

1.1.1 Struktura dat v systému InspIS DATA

Systém InspIS DATA obsahuje za období 2013–2017 celkem 478 formulářů, z toho 182 platných (dosud systémově neuzavřených) a 296 neplatných (systémově již uzavřených) formulářů.³ V rámci rešerše datových zdrojů byl sledován mimo jiné celkový počet záznamů, celkový počet otázek, úroveň agregace, úroveň dotazování a využitelnost pro další analytickou činnost ve zvolených zájmových oblastech v rámci projektu KSH, a to individuálně u každého formuláře.

Celkem šest formulářů v systému Inspis DATA obsahuje 10 000 a více případů. Nejobsáhlejší šetření s názvem *Dotazník pro žáka ZV 2016–2017* obsahuje 21 457 samostatných případů. 22 formulářů potom obsahuje 5 000 a více případů a celkem 80 formulářů pak 1 000 a více případů. V úvahu je nicméně nutné vzít rovněž úroveň agregace, kdy dotazníky pro žáky a učitele budou již z principu obsáhlejší než dotazníky na úrovni školy či na úrovni ředitelů škol. Mezi rozsáhlejší zdroje patří také sdružené formuláře z inspekční činnosti – tzv. *hospitační záznamy*, jež jsou zpravidla shromažďovány na úrovni třídy. Z pohledu celkového počtu otázek v jednotlivých formulářích jsou nejobsáhlejší zejména formuláře s menším počtem případů společně se sdruženými hospitačními záznamy. Pokud se však jedná o úroveň školy, u šetření obsahujících kolem 1 000 škol lze již předpokládat vysokou vypovídající hodnotu vzhledem k podílu na celkovém počtu škol v ČR.

V rámci párování dat ze systému InspIS DATA s daty z mezinárodních šetření PISA 2015 a TIMSS 2015 bylo využitelných přibližně sedm různých datových souborů, a to pro PISA 2015 i TIMSS 2015 (úroveň napárování se zde pohybuje v rozmezí 30–100 %, přičemž soubory s úrovní napárování nižší než 30 % byly stanoveny jako zdroje bez vypovídající statistické hodnoty). Pro šetření PISA 2015 jsou ze sedmi párovaných souborů čtyři využitelné jak z hlediska tematického, tak z hlediska dostatečného počtu případů, počtu otázek, úrovně agregace i úrovně dotazování. Tři zbylé datové soubory jsou pro párování s mezinárodními šetřeními nevhodné buď tematicky, nebo dotazovanou cílovou skupinou, která nezahrnuje skupiny testované v mezinárodních šetřeních. Problémem je pak zejména časové hledisko, kdy žádný z vybraných datových souborů neobsahuje data sbíraná ve stejném časovém období realizace šetření PISA 2015, nýbrž data sbíraná jeden až dva roky po provedení mezinárodního šetření. Ačkoli jsou zejména kontextové faktory v tak krátkém časovém úseku zpravidla stabilní, rozdíl v čase sběru dat by snižoval celkovou validitu dalších analýz využívajících propojování datových souborů. Pro šetření TIMSS 2015 je ze sedmi datových souborů pět využitelných jak tematicky, tak z hlediska dostatečného počtu případů, počtu otázek, úrovně agregace i úrovně dotazování. Dva zbylé datové soubory jsou nevyužitelné především tematicky. Podobně jako u párování s mezinárodním šetřením PISA 2015, tak i u šetření TIMSS 2015 je problémem časové hledisko, které snižuje celkovou validitu dalších provedených analýz.

Z celkového počtu 478 formulářů je 26 využitelných pro samostatné sekundární či dílčí analýzy, případně využitelných po propojení s dalšími datovými soubory ze systému InspIS. Dalších 56 souborů je využitelných částečně, buď jen po výběru konkrétních proměnných, nebo jen po propojení s dalšími datovými soubory ze systému InspIS s podobným tematickým či kontextuálním zaměřením.

³ Uvedený počet formulářů byl aktuální k datu 20. 7. 2017. Uplatnitelnost formulářů zpřístupněných po tomto datu nebyla zjišťována.