

# Špatné dělení dat do distribucí datové sady

Častým problémem je chybné dělení dat do distribucí datové sady. Základním pravidlem je, že pokud má datová sada více distribucí, liší se pouze formou, nikoliv obsahem. Typicky tedy datová sada má pouze jednu distribuci. Více distribucí, které se liší pouze formátem, může datová sada mít pro vyšší komfort uživatele, který si může vybrat takový formát, který mu nejvíce vyhovuje. Distribuce v neotevřených či strojově nečitelných formátech je možné v NKOD mít, ale pouze v případě, že datová sada obsahuje i otevřenou a strojově čitelnou distribuci.

Metadata datové sady se řídí doporučením W3C, tedy webovým standardem, [DCAT - Data Catalog Vocabulary](#), a jeho aplikačním profilem pro Evropu [DCAT-AP](#). Ten určuje základní strukturu katalogizačního záznamu. Část popisu dat která je nezávislá na konkrétním technickém provedení přístupu k datům, jako třeba název, popis, periodičita aktualizace, klíčová slova, prostorové a časové pokrytí, apod. patří na úroveň datové sady. Část popisu dat která se týká konkrétního způsobu přístupu k datům, tedy kde leží data v souboru, v jakém formátu, s jakým schématem, pod jakými podmínkami užití apod., patří na úroveň distribuce datové sady.

Cílem popisu dat metadaty je, aby uživatel, ať už člověk ručně, či aplikace automatizovaně, byl schopen data, tedy soubor ke stažení, nalézt pomocí jeho metadatového popisu.

## Chybný příklad 1: Dělení do distribucí dle času, místa, oddělení apod.

**Datová sada:** Faktury

**Jednotlivé distribuce:** Faktury 2014, Faktury 2015, Faktury 2016 apod.

**Způsobený problém:** Uživatel hledající data za rok 2014 tuto datovou sadu nenajde. Vyhledávání totiž typicky probíhá podle metadat datové sady. U distribuce Faktury 2014 ani žádná metadata o časovém pokrytí NKOD neneviduje - eviduje je u datové sady.

**Řešení:** Každá z distribucí ve skutečnosti reprezentuje samostatnou datovou sadu. Časové pokrytí či prostorové pokrytí jsou části metadat datové sady.

Další špatné příklady:

- [Poskytnuté záštity města \(Statutární město Hradec Králové\)](#)
- [Aplikace SAKO - nahlašování přeplněných kontejnerů \(Statutární město Brno\)](#)

## Chybný příklad 2: Zneužívání datové sady pro seskupování různých souborů

**Datová sada:** Informace o turistických cílech

**Jednotlivé distribuce:** Hrady, Zámky, Infocentra apod.

**Způsobený problém:** Dle metadat (krom názvu distribuce) nelze zjistit, která distribuce obsahuje co. Tyto informace musí být na úrovni datové sady.

**Řešení:** Každá z distribucí ve skutečnosti reprezentuje samostatnou datovou sadu. Každá je o jiném typu objektu, a to se musí projevit minimálně i v názvu, popisu a klíčových slovech dané datové sady.

Další špatné příklady:

- [Data200 \(Zeměměřický úřad\)](#)

## Správný příklad

**Datová sada:** Agendy

**Jednotlivé distribuce:** JSON, JSON-LD, případně XML, CSV, XLSX apod.

V každé distribuci je věcně stejný obsah. Liší se jen datový formát. Speciálním případem mohou být distribuce, které mají sice stejný formát, ale liší se vnitřní strukturou - schématem. Tedy třeba XML dle 2 různých XSD v případě, že pro jedna data existují 2 standardy.

Další správné příklady:

- [Agendy \(Ministerstvo vnitra\)](#)
- [Číselník druhů akcí v tiskopise Oznámení o nástupu do zaměstnání \(Česká správa sociálního zabezpečení\)](#)

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

Permanent link: <https://opendata.gov.cz/%C5%A1patn%C3%A1-praxe:%C5%A1patn%C3%A9-d%C4%9Blen%C3%AD-distribuc%C3%AD>

Last update: **2020/12/10 12:50**

