
PSP Validator Documentation

Release 0.8.4

Jan Stavel

September 10, 2015

1	Ve zkratce	3
2	Co umí	5
3	Instalace	7
3.1	Windows	7
4	Volání	9
4.1	Přehled dostupných validací	10
4.2	příklad validace METS souboru	10
5	Jednotlivé validace	11
5.1	přehled	11
6	rady	13
6.1	Chci přehledný výpis	13
6.2	Chci validovat jen hlavní METS soubor	13
6.3	Chci vidět seznam validací, co program umí	13
6.4	Chci vidět informace o jedné validaci	13
6.5	Chci vidět pouze výsledky na konci	13
6.6	Chci vidět výsledky i průběh validace	14
6.7	Chci provést jen určitou validaci	14
6.8	Chci zachovat rozbalený balíček	14
6.9	Když program vypisuje hodně chyb	14

Contents

- *PSP validátor*
 - *Ve zkratce*
 - *Co umí*
 - *Instalace*
 - * *Windows*
 - *Volání*
 - * *Přehled dostupných validací*
 - * *příklad validace METS souboru*
 - *Jednotlivé validace*
 - * *přehled*
 - *rady*
 - * *Chci přehledný výpis*
 - * *Chci validovat jen hlavní METS soubor*
 - * *Chci vidět seznam validací, co program umí*
 - * *Chci vidět informace o jedné validaci*
 - * *Chci vidět pouze výsledky na konci*
 - * *Chci vidět výsledky i průběh validace*
 - * *Chci provést jen určitou validaci*
 - * *Chci zachovat rozbalený balíček*
 - * *Když program vypisuje hodně chyb*

Ve zkratce

Program validuje PSP balíček, nebo celý adresář PSP balíčků.

PSP se používá k uchování digitálních informací o periodikách a monografiích. Tento balíček používá [Národní knihovna](#) k digitalizaci a uchování informací o periodikách a monografiích.

Podrobnosti o formátech použitých v PSP balíčku jsou na stránkách [Národní knihovny](#).

Program je vystaven v repozitáři [Google Code](#).

- umi jednu verzi schematu
- schema se vyvyji
- budou chodit emaily
- komunikace v cestine
- repozitar na code.google.com
- software bude v anglictine
- prubezne verze standardu se nebudou publikovat
- zapojeni krajske knihovny do testovani - posilali by testovaci data pro testovani
- program by nemel skoncit na prvni chybe. protoze se schema vyvyji a neco uz nemusi byt chybou
- tentyz program by mel byt k dispozici na webu pro krajske knihovny, aby si mohly dopredu zkontrolovat data
- program bude mit volbu:
 - kontrolovat jen metadata (mets soubor)
 - zkontrolvoat vsechno, vcetne metadat
 - vypnout kontrolu ciselniku
- udelat instalator pro windows
- vystup bude v textovem souboru

Co umí

- [X] validuje hlavní soubor METS podle schématu METS
- [X] validuje položky dc v hlavním souboru METS podle schématu Dublin Core
- [X] validuje soubory v adresáři amdSec podle schématu METS
- [X] validuje jednotlivé části `premis:object` a `mix:mix` v souborech v adresáři `amdsec`
- [X] zkontroluje, že všechny linky, které jsou v souboru METS, existují
- [X] zkontroluje MD5 součet všech souborů, na které se soubor METS odkazuje
- [X] validuje soubory v adresáři ALTO podle schématu ALTO

Program vypisuje informace o průběhu validací na konzoli a současně do souboru `log/messages.log`.

Pokud program validuje více balíčků, tak ke každému balíčku doplní samostatný log soubor s výpisy průběhu validace.

Validovaný balíček rozbaluje do adresáře `tmp`.

Pokud se programu místo souboru zadá adresář, předpokládá, že v něm jsou PSP soubory. Ty všechny zvaliduje.

3.1 Windows

1. instalace python 2.7

- stáhnout aktuální verzi pythonu <http://www.python.org/ftp/python/2.7.3/python-2.7.3.msi> a nainstalovat ji.
- nastavit cesty k binárkám pythonu aby se dal volat python v cmd.

2. instalace lxml

- instalace binárek <http://www.lfd.uci.edu/~gohlke/pythonlibs/#lxml>
vybere se lxml-X.X.X.win*.exe Ten se nainstaluje.

3. instalace programu

- uloží se zip balíček ze sekce <https://code.google.com/p/psp-validator/downloads/list>
- rozbálí se do adresáře C:\Opt

Takže program bude v adresáři: C:\Opt\psp-validator

- do proměnné PATH se doplní cesta k programu: C:\Opt\psp-validator\bin

Volání

Pokud jsme ve *Windows*, tak se k cestě k programu přidá název disku:
 C:\Opt\psp-validator\bin\validate-ppsp.py --help

```
usage: validate-ppsp.py [-h] [-m] [-v] [--version] [-l] [-d] [-i INFO] [-p PARTIAL] [-s] [--normdir]
Program validuje PSP balíček.

Umí validovat na třech úrovních:
- METS soubor
- hodnoty z číselníků existují
- linky, na které se v balíčku odkazuje existují

Každá z těchto voleb se dá vypnout.

Program rozbalí zadaný PSP balíček do adresáře /opt/psp-validator/tmp/PSP-Validation-2012-11-19-
Pokud je zadaný adresář, tak vezme všechny soubory, co mají příponu =.zip=
a pokusí se je zkontrolovat.

Pokud se při kontrole souborů v adresáři amdSec, ... objeví chyba,
program skončí u prvního souboru s chybou.
Většinou se chyby opakují a tak by bylo hodně stejných chyb.

positional arguments:
PSP                      cesta k PSP balíčku. Muze to byt soubor, nebo adresar. Pokud to je adresar
                          vezme všechny soubory a zvaliduje je.

optional arguments:
-h, --help                show this help message and exit
-m, --mets                zkontroluje METS soubor
-v, --verbose             hlášky programu budou podrobnější
--version                 verze programu. Vypíše i verzi dokumentace, co popisuje strukturu PSP balí
-l, --list-validators     zobrazí seznam validací, co program umí
-d, --debug              zobrazí ladici hlasky
-i INFO, --info INFO     zobrazí popis vybrané validace
-p PARTIAL, --partial PARTIAL
                          zavola jen jeden konkretni krok validace. Seznam validaci vypisuje argument -
-s, --summary            na konci vypíše přehled testů, co provedl a s jakým skončily výsledkem.
--normdir                na konci se maže pracovní adresář. S tímto argumentem se adresář nesmaže.
-a, --all-files          Kdyz se validuji soubory v adresari, probere vsechny soubory, i kdyz se ob
```

4.1 Přehled dostupných validací

argument `-l` pomůže

```
/opt/psp-validator/bin/validate-ppsp.py -l

seznam validací:
  01_mets
  01_mets_dc
  01_mets_mods
  02_links_checksums
  02_links_exist
  03_techspecs
  03_techspecs_premis_mix
  04_altos
```

4.2 příklad validace METS souboru

program má validovat jen hlavní metadata

```
/opt/psp-validator/bin/validate-ppsp.py -v -s /opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec

2012-10-28 21:24:52,381 PSP_VALIDATION INFO - budu validovat soubor /opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
2012-10-28 21:24:52,381 PSP_VALIDATION INFO - pracuji v adresari: /opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
2012-10-28 21:24:52,381 PSP_VALIDATION INFO - validator: validace hlavního METS souboru
2012-10-28 21:24:55,745 PSP_VALIDATION INFO - validator: validace vnitřku METS souboru, specifikace
2012-10-28 21:24:56,723 PSP_VALIDATION INFO - validator: validace vnitřku hlavního METS souboru
2012-10-28 21:24:56,737 PSP_VALIDATION INFO - validator: kontrola CHECKSUM všech souborů na
2012-10-28 21:24:57,591 PSP_VALIDATION INFO - validator: validace linek v hlavním METS souboru
2012-10-28 21:24:57,653 PSP_VALIDATION INFO - validator: validace souborů ve složce amdSec
2012-10-28 21:24:57,716 PSP_VALIDATION INFO - validator: validace souborů ve složce amdSec
2012-10-28 21:24:57,731 PSP_VALIDATION ERROR - chyba validace: chyby validace souboru /opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
/opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
/opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
/opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
2012-10-28 21:24:57,805 PSP_VALIDATION ERROR - chyba validace: chyby validace souboru /opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
/opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
/opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
/opt/psp-validator/tmp/complete_NDK-0000000000008_1350896484227/amdSec
2012-10-28 21:24:57,806 PSP_VALIDATION INFO - validator: validace souborů v adresáři =ALTO=
2012-10-28 21:24:57,835 PSP_VALIDATION INFO - vysledky validace:
  01_mets : OK
  01_mets_dc : OK
  01_mets_mods : OK
  02_links_checksums : OK
  02_links_exist : OK
  03_techspecs : OK
  03_techspecs_premis_mix: Error
  04_altos : OK
```

Jednotlivé validace

Informace o jednotlivých validacích si můžeme zobrazit volbou `-i`

```
/opt/psp-validator/bin/validate-ppsp.py -i 01_mets

validace hlavního METS souboru
validuje hlavní METS soubor podle specifikace METS
```

5.1 přehled

01_mets	validace hlavního METS souboru	validuje hlavní METS soubor podle specifikace METS
01_mets_dc	validace vnitřku METS souboru, specifikace DC	validuje vnitřní položky DC v METS soubor podle specifikace Dublin Core
01_mets_mods	validace vnitřku hlavního METS souboru, specifikace MODS	validuje vnitřní položky MODS v hlavním METS soubor podle specifikace MODS
02_links_checksum	kontrola CHECKSUM všech souborů na které se v hlavním METS souboru odkazuje	zkontroluje, zda mají soubory, na které linky odkazují, správnou CHECKSUM.
02_links_exists	validace linek v hlavním METS souboru	zkontroluje, zda existují všechny soubory na které se odkazují linky v hlavním souboru METS.
03_tech-specs	validace souborů ve složce amdSec	zkontroluje předběžně jednotlivé soubory ve složce =amdSec= podle formátu METS
03_tech-specs_premis	validace souborů ve složce amdSec na technická metadata	zkontroluje, zda jednotlivé soubory odpovídají použitým schématům. To je formát METS a vevnitř jsou položky =premis:object= podle schématu PREMIS v2.1. a položky =mix:mix= ve formátu MIX

6.1 Chci přehledný výpis

Program na konci vypíše jen přehled provedených validací.

Při validaci jednotlivých souborů skončí při první chybě.

```
/opt/psp-validator/bin/validate-ppsp.py -s -o DIR
```

6.2 Chci validovat jen hlavní METS soubor

Program provede jen validace spojené s hlavním =METS= souborem.

```
/opt/psp-validator/bin/validate-ppsp.py -m -s /opt/psp-validator/tmp/complete_NDK-00000000000008_135089
```

6.3 Chci vidět seznam validací, co program umí

Program vypíše seznam dostupných validací.

```
/opt/psp-validator/bin/validate-ppsp.py -l
```

6.4 Chci vidět informace o jedné validaci

Program vypíše podrobné informace o vybrané validaci.

```
/opt/psp-validator/bin/validate-ppsp.py -i 04_altos
```

6.5 Chci vidět pouze výsledky na konci

Program na konci vypíše přehled provedených validací a jejich výsledek.

```
/opt/psp-validator/bin/validate-ppsp.py -s /opt/psp-validator/tmp/complete_NDK-00000000000008_135089648
```

6.6 Chci vidět výsledky i průběh validace

Program bude průběžně informovat o tom, co provádí a na konci vypíše přehled provedených validací a jejich výsledek.

```
/opt/psp-validator/bin/validate-ppsp.py -v -s /opt/psp-validator/tmp/complete_NDK-00000000000008_135089
```

6.7 Chci provést jen určitou validaci

Program provede jen vybranou validaci a skončí.

```
/opt/psp-validator/bin/validate-ppsp.py -p 03_techspecs /opt/psp-validator/tmp/complete_NDK-0000000000
```

6.8 Chci zachovat rozbalený balíček

Program na konci ponechá rozbalený balíček.

```
/opt/psp-validator/bin/validate-ppsp.py -s -p 03_techspecs --normdir /opt/psp-validator/tmp/complete_N
```

6.9 Když program vypisuje hodně chyb

- pustím program se souhrnným výpisem na konci
- zjistím, které validace neprošly
- pustím program s dotyčnou validací a aby skončil při první chybě a zachoval rozbalený balíček

```
/opt/psp-validator/bin/validate-ppsp.py -v -s -p 03_techspecs --normdir /opt/psp-validator/tmp/co
```