

# Schémata v angličtině

V oblasti IT je častá představa, že cokoliv se programuje nebo strojově zpracovává, musí být anglicky. Tuto představu živí několik důvodů, avšak většina z nich je v dnešní době již přežita. Naopak se více a více ukazují nevýhody tohoto přístupu a problémy, které způsobuje, obzvláště když s daty mají pracovat lidé, kteří je netvořili. Projdeme si důvody tuto představu živící, a postupně si je vyvrátíme.

## Čeština se v kódu nepoužívá, protože to nefunguje

Tvrzení, že pokud se v datech či programech použije čeština, tak „to nebude fungovat“, se týká pouze zastaralých programů, které nejsou připraveny pro práci na Webu. Datové formáty i programovací jazyky dnes používají kódování UTF-8, které bez problémů zvládá nejen češtinu, ale i čínštinu, japonštinu atd.

## Anglicky přece umí každý slušný programátor

Ne každý a ne stejně dobře. Skutečný význam dat pak prochází dvěma překlady - z češtiny do angličtiny na straně poskytovatele, a z angličtiny zpět do češtiny na straně zpracovatele. Jednak to znamená netriviální úsilí, a navíc tak často dochází k efektu „tiché pošty“, tedy že význam dat je překlady názvů položek zbytečně pokřiven. Otevřená data často vychází z legislativy, která je česky. Odpovídající anglické výrazy pak často nejsou oficiálním odborným překladem a nelze zaručit jejich správnost, jednoznačnost a pochopitelnost uživateli.

## Když to bude anglicky, bude s tím umět i zahraniční programátor

Pokud budou například JSON klíče nebo CSV sloupce v angličtině, zvyšuje se riziko, že zahraniční programátor bude jejich význam odvozovat právě z jejich přeloženého názvu. To je ale často zavádějící, protože význam datových položek nelze vystihnout jedním či dvěma slovy, kterými se položky typicky pojmenovávají. Pro pochopení významu dat je potřeba lidsky čitelná dokumentace. A ta, pokud existuje, je často pouze česky, a to i pro data používající anglicky pojmenované položky. Pokud tedy chceme umožnit používat data i zahraničním uživatelům, je třeba zejména zpracovat jejich dokumentaci. Samotné datové položky pak mohou zůstat česky. Těch zahraničních programátorů pracujících s českými daty bude stejně menšina. S výhodou lze pro definici položek ve více jazycích použít [Propojená data](#), která umožňují skloubit data přímo s popisem jejich významu.

## Je to anglicky, protože používáme celosvětové či Webové standardy

Pokud mají data splňovat zahraniční standard, pak pochopitelně budou používat anglické názvy položek. Situace je ale typicky taková, že česká data jsou do reprezentace dle zahraničního standardu

exportována za účelem interoperability. Tato transformace ale bývá ztrátová, co se týče právě významu dat - mezinárodní standardy bývají záměrně obecnější, aby vyhovovaly co nejširšímu počtu uživatelů. V tomto případě je tedy vhodné data poskytovat ve dvou verzích - plnohodnotně česky, kde se neztrácí význam, a anglicky pro potřeby interoperability se zahraničním standardem.

## Závěrem

Ať už tvoříme vlastní datové schéma v angličtině z jakéhokoliv důvodu, ve veřejné správě je třeba zajistit odborný a garantovaný a validovaný překlad termínů ve schématu a také anglickou dokumentaci. Překlad z češtiny do angličtiny programátorem rozhodně nestačí. Bez garantovaných a úplných překladů schématu i dokumentace hrozí vysoké riziko dezinterpretace dat. Dále je na zvážení, zda dává schéma v angličtině smysl v případě dat s ryze českými texty. Pokud na angličtině poskytovatel trvá, měl by pak trvat i na překladu dat samotných, ne jen jejich schémat.

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

Permanent link:  
<https://opendata.gov.cz/%C5%A1patn%C3%A1-praxe:sch%C3%A9mata-v-angli%C4%8Dtin%C4%9B>

Last update: **2021/02/04 09:41**

