

Jak na číselníky

Číselníkem je seznam nějak souvisejících položek. Číselníky se mohou používat například jako seznam možných hodnot pro různé vlastnosti datových entit, což lze využít pro formuláře pro zadávání dat, aplikace vizualizující data apod. Ukázkou číselníku může být například [Číselník důvodů k ukončení pracovní neschopnosti České správy sociálního zabezpečení](#), který obsahuje například následující položky:

- aptus (norm. ukončení nemoci)
- důchod starobní
- mateřská dovolená
- nástup do lázní
- exitus

Použití tohoto seznamu si lze představit třeba v personálním systému nějaké organizace, kde se evidují pracovní neschopnosti zaměstnanců.

Tato podoba, tedy textový seznam, není příliš vhodná pro strojově čitelnou reprezentaci.

Identifikace položek

Největším problémem je totiž to, že jednotlivé položky jsou identifikovatelné pouze svým názvem. Záznamy, které by tyto položky využívaly, by jejich název okopírovaly. Pokud by pak došlo ke změně v názvu položky, rázem by přestaly být validní všechny záznamy tuto položku využívající.

Je tedy třeba zavést identifikaci položky nezávisle na jejím názvu. Pro každou položku se tedy zaznamená její kód:

Kód	Název
0	aptus (norm. ukončení nemoci)
1	důchod starobní
3	mateřská dovolená
5	nástup do lázní
E	exitus

Nyní se ostatní systémy mohou odkazovat na kód položky, a jejich záznamy zůstanou validní i když se název položky změní. Identifikace položek tímto způsobem je ale pouze lokální. Aby je ostatní systémy mohly využívat, potřebují k tomu vědět, o který číselník kterého poskytovatele dat se jedná, protože číselníků, ve kterých je použit kód E je na světě jistě mnoho.

Je zde tedy jasná potřeba identifikovat položky číselníků lépe, ideálně tak, aby byly identifikovány naprosto jednoznačně. Pro tento účel se používají identifikátory IRI - Internationalized Resource Identifier, které jsou založeny na doménových jménech jednotlivých poskytovatelů dat a zaručují jednoznačnou identifikaci. Číselník s položkami identifikovanými IRI pak může vypadat například takto:

IRI	Kód	Název
https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/0	0	aptus (norm. ukončení nemoci)
https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/1	1	důchod starobní
https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/3	3	mateřská dovolená
https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/5	5	nástup do lázní
https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/E	E	exitus

Nyní je každá položka globálně jednoznačně identifikována svým IRI, a to napříč systémy, které tyto položky využívají.

Další vlastnosti položek

Některé položky mohou mít kratší název, například pro zobrazení ve formuláři, a delší popis, který detailně vysvětluje význam dané položky, například:

IRI	Kód	Název	Popis
https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/0	0	aptus	Aptus - Normální ukončení nemoci

Dále je čím dál častější reprezentovat vícejazyčné položky číselníků, tedy název a popis položky jak česky, tak třeba anglicky:

IRI	Kód	Název česky	Název anglicky	Popis česky	Popis anglicky
https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/1	1	důchod starobní	old-age pension	starobní důchod dle § 29 odst. 1 a 3 písm. a), § 74, § 76 a § 94 z. č. 155/1995 Sb. nebo příslušných ustanovení předcházejících předpisů, zahrnut i důchod SIN dle § 29 odst. 4 tétož zákona	Old-age pension in accordance with Section 29(1) and (3)(a) and Sections 74, 76 and 94 of Act No. 155/1995 Coll. or relevant provisions of previous regulations, including pension SIN in accordance with Section 29(4) of the same Act

SKOS - standard pro číselníky

Zde začíná být patrný další problém, a to jak číselníky reprezentovat tak, aby každý uživatel poznal, co je kód, co je název a v jakém je jazyce. I za předpokladu tabulkové reprezentace je variant totiž příliš mnoho, a každý poskytovatel dat si zajisté vybere jinou. Zpravidla se jedná o ruční práci, jelikož se číselník od číselníku liší:

- Názvy sloupců
 - Diakritika
 - Mezery a bílé znaky
 - Speciální znaky
- Pořadí sloupců
- Kódování souboru
- Formáty čísel
- Formáty datumů
- apod.

Je tedy opět potřeba standardizovat podobu samotných číselníků, nejen identifikaci jejich položek. Pro publikaci číselníků na Webu se používá W3C doporučení [SKOS - Simple Knowledge Organization System](#), jehož použití všechny výše uvedené problémy řeší. Jako datový formát používá W3C doporučení RDF - Resource Description Framework, které má vestavěnou podpuru pro jednoznačnou identifikaci entit, jejich vlastností i přirozených jazyků, ve kterých jsou psány názvy a popisy. Položka číselníku popsána pomocí doporučení SKOS pak vypadá takto:

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

<https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/1> a skos:Concept ;
    skos:prefLabel "důchod starobní"@cs, "old-age pension"@en ;
    skos:notation "1" ;
    skos:inScheme
<https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/ConceptScheme> ;
    skos:note "starobní důchod dle § 29 odst. 1 a 3 písm. a), § 74, § 76 a § 94 z. č. 155/1995 Sb. nebo příslušných ustanovení předcházejících předpisů, zahrnut i důchod SIN dle § 29 odst. 4 téhož zákona"@cs,
        "Old-age pension in accordance with Section 29(1) and (3)(a) and Sections 74, 76 and 94 of Act No. 155/1995 Coll. or relevant provisions of previous regulations, including pension SIN in accordance with Section 29(4) of the same Act"@en .
```

Vlastnost <http://www.w3.org/2004/02/skos/core#inScheme> přiřazuje danou položku do daného číselníku, který je zde identifikován jako <https://data.cssz.cz/resource/ciselniky/ciselnik-duvodu-k-ukonceni-pracovni-neschopnosti/ConceptScheme>.

Reprezentace číselníků v RDF dle SKOS má ještě další výhodu. Při správné implementaci při přístupu k URI položky či číselníku, tj. jeho použití jako URL - Uniform Resource Locator, dostane zájemce, ať už člověk či aplikace, přímo reprezentaci této položky, což je v případě jiných reprezentací nemyslitelné.

Položka číselníku dle SKOS

V doporučení SKOS se položce číselníku říká Koncept - skos : Concept a číselníku samotnému Schéma konceptů - skos : ConceptScheme.

Postup pro reprezentaci číselníku pomocí SKOS

1. Zvolte vzorec, dle kterého budete přidělovat IRI číselníkům a jejich položkám (viz. Konvence propojených dat)
2. Ze zdrojového systému převedte číselníky do RDF reprezentace dle doporučení SKOS
3. Volitelně poskytněte reprezentaci v některých nižších formátech - CSV, JSON, XML či Excel
4. Číselník opatřete metadaty a zaregistrujte v NKOD

Další příklady číselníků ve SKOS

[Standardizované číselníky dle doporučení SKOS](#) publikuje také [Úřad pro publikace Evropské Unie](#).

Reprezentace v ostatních otevřených formátech

Je-li číselník reprezentován pomocí SKOS, lze z něj snadno vytvořit i odvozené, zjednodušené reprezentace, které ovšem budou postrádat výhody popsané výše, a pro které standardy nejsou. Opačným směrem je transformace zpravidla výrazně těžší. Doporučujeme tedy primárně reprezentovat číselníky v RDF dle SKOS.

Příklad reprezentace v CSV tabulce:

```
kod_polozky,nazev_polozky,popis_polozky
0,aptus (norm. ukončení nemoci),normální ukončení nemoci
1,důchod starobní,důchod starobní
3,mateřská dovolená,mateřská dovolená
5,nástup do lázní,nástup do lázní
E,exitus,exitus
```

Reprezentace v JSON souboru:

```
[
  {
    "kod": "0",
    "nazev": "aptus (norm. ukončení nemoci)",
    "popis": "normální ukončení nemoci"
  },
  {
    "kod": "1",
    "nazev": "důchod starobní",
    "popis": "důchod starobní"
  },
]
```

```
{
  "kod": "3",
  "nazev": "mateřská dovolená",
  "popis": "mateřská dovolená"
},
{
  "kod": "5",
  "nazev": "nástup do lázní",
  "popis": "nástup do lázní"
},
{
  "kod": "E",
  "nazev": "exitus",
  "popis": "exitus"
}
]
```

Jiná reprezentace v JSON souboru:

```
{
  "0": {
    "nazev": "aptus (norm. ukončení nemoci)",
    "popis": "normální ukončení nemoci"
  },
  "1": {
    "nazev": "důchod starobní",
    "popis": "důchod starobní"
  },
  "3": {
    "nazev": "mateřská dovolená",
    "popis": "mateřská dovolená"
  },
  "5": {
    "nazev": "nástup do lázní",
    "popis": "nástup do lázní"
  },
  "E": {
    "nazev": "exitus",
    "popis": "exitus"
  }
}
```

Reprezentace v XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<polozky>
  <polozka>
    <kod>0</kod>
    <nazev>aptus (norm. ukončení nemoci)</nazev>
    <popis>normální ukončení nemoci</popis>
  </polozka>
```

```
<polozka>
  <kod>1</kod>
  <nazev>důchod starobní</nazev>
  <popis>důchod starobní</popis>
</polozka>
<polozka>
  <kod>3</kod>
  <nazev>mateřská dovolená</nazev>
  <popis>mateřská dovolená</popis>
</polozka>
<polozka>
  <kod>5</kod>
  <nazev>nástup do lázní</nazev>
  <popis>nástup do lázní</popis>
</polozka>
<polozka>
  <kod>E</kod>
  <nazev>exitus</nazev>
  <popis>exitus</popis>
</polozka>
</polozky>
```

From:
<https://opendata.gov.cz/> - **Otevřená data**

Permanent link:
<https://opendata.gov.cz/draft:%C4%8D%C3%ADseln%C3%ADky?rev=1568706005>

Last update: **2020/06/03 09:36**

