

ZAF – Validátor SIP dle Národního standardu pro elektronické systémy spisové služby



ARCHIV
UNIVERZITY KARLOVY
Univerzita Karlova

LightComp
Enhance Your Technology

Petr Cajthaml – Tomáš Dvořák – Petr Pytelka – Zdeněk Vašek



Validátor ZAF

- ▶ Aplikace pro kontrolu SIP balíčků z ESSS dle Nár. standardu, analytický nástroj, bez formální garance shody s validátorem NA
- ▶ Společný vývoj Archivu hl. m. Prahy, LightComp v.o.s. a Archivu UK, uzavřené Memorandum
- ▶ Autory současné aplikace Stanislav Bečvář, Petr Pytelka a Karel Žáček
- ▶ Reakce na aktuální možnosti validátoru NA, potřeba využít validátor v provozu – hromadné testování (test, produkce)
- ▶ Doplněna nová validační pravidla



Odchytky proti řešení NA


- ▶ Samostatná aplikace, lze zapojit do vlastního systému (např. integrovat do ESSS, digitálního archivu)
- ▶ Řádková aplikace
- ▶ Dávkové zpracování
- ▶ Strojově i lidsky čitelný výsledek validace, vysvětlení chyby a odkaz na předpis, lokalizace chyby v rámci SIP
- ▶ Nová validační pravidla pro kontrolu metadat a komponent



Validační pravidla

- ▶ 110 validačních pravidel, 404 vzorových testovacích balíčků
- ▶ Obsah škodlivého kódu
- ▶ Kontrola struktury
- ▶ Logická kontrola obsahu
- ▶ Formátová kontrola dle NSESS a pravidel v InfoListu NA (dále rozvíjena)
- ▶ Pravidla aplikována podle typu SIP

Dokumentovaná a zveřejněná validační pravidla

Kontrola škodlivého kódu			
vir1	Datový balíček SIP neobsahuje hrozbu.	1, 2, 3	§ 21 odst. 6 vyhlášky č. 259/2012 Sb.
Kontrola datové struktury			
data1	Datový balíček SIP je soubor v datovém formátu ZIP (jeho MIME Content-type je detekován jako application/zip) nebo složka.	1, 2, 3	Požadavek 11.2.9 a 11.2.1 NSESSS.
data2	Pokud je datový balíček SIP komprimován do souboru v datovém formátu ZIP, po rozbalení obsahuje právě jednu složku. Ta má stejný název jako je název souboru v datovém formátu ZIP.	1, 2, 3	Požadavek 11.2.9 NSESSS.
data3	Složka obsahuje právě jeden soubor pojmenovaný mets.xml nebo právě jeden soubor pojmenovaný mets.xml a složku pojmenovanou komponenty.	1, 2, 3	Požadavek 11.2.2, 11.2.3 a 11.2.8 NSESSS.
<i>Další kontroly se provádí nad souborem mets.xml.</i>			
Kontrola znakové sady			
kod1	Znakovou sadou souboru je Unicode/UCS v kódování UTF-8 bez BOM (Byte order mark).	1, 2, 3	Požadavek 11.2.7 NSESSS.
Kontrola správnosti XML			
wf1	Soubor je well-formed.	1, 2, 3	Požadavek 11.2.2 NSESSS.
Kontrola jmenných prostorů XML			
ns1	Soubor obsahuje právě jeden kořenový element <mets:mets>.	 1, 2, 3	Bod 2.1. přílohy č. 3 NSESSS.

Výstupy validace (ukázka) – pravidlo 60, chyba

```
<pravidlo id="obs58" text="Jakýkoli element <lt;nse:sss:Identifikator> obsahuje atribut zdroj s neprázdnou hodnotu." zdroj="Příloha č. 2 NSESSS, ř. 288." stav="ok"/>
<pravidlo id="obs59" text="Žádná entita (od spisového plánu po dokument) nebo objekt <lt;nse:sss:Komponenta>, <lt;nse:sss:BezpecnostniKategorie>, <lt;nse:sss:SkartacniRezim> nebo <lt;nse:sss:TypDokumentu> neobsahuje stejné hodnoty elementu <lt;nse:sss:Identifikator> a jeho atributu zdroj a současně odlišné hodnoty v ostatních elementech, jako má jiná entita nebo objekt uvedeného typu, kromě atributu ID uvedené entity." zdroj="Příloha č. 2 NSESSS, ř. 123." stav="ok"/>
<pravidlo id="obs60" text="Pokud jakýkoli element <lt;nse:sss:Dokument> obsahuje v hierarchii dětských elementů <lt;nse:sss:EvidencniUdaje>, <lt;nse:sss:Manipulace> element <lt;nse:sss:AnalogovyDokument> s hodnotou ne, obsahuje element <lt;nse:sss:Dokument> dětský element <lt;nse:sss:Komponenty>." zdroj="Příloha č. 2 NSESSS, ř. 45." popisChyby="Chybi popis komponenty (počítačového souboru) dokumentu v digitální podobě." vypisChyby="Nenalezen povinný element <lt;nse:sss:Komponenty>. Dokumentu (Ident. hodnota: MP12P00BTZ3Z, zdroj: Gordic.Ginis.MP12.X)." mistoChyby="Řádek 17:44, element <lt;nse:sss:Dokument>." stav="chyba"/>
<pravidlo id="obs61" text="Pokud jakýkoli element <lt;nse:sss:Dokument> obsahuje v hierarchii dětských elementů
```


Výstupy validace (ukázka) – pravidlo 66, chyba

```
<pravidlo id="obs65" text="Pokud je základní entitou díl (<nsesss:Dil>) nebo spis (<nsesss:Spis>), potom v hierarchii dětských elementů <nsesss:EvidencniUdaje>, <nsesss:Vyzarovani>, <nsesss:DataceVyzazeni> obsahuje element <nsesss:RokSkartačniOperace> hodnotu, která je rovna nebo větší vyšší hodnotě, přičemž jednou hodnotou je součet hodnoty elementu <nsesss:RokSpousteciUdalosti>, 1 a hodnoty elementu <nsesss:SkartačniLhuta> uvedeného v rodičovském elementu <nsesss:SkartačniRezim> a druhou hodnotou nejvyšší hodnota součtu hodnoty elementu <nsesss:DatumDoruceni> nebo <nsesss:DatumVytvoreni> (v závislosti na tom, zda jde o doručený nebo vlastní dokument), 1 a hodnoty elementu <nsesss:SkartačniLhuta> jakékoli dětské entity dokument (nsesss:Dokument>)." zdroj="$ 15 odst. 5 vyhlášky č. 259/2012 Sb." stav="ok"/>
```

```
<pravidlo id="obs66" text="Pokud je základní entitou díl (<nsesss:Dil>), spis (<nsesss:Spis> nebo dokument (<nsesss:Dokument>), potom v hierarchii dětských elementů <nsesss:EvidencniUdaje>, <nsesss:Vyzarovani>, <nsesss:DataceVyzazeni> obsahuje element <nsesss:RokSkartačniOperace> hodnotu, která je menší nebo rovna aktuálnímu roku." zdroj="$ 20 odst. 1 vyhlášky č. 259/2012 Sb." popisChyby="U dílu, spisu nebo dokumentu nelze provést skartační řízení, protože ještě nenadešel uváděný rok skartační operace." vypisChyby="Hodnota roku elementu <nsesss:RokSkartačniOperace> je větší, než aktuální rok. Hodnota: 2099. Díl (Ident. hodnota: MP12P0000009, zdroj: neGordic)." mistoChyby="Řádek 166:45, element <nsesss:RokSkartačniOperace>." stav="chyba"/>
```



Funkce validátoru a technické možnosti

- ▶ Nenáročný na provoz, možné integrovat do jiných aplikací
- ▶ Testování dávek stovek tisíc SIP
- ▶ Postupná validace podle jednotlivých sad pravidel
- ▶ OS Microsoft Windows, GNU/Linux, OS X nebo Oracle Solaris
- ▶ Java 1.8 a vyšší



Zveřejnění a vývoj

- ▶ Vývoj v AHMP od 2015, v ostrém provozu, od 2021 v Archivu UK a ESSS UK
- ▶ Zveřejnění zdrojového kódu a dokumentace v GitHubu (do konce 2021), možnost zadávání issues, vzorové testovací balíčky
- ▶ Licence Apache License 2.0
- ▶ Otevřené pro všechny uživatele



Možní uživatelé

- ▶ Zřizovatelé digitálních archivů
- ▶ Archiváři testující skartace z ESSS
- ▶ Původci provozující ESSS, digitální správní archivy
- ▶ Vývojáři a dodavatelé ESSS



Možnosti spolupráce

- ▶ Uvítáme využití dalšími uživateli, zpětnou vazbu, hlášení chyb
- ▶ Testování i vývoj (optimalizace, refaktorizace)
- ▶ Metodická spolupráce (zpřesňování pravidel, tvorba nových)

DĚKUJEME ZA POZORNOST!

- ❖ petr.cajthaml@ruk.cuni.cz
- ❖ tomas.dvorak@praha.eu
- ❖ petr.pytelka@lightcomp.cz
- ❖ zdenek.vasek@ruk.cuni.cz



ARCHIV
UNIVERZITY KARLOVY
Univerzita Karlova

LightComp
Enhance Your Technology