



EVROPSKÁ UNIE  
Evropský fond pro regionální rozvoj  
Integrovaný regionální operační program



MINISTERSTVO  
PRO MÍSTNÍ  
ROZVOJ ČR



MINISTERSTVO VNITRA  
ČESKÉ REPUBLIKY

## Dodatek č. 3 Detailního návrhu technického řešení informačních systémů e- Sbírka a e-Legislativa

### Projekt e-Sbírka a e-Legislativa

*Připraveno pro:*

**Ministerstvo vnitra ČR**

21. 6. 2016

Verze 1.0, Finální

*Připravil:*

**MVČR**



# Změny a schválení

## Změny

Datum	Autor	Verze	Popis změn
21. 6. 2016	Tým architekta eSeL	1.0	Finální verze

## Revize

Jméno	Schválená verze	Funkce	Datum
Mgr. A. Gola	1.0	Věcný gestor projektu	21. 6. 2016

# Obsah

<b>1</b>	<b>Úvod</b>	<b>5</b>
<b>2</b>	<b>Proces Tvorby datové báze</b>	<b>6</b>
<b>2.1</b>	<b>Vztah tvorby a verifikace datové báze</b>	<b>6</b>
<b>2.2</b>	<b>Cíle, předpoklady, vstupy a výstupy verifikace datové báze</b>	<b>9</b>
2.2.1	Cíle verifikace datové báze	10
2.2.2	Předpoklady pro verifikaci datové báze	10
2.2.3	Vstupy pro verifikaci datové báze	12
2.2.4	Výstupy procesů verifikace datové báze	16
<b>2.3</b>	<b>Časová osa</b>	<b>17</b>
2.3.1	Časová souslednost tvorby a verifikace datové báze	17
<b>2.4</b>	<b>Akceptační proces tvorby datové báze</b>	<b>17</b>
<b>3</b>	<b>Verifikace datové báze</b>	<b>21</b>
<b>3.1</b>	<b>Verifikace datové báze vyhlášených znění</b>	<b>21</b>
3.1.1	Kontrola rekonstrukce textů	21
3.1.2	Kontrola úplnosti obsahu vyhlášených znění	24
3.1.3	Kontrola správnosti tabulek	25
3.1.4	Kontrola správnosti netextových entit	26
3.1.5	Nalezení překlepů	27
3.1.6	Kontrola indexace předpisů	29
<b>3.2</b>	<b>Verifikace datové báze konsolidovaných znění</b>	<b>30</b>
3.2.1	Ověření protokolů o provedení konsolidace	31
3.2.2	Komparace datové báze vůči nezávislému zdroji	33
3.2.3	Analýza konfliktů v konsolidacích	35
3.2.4	Kontrola odkazového aparátu	36
<b>3.3</b>	<b>Kontrola normalizace obsahu datové báze</b>	<b>38</b>
3.3.1	Cíle kontroly normalizace obsahu datové báze	38
3.3.2	Kvantifikace kontroly normalizace obsahu datové báze	38
3.3.3	Vstupy kontroly normalizace obsahu datové báze	38



3.3.4	Předpoklady procesu kontroly normalizace obsahu .....	40
3.3.5	Popis procesu kontroly normalizace obsahu .....	41
3.3.6	KPI procesu kontroly normalizace obsahu (pro <i>Dodavatele DB</i> ).....	42
<b>3.4</b>	<b>Kontrola tezauru CzechVoc.....</b>	<b>44</b>
3.4.1	Cíle kontroly tezauru CzechVoc .....	46
3.4.2	Kvantifikace kontroly tezauru CzechVoc .....	46
3.4.3	Zvláštní vstupy kontroly tezauru CzechVoc .....	46
3.4.4	Zvláštní předpoklady procesu kontroly tezauru CzechVoc .....	47
3.4.5	Popis procesu kontroly tezauru CzechVoc .....	47
3.4.6	KPI procesu kontroly tezauru CzechVoc (pro <i>Dodavatele DB</i> ).....	47



# 1 Úvod

Tento dokument „Dodatek č. 3 k *Detailnímu návrhu technického řešení informačního systému e-Sbírka a e-Legislativa*“ přímo navazuje na dokument „*Detailní návrh technického řešení informačního systému e-Sbírka a e-Legislativa*“, který doplňuje.

Verifikace datové báze představuje dosažení stavu, kdy datovou bázi českých sbírek právních předpisů může *Zadavatel* prohlásit za hodnověrnou a jako takovou ji

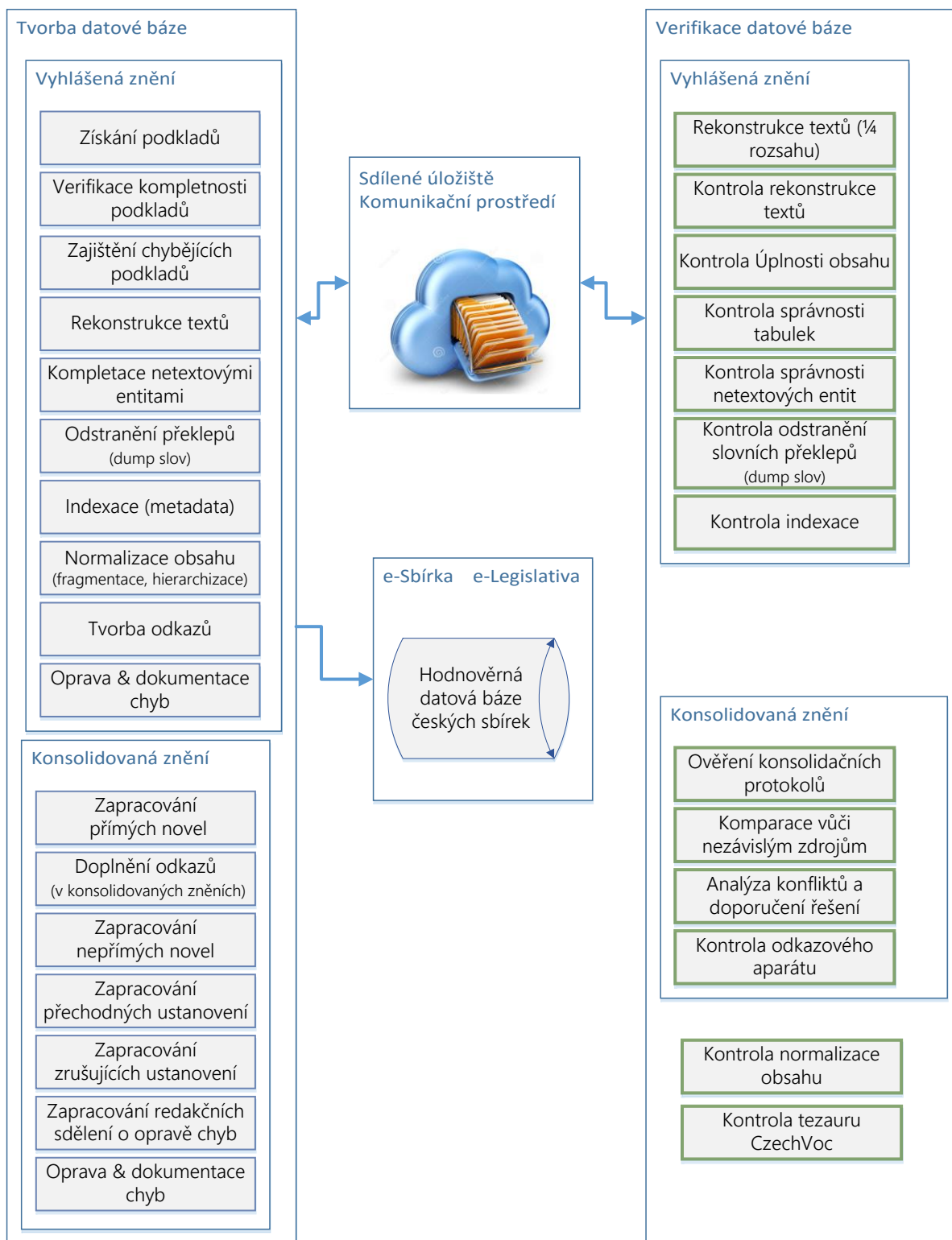
1. Poskytnout publiku prostřednictvím portálu e-Sbírka.
2. Poskytnout k dalšímu rozvoji v legislativních procesech nástroji e-Legislativy.

Verifikace datové báze je tedy

1. Ověřením kvality práce *Dodavatele datové báze*.
2. Souborem činností vedoucím ke korekci databáze tak, aby splňovala kritéria kvality definovaná dále.

Verifikace datové báze musí být koncipována takovým způsobem, aby poskytla *Zadavateli* všechny podklady k prohlášení datové báze za hodnověrnou.

## 2 Proces Tvorby datové báze



Obrázek č. 1: Souhrn fází a procesů tvorby a verifikace datové báze

### 2.1 Vztah tvorby a verifikace datové báze

Procesy verifikace datové budou probíhat současně s procesy tvorby datové báze.

Dodavatel tvorby datové báze je v dalším textu uváděn jako *Dodavatel DB*.

Dodavatel verifikace datové báze je v dalším textu uveden jako *Verifikátor DB*.

*Dodavatel DB* vytváří datovou bázi postupně po ročnících jednotlivých sbírek (od nejstaršího ročníku po nejnovější). Tvorba datové báze je znázorněna levou částí Obrázek č. 1 bude probíhat ve dvou základních fázích, kde každá fáze je rozdělena do několika procesů:

## 1. Tvorba datové báze vyhlášených znění

### 1.1. Získání podkladů, tedy všech částek, všech předpisů, všech sbírek pro následnou rekonstrukci textů

- zajištění listinných podkladů vhodných k naskenování a OCR
- a/nebo
- zajištění digitálních replik listinných podkladů vhodných buď k OCR, nebo extrakci textu

### 1.2. Verifikace kompletnosti a kvality podkladů

- ověření, zda podklady získané v předchozím kroku, jsou z pohledu kompletnosti a kvality použitelné pro rekonstrukci textů

### 1.3. Získání podkladů nových

- získání podkladů alternativních, vyhodnocených v předchozím kroku jako nekompletních nebo nekvalitních

### 1.4. Rekonstrukce textů

- vytěžení textů a tabulek vyhlášených znění z podkladů získaných v předchozích krocích
    - skenováním a OCR
- a/nebo
- extrakcí textů z PDF
- a/nebo
- manuálním přepisem/značkováním
- a/nebo
- kombinací předchozích způsobů

### 1.5. Kompletace netextovými entitami

- doplnění obrázků, souborových příloh, přepis vzorců do strukturovaného popisu

### 1.6. Odstranění překlepů

- specifický proces rozpadu všech rekonstruovaných textů na jednotlivá slova, kontrola těchto slov s cílem nalezení slov chybně rekonstruovaných a jejich následná oprava (pomocí tzv. dumpu slov)

#### 1.7. Indexace

- přiřazení uzlů tematické sítě (subsetu CzechVoc tezauru) k předpisům
- postupný rozvoj tematické sítě (subsetu CzechVoc)

#### 1.8. Normalizace obsahu

- fragmentace: určení struktur všech fragmentů (např. §, odstavec, položka číslovaného seznamu, položka nečíslovaného seznamu, písmeno, část, hlava, nadpis, kapitola, pravidlo atd. ...) extrakce pořadí těchto struktur (např. 1 pro *Část první, Hlava I, § 1* atd.)
- hierarchizace: sestavení fragmentů do hierarchií (tedy stromové struktury)

#### 1.9. Tvorba odkazů

- odkaz: vytvoření obousměrné vazby mezi částí textů, která je citací jiného předpisu nebo jeho části a citovaným předpisem nebo jeho částí
- typizace odkazu, tj. pojmenování vztahu
  - existují např. odkazy interní (uvnitř předpisu), externí (mezi předpisy), prováděcí atd.

#### 1.10. Dokumentace a oprava nalezených chyb

- zadavatel může rozhodnout, že některé chyby předloh (tedy originálů nebo replik originálů) mohou být opraveny, budou-li řádně zdokumentovány

### 2. Tvorba datové báze konsolidovaných znění Zpracování přímých novel

- zpracování každé novely (nebo více novel se stejnou účinností) způsobí vznik konsolidovaného znění s účinností zpracované novely (resp. více novel)
  - přímá novela: explicitní instrukce z vyhlášeného předpisu určující co, jak a v jakém dříve vyhlášeném předpisu změnit
- každé konsolidované znění má účinnost novely (nebo novel), které jeho změny oproti předchozímu konsolidovanému znění způsobily

#### 2.2. Doplnění odkazů v konsolidovaných zněních

- ustanovení modifikovaná zpracováním novel mohou obsahovat nové odkazy a je třeba je dopracovat podle stejného principu jako v bodě 1.9 výše

#### 2.3. Zpracování nepřímých novel

- nepřímá novela: změna dříve vyhlášeného předpisu na základě právního výkladu explicitních instrukcí přímých novel (např. neexistuje novela, která by



explicitně instruovala globální změnu Kčs na Kč, přestože Kčs od roku 1993 byla nahrazena za Kč, různých typů nepřímých novel je celá řada)

#### 2.4. Zapracování přechodných ustanovení

- přechodné ustanovení je pojmenováním instrukce k výkladu právního předpisu, kterou lze někdy zapracovat formou nepřímé novely, většinou to ale možné není a s takovýmto ustanovením je nutno zacházet jako s výjimkou vztaženou k určitým více či méně specifikovaným ustanovením předpisu

#### 2.5. Zapracování zrušujících ustanovení

- zrušující ustanovení je explicitní instrukce ke zrušení primárního předpisu, která má dopad resp. následek nepřímého zrušení také sekundárních novelizačních ustanovení (nebo přímo celých předpisů) k rušenému předpisu s účinností identickou se zrušením primárního předpisu

#### 2.6. Zapracování redakčních sdělení o opravě chyby

- zvláštní typ změny vyhlášeného předpisu, který má za výsledek jeho změnu účinností původního vyhlášení předpisu

#### 2.7. Dokumentace a oprava nalezených chyb

- některé novely nejsou zapracovatelné jednoznačně, případně nejsou zapracovatelné vůbec
- výsledkem tohoto procesu je zachycení těchto stavů, aby mohly být následně *Verifikátorem DB* analyzovány a ve spolupráci se zadavatelem vyřešeny

Paralelně s těmito procesy bude *Dodavatel DB* ještě zjišťovat absenci a dostupnost oznámených ve Sbírce zákonů a Sbírce mezinárodních smluv, tato činnost však nepodléhá verifikaci.

Výstup ročníku každé fáze každé sbírky bude *Dodavatelem DB* předán do procesů verifikace datové báze vykonávané *Verifikátorem DB*.

Procesy verifikace budou probíhat ve třech fázích:

1. Verifikace datové báze vyhlášených znění
2. Verifikace datové báze konsolidovaných znění
3. Verifikace normalizace obsahu datové báze

Tyto tři fáze verifikace datové báze jsou detailně popsány dále v rozdělení na samostatné procesy s jejich podrobným popisem.

## 2.2 Cíle, předpoklady, vstupy a výstupy verifikace datové báze

V této kapitole jsou popsány společné *cíle, předpoklady, vstupy a výstupy* společné bud' pro všechny, nebo pro více dílčích procesů verifikace datové báze *Verifikátorem DB*.

## 2.2.1 Cíle verifikace datové báze

**Cíle** verifikace datové báze vykonávané *Verifikátorem DB* jsou následující:

- Ověřit kvalitu výstupů jednotlivých procesů tvorby datové báze *Dodavatelem DB* prostřednictvím nastavených ukazatelů požadované kvality (KPI) specifikovaných pro jednotlivé procesy.
- Poskytnout *Dodavateli DB* zpětnou vazbu v takové podobě, aby případné chyby byl schopen opravit a po jejich odstranění mohl vytvořená dat použít pro další následující procesy tvorby datové báze.
- Zajistit kvalitu formou splnění KPI metrik jednotlivých procesů – tedy zajistit, aby *Dodavatel DB* předal datovou bázi *Zadavateli* v nejvyšší dosažitelné kvalitě.
- Poskytnout *Zadavateli* všechny podklady k výkonu jeho kontroly s výsledkem prohlášení datové báze vytvořené *Dodavatelem DB* za hodnověrnou.

## 2.2.2 Předpoklady pro verifikaci datové báze

V této části jsou specifikovány **předpoklady** pro verifikaci datové báze sumárně pro všechny dílčí procesy.

### 2.2.2.1 Prostředí pro všechny typy interakcí mezi aktéry verifikace

*Verifikátor DB* vytvoří prostředí pro komunikaci a všechny typy interakcí mezi *Dodavatelem DB*, *Verifikátorem DB* a *Zadavatelem*. *Dodavatel DB* bude povinen toto prostředí používat dle propozic poskytnutých *Verifikátorem DB*.

Rámcové požadavky na prostředí pro komunikaci:

- Sdílené úložiště pro výměnu souborů
  - *Dodavatel DB* publikuje na sdílené úložiště HTML znění v syntaxi rámcově popsané v kapitole 2.2.3. Vždy celý ročník sbírky najednou pro danou fázi procesu tvorby datové báze
  - *Verifikátor DB* ze sdíleného úložiště odebere vždy celý ročník sbírky pro danou fázi procesu tvorby datové báze také najednou.
  - Součástí dodávky ročníku *Dodavatelem DB* bude i postupná dodávka tezauru CzechVoc, který bude vždy dodán v celkové aktualizované podobě.
  - Kolaborativní prostředí pro zajištění interakce *Dodavatele DB* a *Verifikátora DB*.
- Evidence a nezměnitelná historie všech aktivit *Dodavatele DB* i *Verifikátora DB*.
- Rozhraní celkových výsledků pro následnou kontrolu ze strany *Zadavatele*.

Podle uvážení může *Verifikátor DB* takové prostředí využívat také jako interní workflow systém pro výkon procesů verifikace.

### 2.2.2.2 Lidské zdroje pro výkon verifikace

*Verifikátor DB* zajistí dostatek lidských zdrojů, aby byl schopen provést všechny procesy verifikace datové báze nejdéle za jeden týden pro jeden ročník. Je nutné počítat s postupnou rozpracovaností minimálně 3-5 ročníků zároveň.

Celkový objem je následující

- Sbírka zákonů 1945-2019: 74 ročníků
- Sbírka mezinárodních smluv 2000-2019: 20 ročníků
- Úřední list 1945-1962: 17 ročníků

### 2.2.2.3 Technická infrastruktura pro výkon verifikace

*Verifikátor DB* zajistí technickou infrastrukturu pro výkon verifikace, tedy příslušené hardwarové a softwarové vybavení pro svůj tým, rámcově se jedná o tyto bloky infrastruktury:

- OCR 100.000 stran textů sbírek, tedy 200.000 – 250.000 normostran.  
Složitost textů lze označit za vysokou (viz archiv digitálních stejnopisů sbírek zveřejněný na <http://aplikace.mvcr.cz/sbirka-zakonu>).
- Nástroje pro textové porovnání přibližně 40.000 - 45.000 předpisů ze dvou zdrojů s analýzou a evidencí nalezených rozdílů (komparační nástroje).  
Procesní detaily jsou uvedeny v popisech dotčených procesů níže.
- Fulltextovou indexaci předpisů v HTML formátu, kde výsledkem bude seznam všech použitých slov, jejich četnost a přesná pozice v indexovaném souboru.
- Prostor pro řízení a týmový výkon procesů verifikace datové báze:
  - hardware a software pro příslušníky týmu
  - prostředí pro týmovou práci a řízení projektu, jak je uvedeno výše. Podle uvážení *Verifikátora DB* může takovým prostředím být nastavba prostředí popsaného výše v kapitole 2.2.2.1.

### 2.2.2.4 Digitální stejnopisy vyhlášených znění Sbírky zákonů a Sbírky mezinárodních smluv

*Zadavatel* pro procesy verifikace datové báze poskytne digitální stejnopisy Sbírky zákonů a Sbírky mezinárodních smluv (dále jen *digitální stejnopisy*).

*Digitální stejnopisy* jsou zveřejněny ve formátu PDF a na denní bázi aktualizovány zde: <http://aplikace.mvcr.cz/sbirka-zakonu>.

Digitalizaci předpisů, které nejsou v této formě dostupné (např. některé rozsáhlé přílohy, resp. sbírka Úřední list) provede *Dodavatel DB* a poskytne *Verifikátorovi DB*.

### 2.2.3 Vstupy pro verifikaci datové báze

V této části jsou shromážděny *vstupy* společné pro více než jeden proces verifikace datové báze. Vyžaduje-li nějaký proces samostatné vstupy, jsou tyto specifikovány vždy u konkrétního procesu.

1. Soubory jednotlivých znění (tedy vyhlášeného znění a znění konsolidovaných) předpisů sbírek ve formátu HTML vložené *Dodavatelem DB* do sdíleného úložiště.

Syntaxe pojmenování souborů {sb}{yyyy}c{nnn}z{pppp}-v{0/YYYYMMDD}.htm, kde

- {sb} představuje kód sbírky – Sbírka zákonů, Sbírka mezinárodních smluv, Úřední list
- {yyyy} představuje rok vyhlášení předpisu čtyřmístně
- {nnn} je trojciferné číslo částky obsahující vyhlášený předpis zleva doplněné nulami
- {pppp} je čtyřmístné číslo předpisu zleva doplněné nulami
- 0/YYYYMMDD specifikuje znění
  - v0 pro vyhlášené znění
  - vYYYYMMDD pro konsolidované znění, kde YYYYMMDD je datem začátku účinnosti konsolidovaného znění

Příklad:

Soubor s vyhlášeným zněním předpisu Vyhláška č. 78/2013 Sb., o energetické náročnosti budov,

bude pojmenován sb2013c036z0078-v0.htm.

Příklad struktury HTM souboru předpisu

- hlavička s metadaty předpisu
- tělo předpisu, kde každý samostatný řádek je uzavřen elementem <p>...</p>

- orientační příklad pro sb2013c036z0078-v0.htm:

```
sb2013c036z0078.htm x
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
2 <html>
3
4 <head>
5 <meta http-equiv="content-type" content="text/html; charset=WINDOWS-1250">
6 <meta name="title" content="Vyhláška o energetické náročnosti budov">
7 <meta name="year" content="2013">
8 <meta name="reference" content="78">
9 <meta name="type" content="Vyhláška">
10 <meta name="author" content="Ministerstvo průmyslu a obchodu">
11 <meta name="adopted" content="22.3.2013">
12 <meta name="inforce" content="1.4.2013">
13 <meta name="validity" content="ANO">
14 <meta name="volume" content="36">
15 <meta name="keywords" content="">
16 <meta name="amending" content="">
17 <meta name="derogating" content="148/2007 Sb.">
18 <meta name="basedon" content="406/2000 Sb.">
19 <meta name="czechvoc" content="Energetika; Stavby">
20 </head>
21
22 <body>
23 <p>78</p>
24 <p>VYHLÁŠKA</p>
25 <p>ze dne 22. března 2013</p>
26 <p>o energetické náročnosti budov</p>
27 <p>Ministerstvo průmyslu a obchodu stanoví podle § 14 odst. 4 zákona č. 406/2000 Sb., o hospodaření energií, ve znění zákona č. 165/2012 Sb. a zákona č. 318/2012 Sb., k provedení § 7 odst. 8 a § 7a odst. 6 zákona</p>
28 <p>1</p>
29 <p>Předmět úpravy</p>
30 <p>Tato vyhláška zpracovává příslušný předpis Evropské unie<sup>1</sup> a stanoví</p>
31 <p>a) základní optimální úroveň požadavků na energetickou náročnost budovy pro nové budovy, větší změny dokončených budov, jiné než větší změny dokončených budov a pro budovy s téměř nulovou spotřebou energie,</p>
32 <p>b) metodu výpočtu energetické náročnosti budovy,</p>
33 <p>c) vzor posouzení technické, ekonomické a ekologické proveditelnosti alternativních systémů dodávek energie,</p>
34 <p>d) vzor stanovení doporučených opatření pro snížení energetické náročnosti budovy,</p>
35 <p>e) vzor a obsah průkazu a způsob jeho zpracování a</p>
36 <p>f) umístění průkazu v budově.</p>
37 <p>2</p>
38 <p>Základní pojmy</p>
39 <p>Pro účely této vyhlášky se rozumí</p>
40 <p>a) referenční budovu výpočtově definovaná budova téhož druhu, stejného geometrického tvaru a velikosti včetně prosklených ploch a částí, stejné orientace ke světovým stranám, stínění okolní zástavbou a přírodními překážkami, stejného vnitřního uspořádání a se stejným typickým užíváním a stejnými uvažovanými klimatickými údaji jako hodnocená budova, avšak s referenčními hodnotami vlastností budovy, jejích konstrukcí a technických systémů budovy,</p>
41 <p>b) typickým užíváním budovy obvyklý způsob užívání budovy v souladu s podmínkami vnitřního a venkovního prostředí a provozu stanovený pro účely výpočtu energetické náročnosti budovy,</p>
42 <p>c) venkovním prostředím venkovní vzduch, vzduch v přilehlých nevytápěných prostorech, přilehlá zemina, sousední budova a jiná sousední zóna,</p>
43 <p>d) vnitřním prostředím prostředí uvnitř zóny, které je definováno návrhovými hodnotami teploty, relativní vlhkosti vzduchu a objemového toku výměny vzduchu, případně rychlosti proudění vnitřního vzduchu a požadované intenzity osvětlení uvnitř zóny,</p>
44 <p>e) přirozeným větráním větrání založené na principu teplotního a tlakového rozdílu vnitřního a venkovního vzduchu,</p>
45 <p>f) nuceným větráním větrání pomocí mechanického zařízení,</p>
```

- syntaxe zápisu doplňkových entit

- Tabulky

tabulka 2 předpisů Vyhláška č. 78/2013 Sb., o energetické náročnosti budov

**Tab. 2 – Referenční parametry a hodnoty pro měněné stavební prvky obálky budovy**

Parametr	Označení	Jednotka	Referenční hodnota
Součinitel prostupu tepla	$U_R$	$W/(m^2 \cdot K)$	Doporučená hodnota dle ČSN 730540-2:2011

bude zapsána takto:

```
<p>Tab. 2 – Referenční parametry a hodnoty pro měněné stavební prvky obálky budovy</p>
<table border="1" width="800">
<tr>
<th width="200" id="r1_s1">Parametr</th>
<th width="100" id="r1_s2">Označení</th>
<th width="100" id="r1_s3">Jednotka</th>
<th width="400" id="r1_s4">Referenční hodnota</th>
</tr>
<tr>
<td width="200" id="r2_s1">Součinitel prostupu tepla</td>
<td align="center" width="100" id="r2_s2"> $U_{R}$ </td>
<td align="center" width="100" id="r2_s3"> $W/(m^{sup}2</sup> \cdot K)$ </td>
<td align="center" width="400" id="r2_s4">Doporučená hodnota dle ČSN 730540-2:2011</td>
</tr>
</table>
```

v prohlížeči pak takto:

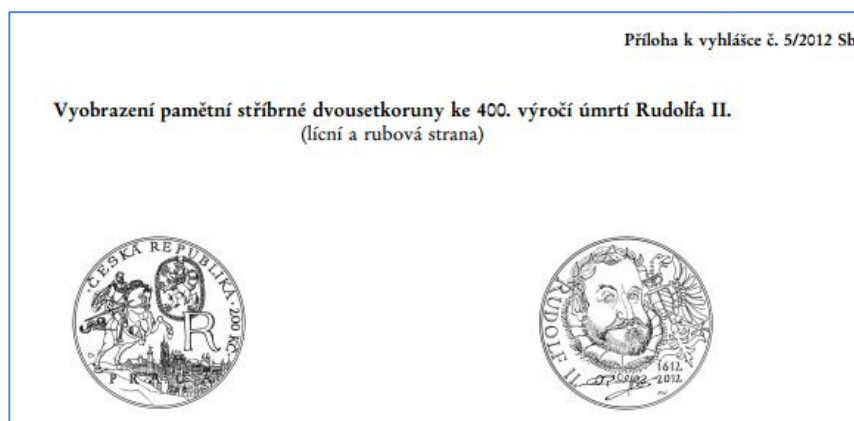
Tab. 2 - Referenční parametry a hodnoty pro měněné stavební prvky obálky budovy

Parametr	Označení	Jednotka	Referenční hodnota
Součinitel prostupu tepla	$U_R$	W/(m <sup>2</sup> ·K)	Doporučená hodnota dle ČSN 730540-2:2011

- Obrázky

budou zapsané prostřednictvím HTML tagu ``

Příklad:



bude zapsána takto:

```
<p>Příloha k vyhlášce č. 5/2012 Sb.</p>
<h3>Vyobrazení pamětní stříbrné dvousetkoruny ke 400. výročí úmrtí Rudolfa II.</h3>
<p>(lící a rubová strana)</p>
<p>


</p>
```

syntaxe pojmenování souborů obrázků

- {sb}{yyyy}c{nnn}z{pppp}o{ooo}.png
  - význam {sb}{yyyy}c{nnn}z{pppp} identický jako u pojmenování souboru předpisu
  - {ooo} je třímístné číslo udávající pořadí obrázku v předpisu zleva doplněné nulami
  - specifikace znění není nutná, neboť obrázek vždy pochází z vyhlášeného znění

- Souborové přílohy

- souborové přílohy se použijí pro např. pro
  - formuláře, u kterých je důležitá originální vizuální podoba
  - skupiny stran s mnoha obrázky např. dopravních značek, vojenských hodností atd.
- budou zapsány prostřednictvím HTML odkazu na PDF souboru přílohy



- příklad:

Příloha č. 4 k vyhlášce č. 78/2013 Sb.

### Vzor průkazu energetické náročnosti budovy

#### PROTOKOL PRŮKAZU

**Účel zpracování průkazu**

<input type="checkbox"/> Nová budova	<input type="checkbox"/> Budova užívaná orgánem veřejné moci
<input type="checkbox"/> Prodej budovy nebo její části	<input type="checkbox"/> Pronájem budovy nebo její části
<input type="checkbox"/> Větší změna dokončené budovy	

#### GRAFICKÉ ZNÁZORNĚNÍ PRŮKAZU

### PRŮKAZ ENERGETICKÉ NÁROČNOSTI BUDOVY

vydaný podle zákona č. 406/2000 Sb., o hospodaření energií, a vyhlášky č. xxx/2012 Sb., o energetické náročnosti budov

<p>Ulice, číslo: .....</p> <p>PSČ, místo: .....</p> <p>Typ budovy: .....</p> <p>Plocha obálky budovy: ..... m<sup>2</sup></p> <p>Objemový faktor tvaru A/V: ..... m<sup>2</sup>/m<sup>3</sup></p> <p>Celková energeticky vztáhná plocha: ..... m<sup>2</sup></p>	
--	--

### ENERGETICKÁ NÁROČNOST BUDOVY

Celková dodaná energie (Energie na vstupu do budovy)	Neobnovitelná primární energie (Vliv provozu budovy na životní prostředí)
Měrné hodnoty kWh/(m <sup>2</sup> ·rok)	
<div style="display: flex; align-items: center;"> <div style="background-color: #4CAF50; color: white; padding: 5px; margin: 5px;">Mimořádně úsporná</div> <div style="margin: 0 10px;">← XXX</div> </div> <div style="display: flex; align-items: center;"> <div style="background-color: #4CAF50; color: white; padding: 5px; margin: 5px;">Velmi úsporná</div> <div style="margin: 0 10px;">← XXX</div> </div> <div style="display: flex; align-items: center;"> <div style="background-color: #4CAF50; color: white; padding: 5px; margin: 5px;">Úsporná</div> <div style="margin: 0 10px;">← XXX</div> </div> <div style="display: flex; align-items: center;"> <div style="background-color: #FFEB3B; color: black; padding: 5px; margin: 5px;">Méně úsporná</div> <div style="margin: 0 10px;">← XXX</div> </div>	<div style="display: flex; align-items: center;"> <div style="border: 1px solid black; padding: 5px; margin: 5px;">Dop. A</div> <div style="margin: 0 10px;">← XXX</div> </div> <div style="display: flex; align-items: center;"> <div style="background-color: black; color: white; padding: 5px; margin: 5px;">XXX B</div> <div style="margin: 0 10px;">← XXX</div> </div> <div style="display: flex; align-items: center;"> <div style="border: 1px solid black; padding: 5px; margin: 5px;">C</div> <div style="margin: 0 10px;">← XXX</div> </div> <div style="display: flex; align-items: center;"> <div style="border: 1px solid black; padding: 5px; margin: 5px;">D</div> <div style="margin: 0 10px;">← XXX</div> </div>

- bude zapsáno takto:

```
<p>Příloha č. 4 k vyhlášce č. 78/2013 Sb.</p>
<p>Vzor průkazu energetické náročnosti budovy</p>
<p><a href="...">Vzor průkazu energetické náročnosti budovy PROTOKOL PRŮKAZU</a></p>
<p>GRAFICKÉ ZNÁZORNĚNÍ PRŮKAZU</p>
<p><a href="...">GRAFICKÉ ZNÁZORNĚNÍ PRŮKAZU</a></p>
```

- Vzorce
  - např. řešení kvadratické rovnice

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

- bude zapsáno v HTML kódu takto

```
<math>
  <mrow>
    <mi>x</mi>
    <mo>=</mo>
    <mfrac>
      <mrow>
        <mo form="prefix">&minus;</mo>
        <mi>b</mi>
        <mo>&PlusMinus;</mo>
        <msqrt>
          <msup>
            <mi>b</mi>
            <mn>2</mn>
          </msup>
          <mo>&minus;</mo>
          <mn>4</mn>
          <mo>&InvisibleTimes;</mo>
          <mi>a</mi>
          <mo>&InvisibleTimes;</mo>
          <mi>c</mi>
        </msqrt>
      </mrow>
      <mrow>
        <mn>2</mn>
        <mo>&InvisibleTimes;</mo>
        <mi>a</mi>
      </mrow>
    </mfrac>
  </mrow>
</math>
```

## 2. tezaurus CzechVoc ve formátu Simple Knowledge Organization System (SKOS)

### 2.2.4 Výstupy procesů verifikace datové báze

Výstupem dílčích procesů verifikace datové bude komunikace

- Verifikátora DB směrem k Dodavateli DB
- Verifikátora DB směrem k Zadavateli.

Výstupem verifikace tedy není jakákoliv úprava datové báze. *Dodavatel DB* je výlučným tvůrcem datové báze, výlučně provádí její tvorbu a také všechny úpravy/opravy na základě komunikace od Verifikátora DB.

Výstupy činnosti Verifikátora DB tedy jsou:

- Podrobné protokoly komunikované *Dodavateli DB*
  - pro každý porovnávaný ročník,
  - s rozdělením na jednotlivé předpisy případně konsolidovaná znění předpisů, je li to v daném procesu relevantní,
  - se stanovením KPI specifikovaných zvláště pro každý proces.



- Sumární protokoly pro *Zadavatele* osvědčující splnění KPI pro každý ročník *Dodavatelem DB*

Konkrétní podoba a obsah protokolů je v péči *Verifikátora DB* a podléhá schválení *Zadavatelem* před zahájením procesů verifikace.

## 2.3 Časová osa

Celková doba procesu tvorby datové báze včetně verifikace datové báze je plánována na 16 měsíců. Do tohoto časového prostoru však nespádají procesy kontroly normalizace obsahu a tezauru CzechVoc.

Kontrola normalizace obsahu a tezauru CzechVoc bude prováděna po dokončení celého procesu tvorby datové báze v časovém rozsahu 2 měsíců.

### 2.3.1 Časová souslednost tvorby a verifikace datové báze

Tvorba datové báze i její verifikace musí probíhat po jednotlivých ročnících sbírek, od nejstarších k nejnovějším. *Verifikátor DB* i *Dodavatel DB* tak musí postupovat ve stejném taktu. Tento takt definuje primárně konsolidace příslušných ročníků, protože musí jít sériově za sebou (jeden ročník za druhým) – na rozdíl od tvorby datové báze vyhlášených znění a jejich verifikace, u kterých na sobě jednotlivé ročníky závislé nejsou. Takt tvorby datové báze a její verifikace bude dán počtem ročníků a dobou určenou provedení těchto činností. Dle hrubých propočtů bude nutné dokončovat týdně 1–4 ročníky. Konkrétní doba „taktu“ pro daný ročník bude dána velikostí ročníků – ročníky z padesátých let jsou mnohem menší než ročníky z let devadesátých (obecně je 80% všech vyhlášených právních aktů v ročnících od roku 1990).

Jen nutné počítat s tím, že *Verifikátor DB* bude v jistém časovém zpoždění oproti *Dodavateli DB*. Jelikož kritickou cestou tvorby i verifikace datové báze je tvorba a verifikace konsolidovaných znění, bude klíčové řídit zpoždění na tomto bodě tak, aby chyby hlášené *Verifikátorem DB* nemusely být opravovány v konsolidaci velkého množství ročníků, které má *Dodavatel DB* nachystány nebo rozpracovány. Jako akceptovatelný považujeme časový rozdíl mezi *Dodavatelem DB* a *Verifikátorem DB* 3 ročníky.

Vzhledem k délce tvorby datové báze jednoho ročníku i jeho verifikace je zřejmé, že bude muset být rozpracováno několik ročníků naráz a mnoho činností bude muset být paralelizováno pomocí většího počtu týmů.

Také je zřejmé, že činnost *Dodavatele DB* a *Verifikátora DB* musí být detailně koordinována, protože zpoždění jednoho implikuje zpoždění druhého a tím zpoždění celého projektu.

## 2.4 Akceptační proces tvorby datové báze



Akceptační proces pro tvorbu datové báze i její verifikaci je jeden společný. Jeho výsledkem budou akceptační protokoly pro *Dodavatele DB* a akceptační protokoly pro *Verifikátora DB*. Jak je datová báze tvořena a verifikována po ročnících, tak bude i akceptována po ročnících. Výjimkou z tohoto pravidla je kontrola normalizace DB a tezauru CzechVoc, která bude probíhat až pro ukončení tvorby datové báze.

V rámci každého ročníku obdrží *Zadavatel*:

- Veškeré vstupy *Dodavatele DB* pro *Verifikátora DB* ve stavu, kdy o nich *Verifikátor DB* prohlásil, že neobsahují chyby:
  - HTML vyhlášených znění včetně tabulek, obrázků, vzorců, souborových příloh či odkazů,
  - HTML konsolidovaných znění včetně tabulek, obrázků, vzorců, souborových příloh či odkazů,
  - Protokol o kontrole překlepů,
  - Protokoly o provedení konsolidace.
- Výstupy *Verifikátora DB*, zejména:
  - Protokol o komparaci vůči nezávislému zdroji,
  - Protokoly o (ne)provedení konsolidace.

*Zadavatel* následně provede manuální kontrolu 1% z celkového objemu dat. Z počtu chyb nalezených *Zadavatelem* v daném ročníku se budou počítat sankce *Dodavatele DB* i *Verifikátora DB*.

Akceptační proces v rámci jedné akceptace (typicky jednoho ročníku) bude probíhat následovně:

- *Dodavatel DB* i *Verifikátor DB* předají podklady pro danou akceptaci (typicky jeden ročník).
- *Zadavatel* provede manuální kontrolu a oznámí *Dodavateli DB* i *Verifikátorovi DB* nalezené chyby (z tohoto počtu se počítají sankce).
- *Dodavatel DB* provede opravu nalezených chyb, *Verifikátor DB* ověří jejich zapracování.
- *Zadavatel* provede kontrolu oprav.

#### Kroky akceptačního procesu jednoho ročníku:

- Kontrola vyhlášených znění
  - Data:
    - HTML vyhlášených znění



- Protokol o kontrole překlepů.
- Způsob: Manuální kontrola vůči originálům či nezávislému datovému zdroji, manuální kontrola protokolu.
- Poznámka: Pro kontrolu je nutné vybrat kontrolované objekty z rozsahu, který kontroloval verifikátor.
- Typy chyb:
  - Chybný znak
  - Chybný odstavec
  - Nenalezená chyba originálu (tzn. nechtěná změna textu)
  - Chybějící strukturální prvek (z kontroly úplnosti)
  - Chyba v rozložení tabulky
  - Chyba indexu buněk tabulky
  - Chyba znaku nebo numerické hodnoty v tabulce
  - Chybějící nebo nedostatečný protokol o kontrole překlepů
  - Chyba v metadatech předpisu
  - Chyba indexace předpisu.
- Kontrola konsolidovaných znění
  - Data:
    - Protokoly o provedení konsolidace
    - Protokol o komparaci vůči nezávislému zdroji
    - Protokoly o (ne)provedení konsolidace
    - HTML konsolidovaných znění
  - Způsob: Manuální kontrola protokolů, Manuální kontrola odkazů
  - Typy chyb:
    - Chyba v zapracování novelizačního bodu
    - Chybějící nebo nedostatečný protokol o komparaci vůči nezávislému zdroji
    - Chybějící nebo nedostatečný protokol o (ne)provedení konsolidace
    - Chybějící nebo chybně směřovaný odkaz

#### Kroky akceptačního procesu kontroly normalizace obsahu:

- Data:
  - Protokoly o normalizaci základních předpisů
  - Protokoly doporučení z normalizace základních předpisů
- Způsob: Manuální kontrola protokolů
- Typy chyb:
  - Chybějící nebo nedostatečný protokol o normalizaci jednoho základního předpisu
  - Chybějící nebo nedostatečný protokol doporučení z normalizace jednoho základního předpisu



### Kroky akceptačního procesu kontroly tezauru CzechVoc:

- Data:
  - Pojmy a asociace v rámci CzechVoc (SW s daty naplněnými dodavatelem)
- Způsob: Manuální kontrola vazeb nad pojmy ve slovníku CzechVoc pomocí SW nástrojů slovníku CzechVoc (již vyvinutého dodavatelem).
- Poznámka: Pro kontrolu je nutné vybrat kontrolované objekty z rozsahu, který kontroloval verifikátor.
- Typy chyb:
  - Chybná extrakce definice pojmu z ustanovení
  - Chybná vazba v tezauru CzechVoc
  - Chybný typ vazby v tezauru CzechVoc

## 3 Verifikace datové báze

### 3.1 Verifikace datové báze vyhlášených znění

Verifikace datové báze vyhlášených znění sestává z šesti následujících dílčích a vzájemně nezávislých procesů:

1. Kontrola rekonstrukce textů
2. Kontrola úplnosti obsahu vyhlášených znění
3. Kontrola správnosti tabulek
4. Kontrola správnosti netextových entit
5. Nalezení překlepů
6. Kontrola indexace předpisů

Tyto procesy mají stejné vstupy, předpoklady i společné cíle, které jsou popsány v kapitole 2.2.

#### 3.1.1 Kontrola rekonstrukce textů

Kontrola rekonstrukce textů je prvním stupněm verifikace datové báze vyhlášených znění - bez tabulek, vzorců, obrázků a jejich popisků, které budou předmětem kontrol v dalších procesech.

##### 3.1.1.1 Kvantifikace kontroly rekonstrukce textů

*Verifikátor DB* zvolí ¼ celkového stranového rozsahu sbírek pro kontrolu rekonstrukce textů a to vždy tak, aby:

- do výběru byl zařazen vždy kompletní text vyhlášeného znění předpisu; tedy ne jednotlivé strany napříč různými předpisy
- bylo zajištěno poměrné zastoupení všech typů předpisů

Celkový rámcový rozsah stran vybraných předpisů: 100.000. Vyjádřeno v normostranách, 200.000 – 250.000 normostran.

Rámcové rozložení v čase bude následující:

- 10.000 stran Sbírky zákonů z ročníků 1945-1989
- 20.000 stran Sbírky zákonů z ročníků 1990-1999
- 45.000 stran Sbírky zákonů z ročníků 2000-2019
- 25.000 stran Sbírky mezinárodních smluv 2000-2019

Přesnější metodika pro výběr předpisů v rámci jednotlivých ročníků jednotlivých sbírek bude stanovena později.

### 3.1.1.2 Popis procesu kontroly rekonstrukce textů

Verifikátor DB provede kontrolu rekonstrukce textů ve výše uvedeném rozsahu a rozložení v následujících krocích:

- OCR digitalizace, resp. vytěžení textů z PDF (je-li to možné)

Výsledkem bude hladký text bez tabulek, obrázků, vzorců. Tyto podklady budou srovnávány komparačními nástroji s texty rekonstruovanými v rámci tvorby datové báze a jsou známy pouze komparační nástroje pro srovnávání hladkých textů.

Každé ustanovení, nadpis atd. bude oddělen „tvrdým enterem“, přechází-li odstavec na další stránku, tak bude spojen do jediného odstavce.

- Porovnání textů vytvořených *Dodavatelem DB* vzniklých z tvorby datové báze, prostřednictvím nástrojů na textové porovnání textových souborů. Nalezené rozdíly budou konfrontovány s listinnou sbírkou případně s jejími *digitálními stejnopisy*.

### 3.1.1.3 KPI kvality textů rekonstruovaných *Dodavatelem DB*

V rámci porovnání textů budou vyhodnocovány následující tři typy rozdílů (odlišností od původního textu předpisů) vzniklé činností *Dodavatele DB* při rekonstrukci textu:

1. chyba textové správnosti předpisů
  - rekonstrukce textů zavlekla do rekonstruovaného a původně bezchybného textu novou chybu (např. v případě záměny  $I \leftrightarrow 1$ ,  $O \leftrightarrow 0$ ,  $r \leftrightarrow ř$  apod.)
2. chyba fragmentace do odstavců, kde
  - původní jeden odstavec je rozdělen na více odstavců (např. v případě přechodu ze sloupce do sloupce nebo přechodu na novou stránku)
  - původních více odstavců je spojeno do jednoho odstavce
3. chyba originálu, kde text původně chybný (zapříčiněný např. písarskou chybou) byl v rámci rekonstrukce opraven na bezchybný
  - v rámci zachování shodnosti rekonstruovaného textu s originálem je třeba původní chyby zachovávat
  - jako vada je tedy považována „oprava“ chyby; naopak její neopravení je žádoucí (tzn. shoda s originálem)

Sledované hodnoty a z nich vyplývající aktivity.

#### Textová správnost předpisů

Bude vyhodnocován počet chybných znaků na ročník.

- **OK:** 0% chybných znaků
- **OK s výhradou:** 0 - 0,05% chybných znaků
  - *Dodavatel DB* opraví nalezené chyby a znovu vygeneruje předpisy do sdíleného úložiště.



- *Dodavatel DB* zajistí, aby se opakované chyby v ostatních předpisech a ostatních ročnících nevyskytovaly.
- *Verifikátor DB* provede kontrolu opravy zapracovaných chyb.
- **NOK:** > 0,05%:
  - *Dodavatel DB* znovu prokazatelně provede rekonstrukci celého ročníku a znovu vloží všechny nově rekonstruované předpisy ročníku do sdíleného úložiště.
  - *Verifikátor DB* provede kontrolu opravy nalezených chyb nad celým ročníkem.

### Správná fragmentace do odstavců

Budou vyhodnocovány spojené odstavce, nebo naopak odstavce rozdělené na nesprávných místech.

- **OK:** 100% odstavců správně
- **OK s výhradou:** .0 – 0,2% chybně fragmentovaných odstavců
  - *Dodavatel DB* tvorby opraví nalezené chyby a znovu vygeneruje předpisy do sdíleného úložiště.
  - *Verifikátor DB* provede kontrolu opravy zapracovaných chyb.
- **NOK:** > 0,2% chybně fragmentovaných odstavců
  - *Dodavatel DB* znovu prokazatelně provede fragmentaci odstavců celého ročníku a znovu vygeneruje všechny předpisy ročníku do sdíleného úložiště.
  - *Verifikátor DB* provede kontrolu opravy nalezených chyb nad celým ročníkem.

### Chyby originálů

Budou vyhodnocovány nezachycené chyby originálů.

- **OK:** 100% chyb originálů bylo nalezeno
- **OK s výhradou:** .0 - 1% chyb originálů nenalezeno
  - *Dodavatel DB* tvorby opraví nalezené chyby a znovu vygeneruje předpisy do sdíleného úložiště.
  - *Verifikátor DB* provede kontrolu opravy zapracovaných chyb.
- **NOK:** > 1% chyb originálů nenalezeno
  - *Dodavatel DB* znovu prokazatelně provede rekonstrukci celého ročníku a znovu vygeneruje všechny předpisy ročníku do sdíleného úložiště.
  - *Verifikátor DB* provede kontrolu opravy nalezených chyb nad celým ročníkem.

Ve všech předchozích bodech *Verifikátor DB* provádí opakované kontroly nalezených chyb. V případě, že nejsou odstraněny výhrady pro stav **OK s výhradou** resp. opakovaně nastane stav **NOK**, informuje *Verifikátor DB* *Zadavatele* a *Dodavatel DB* může být na základě uvážení *Zadavatele* sankcionován.



### 3.1.2 Kontrola úplnosti obsahu vyhlášených znění

Druhým stupněm verifikace je ověření kompletnosti předpisů. Je třeba ověřit, zda rekonstrukce obsahu vyhlášených znění *Dodavatelem DB* bylo provedeno úplně.

#### 3.1.2.1 Kvantifikace kontroly úplnosti obsahu vyhlášených znění

*Verifikátor DB* v tomto procesu zkontroluje všechna vyhlášená znění předpisů v odhadovaném počtu 38.000 napříč všemi ročníky všech sbírek.

Rámcově se jedná o 400.000 stran vyhlášených znění.

#### 3.1.2.2 Popis procesu kontroly úplnosti obsahu vyhlášených znění

Každý předpis dodaný ve vstupním formátu popsaném v kapitole 2.2.3 jako součást celého ročníku bude porovnán s listinným originálem nebo *digitálním stejnopisem*.

Porovnávání proběhne opticky předpis po předpisu a kontrolována bude:

- Úplnost obsahu každé strany.
- Nejedná se přitom o detailní korekturní čtení, nýbrž o zjištění, zda rekonstruovaný text předpisu obsahuje všechny strukturální entity. Zejména části, hlavy, díly, oddíly, paragrafy, články, odstavce, body, přílohy, poznámky pod čarou, tabulky, obrázky resp. ostatní strukturální entity předpisu.
- Dalším cílem je optická kontrola, zda při rekonstrukci jednotlivých stran nebylo něco vynecháno, např. poslední řádek na stránce, neúplně rekonstruovaná poznámka pod čarou atd.

Optická kontrola může být nahrazena nebo doplněna strojovou kontrolou konzistence; např. zjištěním, zda předpis obsahuje úplnou číselnou řadu paragrafů apod. Použití takové strojové kontroly konzistence je na posouzení *Verifikátora DB*.

#### 3.1.2.3 KPI posouzení kompletnosti obsahu vyhlášených znění

Sledované hodnoty a z nich vyplývající aktivity.

##### Kompletnost předpisu

Bude vyhodnocována kompletnost výše zmíněných struktur.

- **OK:** 0% chybějících strukturálních prvků
- **NOK:** nalezena jakákoliv chyba
  - *Dodavatel DB* provede znovu strukturování předpisu.
  - *Dodavatel DB* zajistí, aby se opakované chyby ve zbylých předpisech nevyskytovaly.
  - *Verifikátor DB* provede kontrolu zapracování oprav všech nalezených chyb.



V případě přetrvávajícího **NOK** stavu u předpisu může být *Dodavatel DB* sankcionován ze strany *Zadavatele*.

### 3.1.3 Kontrola správnosti tabulek

Třetím stupněm verifikace datové báze vyhlášených znění je kontrola obsahové a formální správnosti tabulek.

#### 3.1.3.1 Kvantifikace kontroly správnosti tabulek

Všechny sbírky obsahují přibližně 10.000 tabulek, v nichž je přibližně 10.000.000 samostatných buněk.

*Verifikátor DB* zvolí ¼ celkového rozsahu tabulek podle počtu jejich buněk a provede uvedenou optickou kontrolu dle popisu procesu níže.

Metodika principu výběru ¼ tabulek ke kontrole by měla rovnoměrně pokrývat tabulky předpisů napříč celým ročníkem.

#### 3.1.3.2 Popis procesu kontroly správnosti tabulek

Každý předpis je k dispozici ve vstupním formátu popsaném v kapitole 2.2.3 jako součást celého ročníku.

*Verifikátor DB* extrahuje z dodaných vstupů jednotlivé tabulky a provede porovnání s listinným originálem nebo s jeho *digitálním stejnopisem*.

Porovnávání bude zahrnovat kontrolu:

- správnosti rozložení tabulky, tedy počet sloupců, řádků, sloučení buněk, záhlaví atd.,
- správnosti indexace každé buňky v tabulce,
- správnosti textových a numerických hodnot v buňkách tabulek.

Porovnávání bude probíhat opticky. Strojová kontrola zde není možná vzhledem k tomu, že porovnávaný zdroj je k dispozici pouze v listinné podobě nebo jako *digitální stejnopis*. Částečnou strojovou kontrolu může *Verifikátor DB* podle uvážení připravit a využít pro kontrolu správnosti indexace buněk.

#### 3.1.3.3 KPI kontroly správnosti tabulek

Sledované hodnoty a z nich vyplývající aktivity.

##### Správnost rozložení tabulky

Bude vyhodnocován počet chyb na tabulku.

- **OK:** 0 chyb
- **NOK:** jakákoliv chyba v rozložení



- *Dodavatel DB* znovu prokazatelně provede rekonstrukci tabulky a znovu vygeneruje předpis s opravenou tabulkou.
- *Dodavatel DB* přijme opatření, aby se opakovatelné chyby neopakovaly.
- *Verifikátor DB* provede kontrolu opravy nalezených chyb.

#### Správná indexace buněk v tabulce

Bude vyhodnocováno procento chybných indexů buněk.

- **OK:** 100% indexů správně
- **OK s výhradou:** .0 - 10% chybně indexovaných buněk
  - *Dodavatel DB* tvorby opraví nalezených chyby buněk a znovu vygeneruje předpis s opravenou tabulkou.
  - *Verifikátor DB* provede kontrolu opravy nalezených chyb.
- **NOK:** > 10% chybně indexovaných buněk
  - *Dodavatel DB* znovu provede indexaci buněk celé tabulky a znovu vygeneruje předpis s opravenou tabulkou.
  - *Verifikátor DB* provede kontrolu opravy nalezených chyb.

#### Správné textové a numerické hodnoty v buňkách tabulek

Bude vyhodnocována znaková správnost.

- **OK:** 0% chyb
- **OK s výhradou:** .0 – 0,5% znakových chyb
  - *Dodavatel DB* tvorby opraví nalezené chyby a znovu vygeneruje předpis s opravenou tabulkou.
  - *Verifikátor DB* provede kontrolu opravy nalezených chyb.
- **NOK:** > 0,5% znakových chyb
  - *Dodavatel DB* znovu prokazatelně vytvoří celou tabulku a znovu vygeneruje předpis s opravenou tabulkou.
  - *Verifikátor DB* provede kontrolu opravy nalezených chyb.

Ve všech předchozích bodech *Verifikátor DB* provádí opakované kontroly nalezených chyb. V případě, že nejsou odstraněny výhrady pro stav **OK s výhradou** resp. opakovaně nastane stav **NOK** informuje *Verifikátor DB* *Zadavatele* a *Dodavatel DB* může být na základě uvážení *Zadavatele* sankcionován.

### 3.1.4 Kontrola správnosti netextových entit

Čtvrtým stupněm verifikace je kontrola správnosti netextových entit. Je třeba zkontrolovat správnost a čitelnost všech obrázků, vzorců a souborových příloh, kterých rekonstrukce do textu nebude prováděna a zůstanou v binární podobě v připojených souborech.

### 3.1.4.1 Kvantifikace kontroly netextových entit

Všechny sbírky obsahují přibližně 10.000 netextových entit tabulek.

*Verifikátor DB* zkontroluje všechny netextové entity.

### 3.1.4.2 Popis procesu kontroly netextových entit

Každý předpis je k dispozici ve vstupním formátu popsáném v kapitole 2.2.3 jako součást celého ročníku.

*Verifikátor DB* extrahuje z dodaných vstupů jednotlivé netextové entity a provede porovnání s listinným originálem nebo s *digitálním stejnopisem*.

Porovnávání bude zahrnovat verifikaci:

- shodnosti netextové entity s originálem,
- správného způsobu zachycení netextové entity (např. počet stran PDF příloh, kvalita obrazového souboru).

Směrodatná je ve všech případech shodnost vizuální reprezentace vzorce s listinným originálem resp. *digitálním stejnopisem*.

Porovnávání bude probíhat opticky. Strojová kontrola zde není možná vzhledem k tomu, že porovnávaný zdroj je pouze v listinné podobě resp. jako *digitální stejnopis*.

### 3.1.4.3 KPI kontroly správnosti netextových entit

Sledované hodnoty a z nich vyplývající aktivity.

#### Správnost, kompletnost a čitelnost netextové entity

- **OK:** entita kompletní, správná a čitelná
- **NOK:** jakákoliv chyba v kompletnosti, čitelnosti a správnosti
  - *Dodavatel DB* opraví netextovou entitu a poskytne znovu její souborovou formu.
  - *Verifikátor DB* provede kontrolu opravy všech chybných netextových entit v daném ročníku, které označil stavem **NOK**.

V případě neúplné opravy stavu **NOK** zjištěné *Verifikátorem DB* při kontrole opravy, informuje *Verifikátor DB* *Zadavatele* a *Dodavatel DB* může být sankcionován způsobem, který určí *Zadavatel*.

### 3.1.5 Nalezení překlepů

Principem pátého stupně verifikace je nalezení chybných slov s rozlišením, zda se jedná o chyby

- nesprávné rekonstrukce textů ze strany *Dodavatele DB*
- chyby v originálech předpisů

Tento proces je omezen na Sbírku zákonů. V ostatních sbírkách nebude vykonáván.

### 3.1.5.1 Kvantifikace procesu nalezení překlepů

Ve Sbírce zákonů se vyskytuje přibližně 80.000.000 slov v 500.000 slovních tvarech.

Za slovní tvar je považována jakákoliv skupina znaků oddělených mezerami, interpunkčními znaménky, konci řádků, kde počet znaků je roven nejméně jedné.

### 3.1.5.2 Popis procesu nalezení překlepů

Prostřednictvím aplikace (fungující na principech fulltextové indexace, kterou lze buď vyvinout nebo získat formou knihovny nebo balíkového software) bude kompletní textový obsah (opět po celých ročnících sbírek rozložených na soubory jednotlivých předpisů ve vstupním formátu popsaném v kapitole 2.2.3) rozložen na jednotlivá slova se zjištěním četnosti slov a přesné pozice ve zdrojových souborech (tzv. dump slov).

Následná analýza takového seznamu bude probíhat na základě těchto předpokladů:

- Některá slova budou mít velmi malý počet výskytů a jsou tedy potenciálně chybná.
- Jiná slova budou evidentně potenciálně chybná (v nesouladu s pravidly pravopisu, překlepem apod.)
- Každé podezřelé slovo s malým počtem výskytů, případně s jinou potenciální chybou, je třeba konfrontovat s listinným originálem nebo *digitálním stejnopisem* a zjistit, zda je shodné nebo odlišné.
- Pokud chybné slovo není shodné s originálem, jedná se o chybu tvorby dat.
- Pokud chybné slovo je shodné s originálem, jedná se o chybu originálu, kterou je třeba zaevidovat.

Porovnávání potenciálně chybných slov bude probíhat opticky. Strojovou kontrolu (např. spellcheck) lze sice použít, ale pouze pomocným způsobem, vzhledem k tomu, že Sbírka zákonů obsahuje velký počet cizích slov (např. názvů léků, chemických sloučenin, názvů různých entit apod.).

### 3.1.5.3 KPI kontroly nalezení překlepů

Sledované hodnoty a z nich vyplývající aktivity

#### Správnost potenciálně chybného slova

- **OK:** slovo je shodné s listinným originálem
- **NOK:** slovo není shodné s listinným originálem
  - *Dodavatel DB* provede opravu všech slov ve stavu **NOK** v rekonstruovaném obsahu celého ročníku.
  - *Verifikátor DB* provede kontrolu zapracování všech slov ve stavu **NOK**.

V případě neúplné opravy stavu **NOK** zjištěné *Verifikátorem DB*, informuje *Verifikátor DB* *Zadavatele* a *Dodavatel DB* může být sankcionován způsobem, který určí *Zadavatel*.

### 3.1.6 Kontrola indexace předpisů

Kontrola indexace je šestým stupněm verifikace datové báze vyhlášených znění. Indexací je myšleno správné přiřazení metadat předpisu *Dodavatelem DB*. Jedná se o

„Tvrdá“, tedy jednoznačná metadata

- příslušnost k části
- číslo předpisu
- název předpisu
- datумы schválení a vyhlášení
- datum začátku platnosti
- datum začátku účinnosti
- autor předpisu

„Měkká“ metadata přiřazená právní analýzou

- tematická indexace podle CzechVoc (zařazení předpisu do tematických právních oblastí, nikoliv teaurus pojmů v předpisu)
- územní platnost předpisu
- výjimky z účinnosti předpisu

#### 3.1.6.1 Kvantifikace kontroly indexace předpisů

*Verifikátor DB* v tomto procesu zkontroluje všechny předpisy v odhadovaném počtu 38.000 napříč všemi ročníky všech sbírek.

#### 3.1.6.2 Popis procesu kontroly indexace předpisů

Metadata jsou označována v sekci <head>...</head> HTML souborů každého předpisu dodaných po celých ročnících sbírek rozložených na soubory jednotlivých předpisů ve vstupním formátu popsáném v kapitole 2.2.3.

Kontrola indexace předpisů opět probíhá nad celým ročníkem předpis po předpisu.

„Tvrdá“ metadata každého předpisu budou konfrontována s listinným originálem nebo s *digitálním stejnopisem*.

„Měkká“ metadata budou posuzována právní analýzou. Pro tematické zařazení předpisu, tedy pro přiřazení předpisu k uzlům CzechVoc se předpokládá, že CzechVoc bude vznikat rovněž postupně po ročnících v rámci tvorby datové báze.

Kontroly indexace předpisů budou probíhat optickou kontrolou metadat v sekci <head>...</head> HTML souborů. Strojová kontrola zde není možná vzhledem k tomu, že porovnávání zdroj je pouze v listinné nebo binární PDF podobě.

### 3.1.6.3 KPI kontroly indexace předpisů

Sledované hodnoty a z nich vyplývající aktivity.

#### „Tvrdá“ metadata

- **OK: 0 chyb v ročníku**
    - *Verifikátor DB* akceptuje ročník.
  - **NOK:** jakákoliv chyba v jakékoliv metainformaci jakéhokoliv předpisu
    - *Verifikátor DB* informuje *Dodavatele DB* sumárně pro celý ročník.
    - *Dodavatel DB* znovu prokazatelně provede indexaci předpisu.
    - *Dodavatel DB* přijme opatření, aby se opakovatelné chyby neopakovaly.
    - *Verifikátor DB* provede kontrolu opakované indexace předpisu.
- Cyklus se opakuje až do dosažení stavu **OK: 0 chyb v ročníku**

#### „Měkká“ metadata přiřazena právní analýzou

Bude vyhodnocována stejnost právního názoru *Verifikátora DB* a *Dodavatele DB*.

- **OK: 100% shoda nad celým ročníkem**
- **OK s výhradou:** jakýkoliv odlišný názor na tematickou indexaci předpisu
  - *Dodavatel DB* zváží odlišný názor, a buď opraví indexaci předpisu, nebo zůstane u svého stanoviska, které zdůvodní. Pokud takové zdůvodnění *Verifikátor DB* neuzná, rozhodne *Zadavatel*.  
Rozhodnutí *Zadavatele* je konečné a *Dodavatel DB* i *Verifikátor DB* jej přijmou.  
*Dodavatel DB* provede akci dle rozhodnutí *Zadavatele*.
- **NOK:** jakákoliv evidentní chyba
  - *Dodavatel DB* znovu provede tematickou indexaci předpisu.

**OK s výhradou** a **NOK** jsou postupně vypořádány až do dosažení stavu **OK: 100% shoda nad celým ročníkem**.

V případě neúplné opravy stavu **NOK** zjištěné *Verifikátorem DB*, informuje *Verifikátor DB* *Zadavatele* a *Dodavatel DB* může být sankcionován způsobem, který určí *Zadavatel*.

## 3.2 Verifikace datové báze konsolidovaných znění

Tvorba datové báze konsolidovaných znění představuje postupné zapracování novel resp. redakčních sdělení o opravě chyby do ustanovení vyhlášených předpisů resp. do dříve vytvořených ustanovení konsolidovaných znění.

Verifikace datové báze konsolidovaných znění pak představuje další a samostatnou skupinu procesů verifikace datové báze.

Verifikace datové báze konsolidovaných znění sestává ze čtyřech nezávislých procesů:

1. Ověření protokolů o provedení konsolidace

2. Komparace datové báze vůči nezávislému zdroji
3. Analýza konfliktů v konsolidacích
4. Kontrola odkazového aparátu

Tyto procesy mají doplňkové vstupy specifikované u jednotlivých procesů. V logice verifikace datové báze konsolidovaných znění budou vykonávány sekvenčně v rámci každého ročníku od nejstaršího předpisu po nejmladší.

Jednotlivé procesy jsou popsány v následujících samostatných kapitolách 3.2.1 - 3.2.4.

## 3.2.1 Ověření protokolů o provedení konsolidace

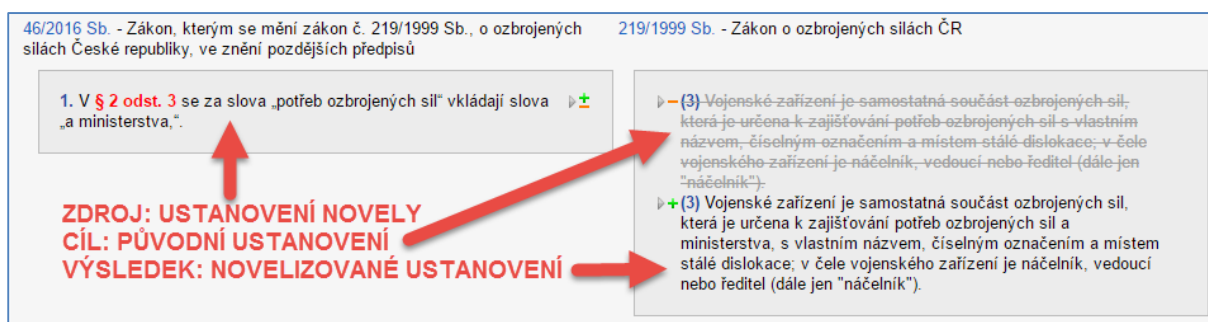
Kontrola provedení konsolidace je prvním stupněm verifikace datové báze konsolidovaných znění.

**Zdrojem** novelizace je novelizační bod popisující změnu, která má být v ustanovení konkrétního předpisu provedena. **Cílem** je ustanovení před novelizací. **Výsledkem** je pak novelizované ustanovení.

### 3.2.1.1 Vstupy pro ověření protokolů o provedení konsolidace

Ze zpracování každého novelizačního bodu bude *Dodavatelem DB* pořízen (případně strojově vygenerován) *Protokol o provedení konsolidace*. Takový protokol bude pořízen i v případech, kdy novelizační bod nebylo možné zpracovat, případně zpracování není jednoznačné.

*Protokol o provedení konsolidace* má formu „novelizační trojice“: **původní ustanovení**  $\leftrightarrow$  **ustanovení novely**  $\leftrightarrow$  **novelizované ustanovení**. Příklad nejjednoduššího možného případu „novelizační trojice“ je uveden na následujícím obrázku.



*Protokol o provedení konsolidace* má formu samostatně čitelného souboru.

*Protokoly o provedení konsolidace* jsou zvláštním doplňkovým vstupem pro tuto fázi verifikace datové báze konsolidovaných znění.

Speciálním případem novelizace jsou oznámení o opravě tiskové chyby, které se do původního ustanovení přenáší ne s účinností novely nýbrž s účinností původního ustanovení.



Nebude-li v možnostech dodavatele tvorby datové báze zapracování některého novelizačního bodu, předá dokumentaci tohoto problému *Verifikátorovi DB* formou specifického *protokolu o (ne)provedení konsolidace* s popisem problému.

### 3.2.1.2 Kvantifikace ověření protokolů o provedení konsolidace

Odhad počtu takových *protokolů o provedení konsolidace* je přibližně 50.000.

Novela v daném ročníku, která způsobí nové konsolidované znění předpisů z minulých ročníků, bude formou protokolů kontrolována v rámci ročníku novely.

Celkový objem je následující

- Sbírka zákonů 1945-2019: 74 ročníků
- Sbírka mezinárodních smluv 2000-2019: 20 ročníků
- Úřední list 1945-1962: 17 ročníků

Celkový počet konsolidovaných znění všech sbírek pro všechny ročníky je odhadován na 32.000.

### 3.2.1.3 Předpoklady ověření protokolů o provedení konsolidace

*Dodavatel DB* bude na sdílené úložiště publikovat *protokoly o provedení konsolidace*.

*Verifikátor DB* ze sdíleného úložiště odebere vždy celý ročník najednou protokolů i konsolidovaných znění, také najednou.

### 3.2.1.4 Popis procesu ověření protokolů o provedení konsolidace

*Verifikátor DB* verifikuje *Protokoly o provedení konsolidace* zapracování novelizačních bodů každé novely v pořadí, jak byly novely v ročníku vyhlášovány.

Každý *Protokol o provedení konsolidace* je třeba důkladně analyzovat s patřičnou mírou ostražitosti k ostatním nekonzistencím a chybám, které se mohou vyskytnout.

Tým, který bude vykonávat tuto činnost, by měl disponovat schopností pro právně analytické posuzování situací zachycených v protokolech.

### 3.2.1.5 KPI ověření protokolů o provedení konsolidace

Sledované hodnoty a z nich vyplývající aktivity.

Správnost provedení konsolidace bude vyhodnocována stavem OK/NOK pro každý *Protokol o provedení konsolidace*.

- **OK:** novelizační bod správně zapracován
- **NOK:** nesprávné zapracování
  - *Dodavatel DB* provede opravu provedení novelizace.



- *Verifikátor DB* provede kontrolu opravy prostřednictvím nového *protokolu o provedení konsolidace*.

V případě neúplné opravy stavu **NOK** zjištěné *Verifikátorem DB*, informuje *Verifikátor DB* *Zadavatele* a *Dodavatel DB* může být sankcionován způsobem, který určí *Zadavatel*.

### 3.2.2 Komparace datové báze vůči nezávislému zdroji

Komparace datové báze vůči nezávislým zdrojům je druhým stupněm verifikace datové báze konsolidovaných znění.

Předchozí stupně kontrol pracují pouze s datovou bází nově vytvořenou *Dodavatelem DB*. Je tedy možné, že chyba v datové bázi může přetrvat. Proto je třeba texty nově vytvořených předpisů ve vyhlášených a konsolidovaných zněních konfrontovat s jiným nezávisle vytvořeným zdrojem.

#### 3.2.2.1 Vstupy komparace datové báze vůči nezávislému zdroji

Vstupem pro komparaci datové báze nad rámce celkových vstupů popsaných v kapitole 2.2.3 bude nezávislý zdroj pro komparaci – tedy vyhlášená a konsolidovaná znění právních předpisů některého z renomovaných právních systému na českém trhu.

#### 3.2.2.2 Předpoklady komparace datové báze vůči nezávislému zdroji

Předpoklady pro tuto fázi verifikace nad rámec celkových předpokladů popsaných v kapitole 2.2.2 jsou:

- Zajištění doplňkového vstupu popsaného v předchozí kapitole 3.2.2.1, tedy nezávislého zdroje pro komparaci.
- Předzpracování (strojové) tohoto zdroje do podoby komparovatelné nástroji pro porovnání textů s předpisy datové báze dodané *Dodavatelem DB*.

#### 3.2.2.3 Kvantifikace komparace datové báze vůči nezávislému zdroji

Komparace bude probíhat pouze nad předpisy Sbírky zákonů. Ostatní sbírky, tedy Sbírka mezinárodních smluv ani Úřední list, nebudou komparovány.

Budou komparovány vždy 3 znění každého předpisu Sbírky zákonů:

- vyhlášené znění,
- aktuální konsolidované znění, resp. poslední konsolidované znění před zrušením předpisu,
- náhodně vybrané minulé konsolidované znění starší než znění z předchozího bodu, existuje-li takové.

Odhad počtu s prognózou do konce roku 2019:

- počet vyhlášených znění pro komparaci: 20.000
- počet předpisů, které mají alespoň jedno konsolidované znění: 10.000
- počet předpisů, které mají více než jedno konsolidované znění (pro náhodný výběr minulého konsolidovaného znění): 7.000
- celkový odhad znění ke komparaci (součet výše uvedených počtů): 37.000

#### 3.2.2.4 Popis procesu komparace datové báze vůči nezávislému zdroji

*Verifikátor DB* otevře v nástroji pro porovnání textu současně

- znění předpisu dodané *Dodavatelem DB*,
- znění předpisu z nezávislého zdroje se stejnou účinností.

*Verifikátor DB* provede komparaci celého znění po jednotlivých ustanoveních. Neukončí tedy komparaci na první chybě.

Mohou nastat tyto situace:

- ustanovení jsou totožná,
- ustanovení nejsou totožná, v takovém případě je chyba buď ve vytvořené datové bázi, nebo v nezávislém zdroji.

V případě rozdílů je třeba analyzovat příčiny s použitím

- listinných originálů nebo *digitálních stejnopisů*
- protokolů o provedení konsolidace z předchozího stupně verifikace (viz kapitola 3.2.1)

Výsledným zjištěním je stav každého rozdílu identifikující zda je chyba v datové bázi vytvořené *Dodavatelem DB* nebo v porovnávaném zdroji.

Chyba v datové bázi vytvořené *Dodavatelem DB* musí být prokazatelným způsobem zdůvodněna a zaznamenána do protokolu o komparaci předpisu.

Orientační odhad počtu rozdílů mezi datovou bázi vytvořenou *Dodavatelem DB* a nezávislým zdrojem na jeden ročník, které bude třeba analyzovat je v intervalu  $> 100$  a  $< 1.000$ .

Poznámka k fyzickým možnostem procesu komparace: Stejná chyba v obou porovnávaných zdrojích nebude komparací odhalena.

#### 3.2.2.5 KPI komparace datové báze vůči nezávislému zdroji

Sledované hodnoty a z nich vyplývající aktivity.

Bezchybnost textu jakéhokoliv znění předpisu vytvořeného *Dodavatelem DB* na základě analýzy rozdílů.

- **OK:** komparace neodhalila žádnou chybu v textu jakéhokoliv znění předpisu
- **NOK:** 1 nebo více chyb v textu jakéhokoliv znění předpisu
  - *Dodavatel DB* provede opravu.

- *Dodavatel DB* opakovaně poskytne na sdílené úložiště všechna znění předpisů, ve kterých byla nalezena chyba.
- *Verifikátor DB* provede opakovanou komparaci znění předpisů za účelem kontroly odstranění chyb.

V případě neúplné opravy stavu **NOK** zjištěné *Verifikátorem DB*, informuje *Verifikátor DB* *Zadavatele* a *Dodavatel DB* může být sankcionován způsobem, který určí *Zadavatel*.

### 3.2.3 Analýza konfliktů v konsolidacích

Třetím stupněm verifikace datové báze konsolidovaných znění je analýza konfliktů v konsolidacích a doporučení jejich řešení (dále označeno jako *Konsolidační konflikt*).

Ověřování protokolů o provedení konsolidace (viz kapitola 3.2.1) i komparace datové báze vůči nezávislým zdrojům (viz kapitola 3.2.2) ukáží na situace, v kterých některé novelizační body není možno zapracovat z důvodů nejasně formulovaných, protichůdných nebo duplicitních novelizačních instrukcí v novelizačních bodech – *Konsolidačních konfliktů*. Takové situace jsou „hazardní“ v tom smyslu, že zavádí potřebu výkladu uživatelem a tím i riziko rozdílných výkladů ustanovení právních předpisů, ke kterým jsou kompetentní pouze příslušné soudy.

V takovém případě je třeba provést právní analýzu důvodů, pro které správné zapracování novelizačních bodů není možné. Výsledkem takové právní analýzy bude

- buď popis *konsolidačního konfliktu*, který bude následně zveřejněn na portálu e-Sbírka,
- nebo doporučení, jak konsolidační konflikt vyřešit, např. v příští novelizaci předpisu.

Analýza *Konsolidačních konfliktů* je v působnosti dodavatele verifikace datové báze a nemá souvislost s dodavatelem její tvorby. KPI tedy nejsou navrhovány.

#### 3.2.3.1 Kvantifikace analýzy konfliktů v konsolidacích

Počet konsolidačních konfliktů lze pouze odhadnout na < 1.000.

#### 3.2.3.2 Vstupy pro analýzu konfliktů v konsolidacích

Specifickými vstupy tohoto procesu jsou zde výsledky předchozích procesů.

- Specifické *Protokoly o (ne)provedení konsolidace*.
- Zjištění z komparace datové báze vůči nezávislému zdroji, které nebudou vyhodnoceny jako **NOK**.

#### 3.2.3.3 Předpoklady analýzy konfliktů v konsolidacích

Specifickými předpoklady pro tuto fázi verifikace jsou:

- schopnost a kvalifikace *Verifikátora DB* k potřebným právním analýzám,
- zapojení právních autorit v daných oblastech v případě potřeby,
- zapojení právních expertů *Zadavatele* v roli arbitra.

#### 3.2.3.4 Cíle analýzy konfliktů v konsolidacích

Cílem analýzy konfliktů v konsolidacích je:

- vyřešení *Konsolidačních konfliktů*, jsou-li řešitelné,
- popis neřešitelných *Konsolidačních konfliktů*, který bude následně publikován v e-Sbírce.

#### 3.2.3.5 Popis a výstupy procesu analýzy konfliktů v konsolidacích

Pro každý *Konsolidační konflikt* bude vypracován právně-analytický dokument, který bude předán *Dodavateli DB* a který

- buď poskytne instrukci k zapracování řešení *konsolidačního konfliktu* v datové bázi,
- nebo poskytne formulaci, která pak bude ve formě komentáře k nezpracovanému *Konsolidačnímu konfliktu* připojena k dotčeným ustanovením,
- případně obojí, tedy instrukce k zapracování doplněná komentářem.

*Dodavatel DB* následně datovou bázi aktualizuje dle právně-analytických dokumentů od *Verifikátora DB*.

### 3.2.4 Kontrola odkazového aparátu

Čtvrtým a posledním stupněm verifikace datové báze konsolidovaných znění je kontrola odkazového aparátu.

V rámci tvorby datové báze bude vytvořena síť odkazů mezi ustanoveními uvnitř jednoho předpisu a mezi různými předpisy. Na portálu e-Sbírka a v nástrojích pro tvorbu legislativního procesu bude odkaz reprezentován hypertextovým linkem.

Tuto síť odkazů je třeba zkontrolovat. Tato kontrola bude probíhat pouze ve Sbírce zákonů.

Některé typy odkazů jsou přitom důležitější než jiné např. mezi nadřazenými a prováděcími předpisy.

#### 3.2.4.1 Kvantifikace kontroly odkazového aparátu

Lze odhadnout:

- počet unikátních odkazů: 1.000.000,
- tyto unikátní odkazy budou namnoženy do všech znění předpisů, přičemž lze předpokládat celkový počet takto namnožených odkazů: 10.000.000.

Rámcové vysvětlení „namnožení unikátních odkazů“:

- odkaz z ustanovení A1 vyhlášeného znění X1 předpisu C míří na ustanovení B1 ve vyhlášeném znění Y1 předpisu Y
- novela N1 způsobí vznik konsolidovaného znění X2 předpisu X, ustanovení A1 novelizováno není a ve znění X2 je tedy v identické podobě označeno jako A2
- unikátní odkaz A1 → B1 ve znění X1 je „namnožen“ v identické podobě do znění X2 jako A2 → B1

### 3.2.4.2 Vstupy pro kontrolu odkazového aparátu

Specifickými vstupy pro kontrolu odkazového aparátu jsou:

- Přesná metodika dodaná *Dodavatelem DB*, podle které jsou odkazy provedeny.
- Soubory všech znění předpisů Sbírky zákonů.  
Pro jednoduchost lze předpokládat, že tyto soubory budou ve formátu identickém k celkovým vstupům popsaným v kapitole 2.2.3 a odkazy v nich budou zapsány v „hyperlinkové notaci“ HTML.

### 3.2.4.3 Popis procesu kontroly odkazového aparátu

Základním způsobem kontroly je optické interaktivní ověření kliknutím na odkaz a kontrola, zda odkaz směřuje na správné místo.

Při takovémto způsobu kontroluje *Verifikátor DB* postupně odkaz po odkazu a postupně buduje KPI pro kontrolovaný ročník.

Je evidentní, že *Verifikátor DB* může po důkladné analýze vybudovat pomocné nástroje pro alespoň poloautomatickou kontrolu odkazového aparátu.

### 3.2.4.4 KPI kontroly odkazového aparátu

Sledované hodnoty a z nich vyplývající aktivity.

#### Správné zacílení odkazu

Bude vyhodnocováno, zda odkaz míří na správné místo citované ve zdroji odkazu.

- **OK:** 100% odkazů míří správně
- **OK s výhradou:** 0 – 3% chyb, tedy odkazů zaměřených nesprávně
  - *Dodavatel DB* tvorby opraví vadné odkazy a znovu vygeneruje předpisy do sdíleného úložiště.
  - *Verifikátor DB* provede kontrolu opravy zpracovaných chyb.
- **NOK:** > 3% odkazů je zaměřeno nesprávně
  - *Dodavatel DB* znovu vytvoří odkazový aparát pro celý ročník.
  - *Verifikátor DB* provede kontrolu opravy dříve nalezených chyb.

V případě přetrvávajícího **NOK** stavu u předpisu může být *Dodavatel DB* sankcionován ze strany *Zadavatele*.

## 3.3 Kontrola normalizace obsahu datové báze

Souhrnný pojem *kontrola normalizace obsahu datové báze* představuje ověření fragmentace, hierarchizace a konzistence datové báze.

Správná normalizace obsahu je nutnou podmínkou pro funkčnost e-Legislativy, konkrétně e-šablony pro plusovou osu, tedy editoru právních předpisů.

Jedná se o specifický proces oddělený od ostatních kontrol.

Tento proces bude vykonán po ukončení tvorby datové báze, tedy po úspěšném průběhu všech dříve pospaných procesů, které se de facto zabývaly z různých úhlů kontrolami správnosti obsahu datové báze.

### 3.3.1 Cíle kontroly normalizace obsahu datové báze

Konkrétní cíle verifikace normalizace datové báze jsou následující:

- Kontrola, zda každý fragment má správnou strukturu (tedy např. paragraf, odstavec, písmeno, bod, nadpis apod.) a zda má každá číslovaná struktura správně přiřazené pořadí (např. § 1 odst. 2, písm. 3).
- Kontrola, zda je správně sestavena stromová struktura předpisu (hierarchie „rodič – děti“).
- Kontrola souladu vytvořené datové báze s pravidly tvorby právních předpisů, zejména s Legislativními pravidly vlády platnými v čase provádění kontroly<sup>1</sup> (tzn., s kterými bude pracovat eLegislativa). Tento typ kontroly je doplňkový a jím nalezené „chyby“ se zaznamenávají do protokolů, nejsou však předmětem hodnocení (nezapočítávají se do KPI)

### 3.3.2 Kvantifikace kontroly normalizace obsahu datové báze

Výše uvedené kontroly budou provedeny nad všemi zněními (tj. znění vyhlášená a znění konsolidovaná) platných předpisů (tedy těch, kterým nebyla explicitním zrušením ukončena účinnost) s omezení pouze na Sbírku zákonů. V ostatních sbírkách nebude kontrola normalizace obsahu prováděna.

Celkový počet znění ke kontrolám lze odhadnout na 35-40.000.

### 3.3.3 Vstupy kontroly normalizace obsahu datové báze

Je evidentní, že kontroly normalizace verifikaci nelze provést nad vstupy souhrnně používanými pro předchozí kontroly, tedy nad soubory znění označované jednoduchým

---

<sup>1</sup> <http://www.vlada.cz/cz/ppov/lrv/dokumenty/legislativni-pravidla-vlady-91209/>





Tag	Value
Xml	
Engine	DodavalDB
Version	N.NNN
Doc	
DocId	sb2011c001z0001
Head	
Frag	
Frag	
FID	f00000007
Struct	Virtual_Novela
Frag	
FID	f00000008
Struct	Clanek
Number	1
Anchor	c1
Quote	Čl. I
Content	Čl. I
Frag	
Frag	
FID	f00000010
Struct	Bod_Dd
Number	1
Anchor	c1-1
Quote	Čl. I bod 1
Content	<var>1.</var> V § 14 odst. 1 se na konci písmena a) slovo „nebo“ zrušuje, na konci písmena b) se tečka nahrazuje čárkou a doplňují se písmena c) a d), která znějí:
Frag	
FID	f00000011
Struct	Pismo_Lb
Number	3
Content	„c) střídavě modrá pole s černým nápisem „POLICIE ČR“ a žlutá pole s černým nápisem „VSTUP ZAKÁZÁN“, nebo
Frag	
FID	f00000012
Struct	Pismo_Lb
Number	4
Anchor	c1-1-d
Quote	Čl. I bod 1 písm. d)
Content	<var>5.</var> střídavě žlutá pole s černým nápisem „POLICIE ČR“ a modrá pole s černým nápisem „VSTUP ZAKÁZÁN“.
Frag	
Frag	
Frag	
FID	f00000021
Struct	Bod_Dd
Number	5
Anchor	c1-5
Quote	Čl. I bod 5
Content	<var>5.</var> V příloze č. 1 bodu 2 písm. b) se na konci doplňují vyobrazení:
Frag	
FID	f00000022
Struct	Priloha
Number	1
Content	„<a class=“linkfile” rel=“nofollow” target=“_blank” title=“” href=“/sb/2011/sb2011c001z001o006.jpg”><img src=“/sb/2011/nahled/2011c001z001o006(N).png” alt=“”/></a>“.
Frag	
Frag	
Frag	
Version	
Version	1
EffectFrom	2011-01-22
Links	
LinkType	
TypeId	8
Name	Mění
LinkItem	
DocId	3912532
Path	/cs/2008-460
Quote	460/2008 Sb.
Title	Vyhlaška o způsobu vnějšího označení, služebních stejnojmenných a zvláštním barevném provedení a označení služebních vozidel, plavidel a letadel Policie České republiky a o pro

Obrázek č. 3: Příklad stromového zápisu fragmentace a hierarchizace znění předpisu v XML

Rámcový popis XML schémy z příkladů na obrázcích uvedených výše:

- <Xml> kontejner
- <Doc> dokument, jeden předpis sbírky
- <Head> veškeré vlastnosti (metadata) předpisu
- <Frag> fragmenty, hierarchie, znění předpisu, struktury, content, citace, kotvy
- <Links> souvislosti, odkazy na jiné předpisy dle typu
- <Versions> verze, přehled všech znění předpisu včetně novel a datuju účinnosti

### 3.3.4 Předpoklady procesu kontroly normalizace obsahu

Verifikátor DB vytvoří nástroj pro kontroly normalizace obsahu Sbírky zákonů, což obnáší kontrolu velké řady situací např.



- **pro fragmentaci:**
  - Je předpis správně logicky rozčleněn? Má správně určenou normativní, novelizační, přílohovou část?
  - Jsou struktury, kompletně, správně, unikátně a konzistentně číslovány?
  - Nevyskytují se společně takové struktury fragmentů, které jsou v rozporu s legislativně technickými pravidly<sup>1</sup> (např. paragrafy a články) současně v jednom předpisu?
  - Nevyskytují se fragmenty, které by měly být strukturovány, avšak nejsou?
  - Nejsou zaměněny struktury nadpisů se strukturami nečíslovaných odstavců?
  - Nejsou novelizovány typy předpisů, které z principu novelizovány nemohou být?
  - ... atd. ...
- **pro hierarchizaci:**
  - má každá část předpisu vše, co má mít a nemá, co mít nemá? – např.
    - novelizační část má aspoň jeden novelizační bod
    - každý zákon má účinnostní ustanovení
    - ústavní zákon má pouze články a nikoliv paragrafy
  - má každá příloha svou identifikaci?
  - je každý fragment v hierarchii zařazen tak, aby byl jednoznačně citovatelný?
  - ... atd. ...

Je předpokládána strojová kontrola s následným manuálním vyhodnocením validity incidentů identifikovaných nástrojem.

Výstupem procesu bude

1. stanovení KPI pro každé znění předpisu – pro nápravu *Dodavatelem DB*
2. dokumentace nesouladu s legislativně technickými pravidly (viz kapitola 3.3.1) – jako informace pro zadavatele

### 3.3.5 Popis procesu kontroly normalizace obsahu

Prostřednictvím nástroje na kontroly normalizace obsahu, jehož rámec požadavků je popsán výše (kapitola 3.3.4) provede *Verifikátor DB* kontrolu každého znění každého předpisu (dle specifikace kvantifikace (kapitola 3.3.2)). Následně manuálně vyhodnotí incidenty identifikované nástrojem tak, aby *Dodavateli DB* neodeslal falešná hlášení chyb.

Výstupy procesu kontroly normalizace obsahu komunikované *Dodavateli DB* jsou opět protokoly analogické k protokolům výstupů ostatních dílčích procesů s tím rozdílem, že základní vyhodnocovanou jednotkou zde je předpis a všechna jeho znění (tedy ne celý ročník jako v předchozích procesech).

Výstupy procesu kontroly normalizace obsahu pro zadavatele budou navíc obsahovat protokoly dokumentujícími každý nesoulad s legislativně technickými pravidly. Takto vzniklá dokumentace bude sloužit jako podklad k nápravě takového při následující legislativní

úpravě (novelizaci) dotčeného předpisu. Tvůrce legislativní úpravy tento podklad zohlední a podle svého uvážení provede nápravu.

Pro osvětlení celého procesu uvádíme i příklad nad zákonem č. 191/1950 Sb. (Zákon směnečný a šekový):

- Každý článek zákona znovu začíná § 1, takže ve výsledku jsou tam 3ks § 1-7, 60ks § 8-68.
- Strojová kontrola strukturování to vyhodnotí jako jednu chybu (tak to musí být uděláno).
- Strojová kontrola hierarchizace identifikuje další chybu: § nesmí být podřazen článku.
- Operátor *Verifikátor DB* se na obojí podívá, otevře si PDF a zkonstatuje „není to chyba“ a nenastavuje KPI (nehlásí chybu *Dodavatel DB*).
- V protokolu s doporučeními zadavateli bude uvedeno něco jako „duplicitní paragrafy + § v Čl.“, zvažte v budoucnu opravu“.
- *Zadavatel* nazná, že je to nemožné a odloží tento protokol do archivu.

### 3.3.6 KPI procesu kontroly normalizace obsahu (pro *Dodavatele DB*)

Správnost normalizace datové báze bude vyhodnocována počtem kritických a nekritických chyb strukturování každého předpisu, tj. všech jeho znění, kde:

- kritickou chybou je chyba ve fragmentaci a hierarchizaci normativních částí předpisů s výjimkou příloh (resp. vložených dokumentů typu mezinárodních smluv)
- nekritickou chybou je chyba ve fragmentaci a hierarchizaci příloh

Sledované hodnoty a z nich vyplývající aktivity:

Kritické chyby

- **OK:** žádná kritická chyba
- **NOK:** 1 a více kritických chyb
  - *Dodavatel DB* provede novou fragmentaci a hierarchizaci všech znění dotčeného předpisu.
  - *Verifikátor DB* provede kontrolu opravy zpracovaných chyb.

Nekritické chyby

- **OK:** žádná nekritická chyba
- **OK s výhradou:** < 5% fragmentů chybně strukturováno nebo hierarchizováno
  - *Dodavatel DB* odstraní nalezené nekritické chyby.
  - *Verifikátor DB* provede kontrolu opravy nalezených nekritických chyb.
- **NOK:** 5 a více % fragmentů chybně strukturováno nebo hierarchizováno
  - *Dodavatel DB* provede nové strukturování a hierarchizaci všech znění dotčeného předpisu.
  - *Verifikátor DB* provede opakovanou kontrolu všech znění dotčeného předpisu.



V případě neúplné opravy stavů **NOK** zjištěných *Verifikátorem DB*, informuje *Verifikátor DB* *Zadavatele* a *Dodavatel DB* může být sankcionován způsobem, který určí *Zadavatel*.

## 3.4 Kontrola tezauru CzechVoc

Tezaurus CzechVOC je digitálním terminologickým výkladovým slovníkem pojmů užívaných v právních předpisech. Asistuje adresátům práva při orientaci a vyhledávání v e-Sbírce a pomáhá ke správnému porozumění pojmům užívaným ve Sbírce zákonů a mezinárodních smluv. Tvůrcům legislativy pak napomáhá ke správnému a jednoznačnému používání pojmů ve správných kontextech jako součást e-Šablony.

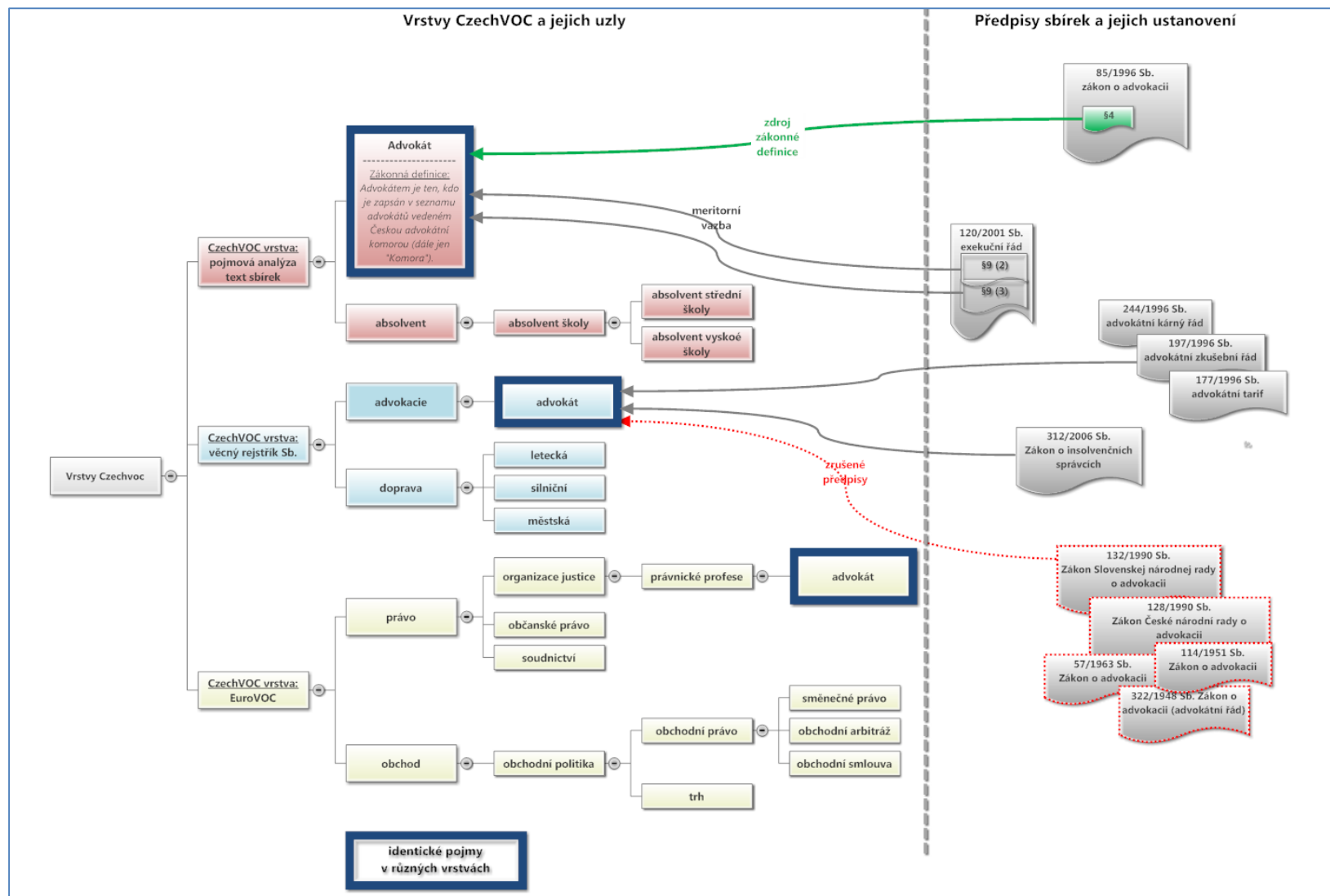
CzechVoc vznikne autorskou činností Dodavatele DB. Bude CzechVoc bude organizován do vrstev organizované do tezauru dle normy ISO 25964<sup>2</sup>.

1. vrstva bude obsahovat pojmy používané ve sbírkách asociované na fragmenty, které pojmy buď definují (definiční vazba), nebo jsou pro význam pojmů jinak podstatné (meritorní vazba)
  - předpokládáno využití zejména v systému e-Legislativa při tvorbě/novelizaci právních předpisů jako asistenční nástroj pro tvůrce legislativy
2. vrstva bude obsahovat témata právních předpisů asociované na relevantní předpisy (v historii Sbírky zákonů je používán také pojem *věcný rejstřík*)
  - předpokládáno využití zejména v systému e-Sbírka jako asistenční nástroj pro dohledání právních norem
3. vrstva bude obsahovat zrcadlo terminologického tezauru EuroVoc<sup>3</sup>
  - předpokládáno využití zejména v uživatelských rozhraních různých aspektů souvislosti předpisů Sbírky zákonů a mezinárodních smluv s právními předpisy EU

Struktura a vzájemné souvislosti zmíněných vrstev CzechVoc jsou ilustrovány na následujícím Obrázek č. 4:

<sup>2</sup> [https://en.wikipedia.org/wiki/ISO\\_25964](https://en.wikipedia.org/wiki/ISO_25964)

<sup>3</sup> <http://eurovoc.europa.eu/drupal/?q=cs>



Obrázek č. 4: Rámcová struktura CzechVoc

### 3.4.1 Cíle kontroly tezauru CzechVoc

Konkrétním cílem kontroly tezauru CzechVoc v rámci tohoto procesu je kontrola pojmové báze, tedy 1. vrstvy CzechVoc, její organizace do tezauru a sít' asociací na fragmenty, tedy na konkrétní ustanovení konkrétních předpisů.

Kontrola 2. vrstvy, tedy vlastního tematického zařazení předpisů, je předmětem samostatné kontroly popsané v kapitole 3.1.6. kde probíhá po jednotlivých ročnících, v rámci kterých je tato vrstva (tedy stromová struktura témat) postupně budována. Proces dle této kapitoly (**Error! Reference source not found.**) naopak probíhá po kompletaci tvorby a verifikace d datové báze jako celku a bude v něm provedena právní analýza celkového sestavení této vrstvy CzechVoc.

Ukáží-li výše uvedené kontroly na nekonzistence, budou tyto předány zadavateli ve formě zdůvodněných doporučení resp. návrhů ke změnám.

Kontrola 3. vrstvy (EuroVoc) nebude prováděna, jak již řečeno výše jedná se o zrcadlo datových struktur EuroVoc.

### 3.4.2 Kvantifikace kontroly tezauru CzechVoc

V 1. vrstvě je předpokládáno orientačně 10.000 pojmů. Velmi zhruba odhadovaný počet asociačních vazeb na fragmenty sbírek je mezi 250.000 a 500.000. Pro kontrolu bude vybráno 10% pojmů, tedy orientačně 1.000 pojmů a s nimi všechny jejich definiční i meritorní vazby.

V 2. vrstvě je předpokládáno orientačně 500-1000 pojmů.

### 3.4.3 Zvláštní vstupy kontroly tezauru CzechVoc

*Dodavatel DB* předá *Verifikátorovi DB* 1. a 2. vrstvu CzechVoc po jeho dokončení, tedy po zpracování všech předpisů sbírek ve formátu standardizovaném a využívaném pro zápis tezaurů, např. SKOS<sup>4</sup>.

Dále *Dodavatel DB* předá pro kontrolu 1. vrstvy CzechVoc *Verifikátorovi DB* asociační vazby mezi fragmenty a pojmy CzechVoc tak, aby z něj byla patrná

- typ vazby (definiční/meritorní)
- jednoznačná identifikace fragmentu (např. FID podle Obrázek č. 3, v případě použití takového XML schématu).

---

<sup>4</sup> [https://en.wikipedia.org/wiki/Simple\\_Knowledge\\_Organization\\_System](https://en.wikipedia.org/wiki/Simple_Knowledge_Organization_System)

### 3.4.4 Zvláštní předpoklady procesu kontroly tezauru CzechVoc

*Verifikátor DB* vytvoří nástroj pro interaktivní kontroly konzistence sítě asociačních vazeb.

Interaktivní kontrola konzistence je primárně strojová dle algoritmických pravidel (zejména výskytu pojmu v asociovaném fragmentu), kde asociační vazba, která nenaplnuje algoritmická pravidla, bude posouzena interaktivně *Verifikátorem DB*.

Výstupem nástroje bude

1. stanovení KPI pro každý pojem první vrstvy – pro nápravu *Dodavatelem DB*
2. souhrnná zpráva pro zadavatele obsahující zdůvodněná doporučení resp. návrhů ke změnám v organizaci tezauru (zvlášť pro 1. vrstvu a pro 2. vrstvu) – tato zpráva je vstupem pro *Zadavatele* a nehodnotí přímo práci *Dodavatele DB*.

### 3.4.5 Popis procesu kontroly tezauru CzechVoc

Kontrola tezauru CzechVoc bude probíhat po ukončení tvorby datové báze současně s předchozí kontrolou normalizace obsahu a se stejnou dvouměsíční lhůtou jako kontrola normalizace obsahu.

Prostřednictvím nástroje pro interaktivní kontroly správnosti konzistence sítě asociačních vazeb provede *Verifikátor DB* nad vzorkem 10% pojmů 1. vrstvy kontrolu

- zákonné definice pojmu, zda je správně extrahovaná z ustanovení
- přiřazení fragmentů k pojmům
- kontrolu typu vazby (definiční/meritorní)

s výsledným stanovením KPI pro každý pojem a následné komunikace k *Dodavateli DB*.

Paralelně a nezávisle na předchozí kontrole sítě asociačních vazeb provede právní tým *Verifikátora DB* analýzu a posouzení sestavení zvlášť 1. vrstvy a zvlášť 2. vrstvy CzechVoc s výsledným zpracováním výstupu pro zadavatele: souhrnné zprávy obsahující zdůvodněná doporučení resp. návrhů ke změnám v organizaci tezauru.

### 3.4.6 KPI procesu kontroly tezauru CzechVoc (pro *Dodavatele DB*)

Kontrola správnosti pojmů bude posuzovat tato hlediska:

- správnost extrakce definic pojmů z ustanovení
- správnost asociačních vazeb konkrétní pojem  $\leftrightarrow$  fragment
- správnost typu asociační vazby (definiční/meritorní)

Kontrola bude sledovat následující hodnoty vždy pro jednotlivé pojmy a na jejich základě generovat následující aktivity:



- **OK:** 100% kontrolovaných hledisek (viz výše) je u pojmu správně
- **OK s výhradou:** 0 – 3% chyb asociačních vazeb konkrétní pojem  $\leftrightarrow$  fragment nesprávně
  - *Dodavatel DB* tvorby opraví nesprávné asociační vazby a znovu opravený výsledek opakovaně předá *Verifikátorovi DB*
  - *Verifikátor DB* provede kontrolu opravy zpracování *Dodavatelem DB*
- **NOK 1.varianta:** > 3% chyb asociačních vazeb konkrétní pojem  $\leftrightarrow$  fragment nesprávně
  - *Dodavatel DB* znovu vytvoří asociační vazby daného konkrétního pojmu na fragmenty
  - *Verifikátor DB* provede opakovanou kontrolu asociačních vazeb pro konkrétní pojem
- **NOK 2. varianta:** definice pojmu je extrahována nesprávně
  - *Dodavatel DB* znovu extrahuje definici pojmu
  - *Verifikátor DB* provede opakovanou kontrolu definice pojmu

Poznámka: Proces kontroly ani KPI nezachycují potenciálně chybějící vazby, pouze ověřují správnost přítomných vazeb.