

Limity počtu přístupů z dané adresy

Někdy jsou rozsáhlejší datové sady zpřístupňovány jako soubory ke stažení pro jednotlivé entity, například pomocí jejich identifikátoru. To může být implementováno buďto jako dotaz do databáze, nebo jako předgenerovaný soubor pro každou entitu. Na takto zveřejněná data pak jejich poskytovatel implementuje omezení na počet přístupů za jednotku času takovým způsobem, že tím znemožní pravidelné získávání celé datové sady. **Takto zveřejněná data nelze chápat jako data otevřená.** Příkladem takto nevhodně zveřejněných dat je [registr ARES](#).

Nejčastějšími argumenty pro toto řešení bývá obava z přetížení serveru a nákladnost jiného řešení. Oba zde vyvracíme.

Cílem otevírání dat je umožnit uživatelům udržovat si jejich co nejaktuálnější kopii. Pokud jsou data poskytována pouze jako dotazy do databáze, nebo jako velké množství souborů ke stažení, uživatelům dat nezbyvá nic jiného, než si pravidelně říkat o každý soubor či záznam zvlášť. To opravdu může vést k zahlcení serveru poskytovatele. Řešením ale není omezit počet přístupů, protože to nevede k cíli otevírání dat - s daty nebude moci v plném rozsahu pracovat nikdo.

Řešením je přidat alternativní způsob publikace této datové sady a poskytovat ji také jako jeden soubor ke stažení obsahující všechny záznamy. Tím dojde k radikálnímu snížení počtu dotazů a zatížení serveru - každý záznam se vygeneruje jednou, na straně poskytovatele se vloží do souboru ke stažení, a ten následně stahují uživatelé, každý jedním požadavkem. Uživatelé, kteří mají zájem o celou datovou sadu, stahují tento soubor. Uživatelé, kteří mají zájem o záznam o jedné entitě stahují dále běžným způsobem.

Tento způsob publikace lze nadále optimalizovat tím, že se bude publikovat například:

1. Kompletní soubor jednou měsíčně
2. Změnový soubor od začátku měsíce jednou týdně
3. Změnový soubor od začátku týdne jednou denně
4. Změnový soubor od začátku dne jednou za hodinu
5. atd.

Tím se dosáhne toho, že si každý uživatel bude moci zvolit, jak aktuální kopii si chce udržovat, minimalizuje se zátěž serveru i sítě. Navíc je toto řešení velice snadno a levně realizovatelné, typicky se jedná pouze o pravidelné spouštění dotazu nad databází.

From:

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

Permanent link:

<https://opendata.gov.cz/%C5%A1patn%C3%A1-praxe:limity>



Last update: **2017/11/28 00:00**