 3VM je pole 24510$$p.

### Vícenásobná vlastnost

Hodnoty vícenásobných vlastností, které se liší list od listu (například pokrytá oblast), je ideální zapsat do jednoho sloupce tabulky a oddělit je znakem „|“. Příkladěm ze šablony 3VM jsou pole 080 nebo 651.

### Závislá vlastnost

U závislých vlastností je situace složitější.

Pokud je hodnota závislé vlastnosti rovna hodnotě, na níž závisí, není třeba žádnou hodnotu uvádět. Příkladěm ze šablony 3VM jsou pole 260$$c, 4901.2 nebo 830.

Pokud se hodnota závislé vlastnosti liší, je nutné uvést možné hodnoty nebo odkazy do proměnné valuesByBaseIndex. V průběhu katalogizace je potom automaticky vybrána hodnota, která má takovou pozici (index) v rámci možných hodnot, jakou má zvolená hodnota proměnné, na které závisí, v rámci svých možných hodnot.

U proměnných, které mají atribut valuesByBaseIndex, je pak třeba se rozhodnout, zda budou jejich hodnoty zapsány přímo do šablony (příkladem ze šablony 3VM jsou pole 008.2, 008.3, 110, 110$$7, 24510$$h, 260$$a, 4901.1, 655) nebo do tabulky (pole 2463 druhá hodnota, 4901 druhá hodnota, 830 druhá hodnota) nebo budou získány převodem z původní hodnoty (2463 první hodnota, 4901 první hodnota, 830 první hodnota). Rozhodovací klíč je v tomto případě obdobný, jako u variantních vlastností.

# Příprava kladu mapových listů

Klad mapových listů je nutné **připravit jako soubor ve formátu ShapeFile (\*.shp)**. Každý mapový list je v něm reprezentován jedním polygonem a několika sloupci (atributy).

V tabulce kladu musí existovat všechny sloupce odkazované ze šablony a jejich hodnoty je třeba vyplnit. Názvy a typy sloupců mohou být libovolné (nejbezpečnější jsou typy text nebo číslo). Dva sloupce jsou povinné (vyžadované aplikací):

* SHEET, v němž je uloženo číslo mapového listu
* TITLE, v němž je uložen název mapového listu

Klad mapových listů je nejbezpečnější převést do souřadného systému EPSG:4326, ale GeoServer si většinou poradí i s jiným.

# Publikace kladu pomocí GeoServeru

ShapeFily s klady mapových listů jsou umístěny na Dropboxu ve složce /mapseries (neplést s /TEMAP/mapseries !!!). Přístup přiděluje Klokan. Ve složce /mapseries je třeba najít nebo vytvořit adresář reprezentující oblast, kterou mapové dílo pokrývá, a do něj nahrát připravený ShapeFile kladu mapových listů. Název ShapeFilu může být libovolný.

Po nahrání ShapeFilu na Dropbox je třeba se přihlásit na GeoServer: <http://mapseries.georeferencer.org/geoserver/>  
Přístup přiděluje Klokan.

Podrobnější návod, jak nahrát ShapeFile na GeoServer, je například zde: <http://docs.geoserver.org/2.1.4/user/gettingstarted/shapefile-quickstart/index.html>

## Jak vytvořit sklad

Prvním krokem je vytvořit nový sklad (**Store**) typu „Vector Data Source / Shapefile“. Parametr **Workspace** je třeba nastavit na hodnotu „mapseries“.

Parametr **Data Source Name** by měl vycházet z adresářové struktury, v níž je Shapefile umístěn a měl by obsahovat území mapového díla a název mapového díla. Doporučuji pužívat pouze malá písmena, čísla a pomlčky. Příklad pro 3VM 1:200K: „austria-hungary--3rd-military-survey--200“.

Parametr **Shapefile** location je cesta ke ShapeFilu, například „file:mapseries/austria-hungary/3rd-military-survey/3VM\_200\_4326\_join.shp“.

Parametr **DBF charset** nese informaci o použítém kódování znaků. U českých Windows to bude pravděpodobně hodnota „windows-1250“.

## Jak vytvořit vrstvu

Následuje vytvoření vrstvy (**Layer**). Vrstvu je třeba vybrat z vytvořeného skladu.

Parametr **Name** musí odpovídat názvu skladu.

Parametr **Native SRS**, resp. **Declared SRS**, musí odpovídat souřadnému systému daného Shapefilu.

Dále je třeba dopočítat oba parametry **Bounding Boxes** (na automatický výpočet zde dva odkazy).

V kartě **Publishing** je třeba ověřit, že je zvolen **Default Style** s hodnotou „polygon“.

# Zaregistrování mapového díla do nástroje

Samotný nástroj je přístupný na Dropboxu /mapseries‑www/jk (přístup přiděluje Klokan).

Zde umístěné soubory jsou viditelné na webu, například soubor /mapseries‑www/jk/mapseries.html je viditelný na adrese <http://mapseries.georeferencer.org/jk/mapseries.html>

Pozor na to, že se jedná o ostrou verzi nástroje, jakákoliv změna se okamžitě projeví i na webu. Proto je dobré před každou větší změnou celou složku /mapseries‑www/jk zálohovat.

## Nahrání šablony

Soubor se šablonou je třeba nahrát do adresáře /mapseries‑www/jk/templates

Jeho název by měl být bez diakritiky, mezer, apod.

## Přidání mapového díla do konfiguračního souboru

Konfigurační soubor je běžný JavaScriptový soubor umístěný zde: /mapseries‑www/jk/config.js

Každé mapové dílo je reprezentováno jedním objektem v poli mapseries.config.series. Do tohoto pole je tedy třeba přidat nový objekt a nastavit mu minimálně první tři z následujících atributů.

Jelikož se jedná o standardní JavaScriptový soubor, není třeba používat striktní notaci jako v případě JSON objektů v šabloně.

|  |  |  |
| --- | --- | --- |
| Atribut | Typ | Popis |
| title | Text | Skládá se z názvu území mapového díla, dvojtečky a z názvu samotného mapového díla.  Příklad: "Austria-Hungary: 3rd Military Survey, 1:200 K" |
| layer | Text | Název vrstvy kladu publikované na GeoServeru.  Příklad: „austria-hungary--3rd-military-survey--200“ |
| template | Text | Název souboru se šablonou.  Příklad: „ah-ms3-200.txt“ |
| formatFunctions | Objekt, jehož hodnoty jsou JavaScriptové funkce | Obsahuje formátovací funkce specifické pro dané mapové dílo. Parametry a návratová hodnota těchto funkcí jsou popsány výše v části věnující se šabloně.  Nepovinný atribut. |