



NA vs. ZAF

Souboj validátorů balíčků SIP

Tomáš Dvořák, Daniel Šurda
Archiv hlavního města Prahy



Validátory

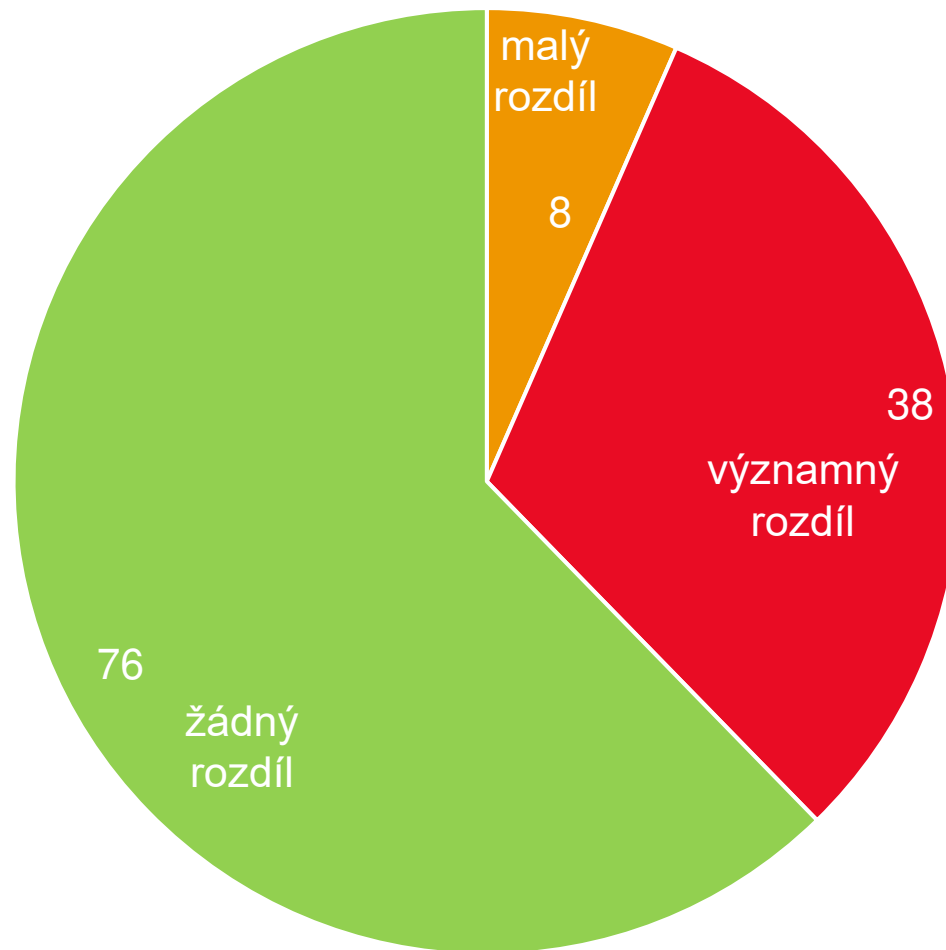
 Národní archiv		
2014	vznik	2021 (předchůdce SipValid 2015)
10^7	SIPy	10^5-10^6
124	pravidla	116
Národní archiv	vývoj	Archiv hl. m. Prahy, Archiv UK, LightComp, v.o.s.
https://validatorsip.nacr.cz/	odkaz	https://validatorzaf.github.io/zaf/

Disciplína

- validace balíčků SIP podle přílohy č. 2 národního standardu pro elektronické systémy spisové služby (schéma XML pro vytvoření datového balíčku SIP a pro zaznamenání popisných metadat uvnitř datového balíčku SIP) s účinností od 1. 7. 2023
- posouzení pravidel validace (hodnocení rozsahu a správnosti validace)
- neposuzuje se výkon, použitelnost, uživatelská přívětivost apod.

Výsledek

– 122 srovnání



Typy rozdílů a důvody (1)

– opominutí

- **přehlednutý nedodělek** → **bude opraveno**
- **záměr** (je nastavena větší tolerance, než stanovují předpisy) → ponechá se

– odlišné interpretace

- **chyby** → bude opraveno
- **nejasné záležitosti** (nelze nalézt jednotný výklad) → předá se k řešení pracovní skupině pro NSESSS

Příklad (1)

NSESSS definuje právě jeden výslyt elementu <fileGrp>, pokud je uveden element <fileSec>.

1.14 Skupina souborů (komponent)

Element zaznamenává soubory (komponenty).

Element	<mets:fileGrp>
Typ	složený datový typ (kontejner)
Povinnost	ANO
Opakovatelnost	NE

Jeden validátor kontroluje, druhý nikoli.

Typy rozdílů a důvody (2)

– opominutí

- přehlédnutý nedodělek → bude opraveno
- **záměr (je nastavena větší tolerance, než stanovují předpisy) → ponechá se**

– odlišné interpretace

- chyby → bude opraveno
- nejasné záležitosti (nelze nalézt jednotný výklad) → předá se k řešení pracovní skupině pro NSESSS

Příklad (2)

Pro oddělovač v URL v atributu `xlink:href` v rámci elementu `<FLocat>` se používá znak lomítka („/“) viz RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax.

```
A path consists of a sequence of path segments separated by a slash ("/") character. A path is always defined for a URI, though the defined path may be empty (zero length). Use of the slash character to indicate hierarchy is only required when a URI will be used as the context for relative references. For example, the URI <mailto:fred@example.com> has a path of "fred@example.com", whereas the URI <foo:///info.example.com?fred> has an empty path.
```

Jeden validátor kontroluje, druhý umožňuje i backslash („\“) a nebude zpřísněno.

Typy rozdílů a důvody (3)

– opominutí

- přehlédnutý nedodělek → bude opraveno
- záměr (je nastavena větší tolerance, než stanovují předpisy) → ponechá se

– odlišné interpretace

- **chyby** → **bude opraveno**
- nejasné záležitosti (nelze nalézt jednotný výklad) → předá se k řešení pracovní skupině pro NSESSS

Příklad (3)

Informační list NA, čá. 6/2020, č. 3/2020 usměrňuje vyplňování metadat komponent.

Komponenta->Verze

- Prvek pro uvedení pořadového čísla verze komponenty. Pořadové číslo komponent v dokumentu tvoří souvislou řadu od 1 do n, kde 1 je prvotní koncept nebo doručená komponenta a n je komponenta poslední, která je vždy ve výstupním datovém formátu. Verze s nejvyšším číslem je poslední.
- V SIP se může vyskytovat více verzí jedné komponenty (musí mít stejné číslo pořadí a stejný druh).
- Číselná řada verzí nemusí být v SIP souvislá, ale musí být použito pořadí verzí z eSSL (nikoliv nově generované pro balíček SIP); když bude např. předána jen první a poslední verze z třinácti, budou mít číslo verze 1 a 13.

Jeden validátor chybně vyžaduje verzi 1 komponenty, druhý nikoli.

Typy rozdílů a důvody (4)

– opominutí

- přehlédnutý nedodělek → bude opraveno
- záměr (je nastavena větší tolerance, než stanovují předpisy) → ponechá se

– odlišné interpretace

- chyby → bude opraveno
- **nejasné záležitosti (nelze nalézt jednotný výklad) → předá se k řešení pracovní skupině pro NSESSS**

Příklad (4)

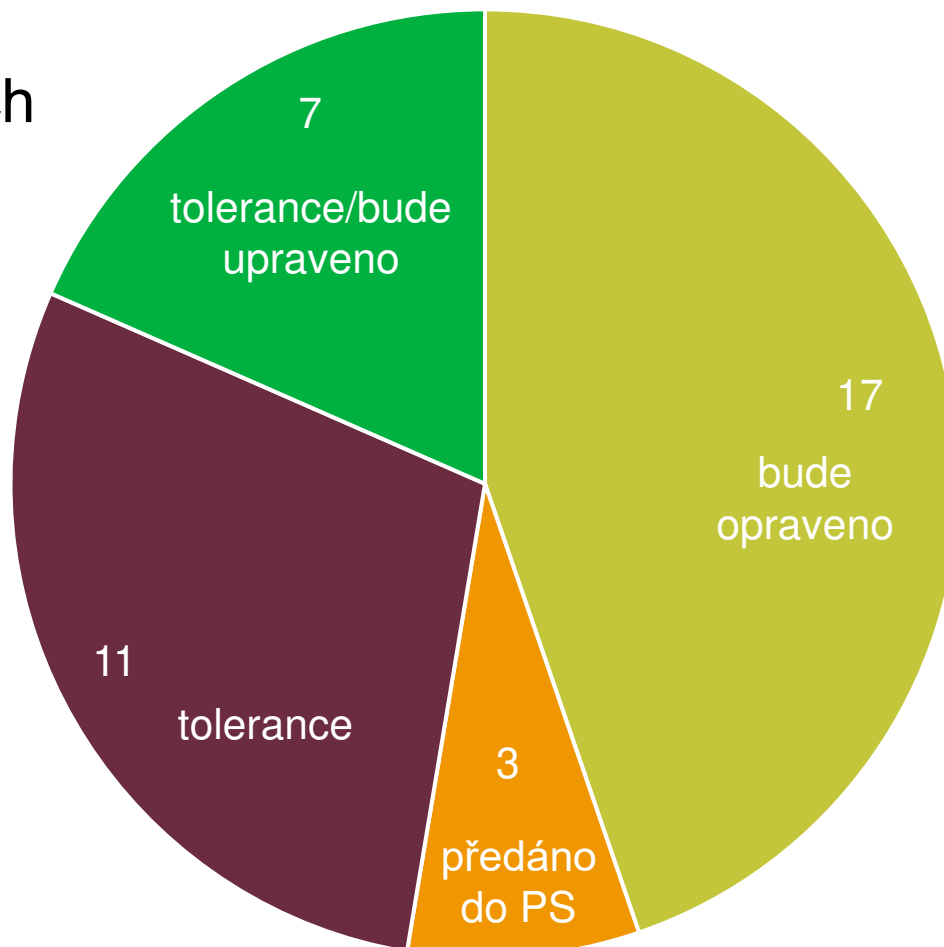
NSESSS definuje element <MaterskeEntity> jako nepovinný s komentářem „Prvek je povinný pouze v případě, že je entita "dokument" použita jako základní entita.“

```
<xs:element name="MaterskeEntity" type="tMaterskeEntityDokumentu" minOccurs="0">  
  <xs:annotation>  
    <xs:documentation>Prvek je povinný pouze v případě, že je entita "dokument" použita jako  
    |základní entita.</xs:documentation>  
  </xs:annotation>  
</xs:element>
```

Jeden validátor kontroluje vazbu dokumentu jako základní entity a <MaterskeEntity>. Druhý kontroluje, že neexistuje <MaterskeEntity>, pokud je dokument ve spisu/dílu.

Co s tím?

– 38 významných rozdílů



Závěr

- přehlédnuté nedodělky a chyby budou opraveny
- nejasné záležitosti jsou předány do pracovní skupiny pro NSESSS
- zůstanou malé (nevýznamné) rozdíly a určitá míra tolerance

**Díky J. Bernasovi
a P. Nimrichtrové z NA**

