

Žádné nebo nejednoznačné identifikátory

Častým problémem v datových sadách jsou odkazy na nedostatečně identifikované entity. Příkladem může být datová sada příjemců dotací, kde příjemce dotace je identifikován pouze názvem, nikoliv pomocí IČO, nebo identifikace adresy několika textovými poli místo IRI adresního místa.

Řešení

Řešením je si nejprve rozmyslet, na které entity (firmy, školy, adresy, ...) se v datové sadě odkazují, nejlépe formou konceptuálního modelu dat ([ER model](#) či [UML diagram tříd](#)), který je následně součástí dokumentace datové sady. Druhým krokem je zjistit, jak se tyto entity obvykle jednoznačně identifikují, a tyto identifikátory v datové sadě použít.

Nejlepším řešením je použít globální jednoznačné identifikátory ve formě IRI tam, kde je primární správce dat o tomto typu entit již zdefinoval. Příklady:

- Registr územní identifikace, adres a nemovitostí RÚIAN, který pro každý územní prvek definuje jednoznačné IRI v metadatovém profilu, např.
<https://linked.cuzk.cz/resource/ruian/adresni-misto/16135661>
- Registr práv a povinností, který definuje mimo jiné identifikátory pro Orgány veřejné moci, např.
<https://rpp-opendata.egon.gov.cz/odrpp/zdroj/organ-veřejné-moci/00007064>

Tam, kde správce ještě nedefinoval IRI pro globální jednoznačnou identifikaci jeho entit, je třeba o tuto definici správce požádat a do té doby prozatimně použít lokální identifikátor, např. 00007064 pro IČO, který je pro daný typ běžně používaný.

Výhody identifikace entit pomocí IRI

- Identifikátor je globální - datové sady, které ho používají, jsou automaticky integrovatelné
- Není třeba vysvětlovat, čemu daný identifikátor náleží - je to jasné přímo z něj
- V případě správné implementace lze v případě zájmu získat o dané entitě další informace pouze na základě jejího identifikátoru.

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

Permanent link:
<https://opendata.gov.cz/%C5%A1patn%C3%A1-praxe:nejednozna%C4%8Dn%C3%A9-identifik%C3%A1tory>

Last update: **2020/06/03 09:37**

