

2 Datové zdroje v ČR

Předchozí rešerše identifikovala jednotlivé instituce provádějící sběr dat, která by se následně mohla využít pro další analýzy (především pro analýzy prováděné ze strany ČŠI, dále pak byly identifikovány vybrané datové zdroje potenciálně využitelné pro potřeby aktivit KA5 v rámci projektu KSH). Ať už při propojení s daty z mezinárodních šetření, tak pro specifické tematické analýzy. Instituce z hlediska upotřebitelnosti datových zdrojů můžeme třídit do několika kategorií.

1. **Vzdělávací výsledky** – Cermat, ČŠI InspIS, Scio, Kalibro
2. **Podmínky vzdělávání** – ČŠI InspIS, ČSÚ (financování škol, náklady na žáka atd.), NÚV
3. **Socioekonomický kontext území** – ČSÚ, Agentura pro sociální začleňování Úřadu vlády ČR (identifikace vyloučených lokalit)
4. **Potřeby trhu práce** – MPSV (data o nezaměstnanosti absolventů dle typů škol)

Již předchozí kapitola se detailněji věnovala některým datovým zdrojům. Cílem této kapitoly je nalezené poznatky syntetizovat, utřídit a zpřehlednit. K tomu slouží analýza jednotlivých datových souborů, které se dotýkají oblasti vzdělávání. Zařazeny jsou i dílčí datové soubory institucí, které nebyly v předchozí kapitole detailně popsány ať už z důvodu nižší relevance, tak i z důvodů, že představují pouhý zdroj dat bez dalšího využití. Dále jednotlivé identifikované datové zdroje z hlediska zdroje, vývoje časových řad, typů proměnných a relevance dat.¹⁹ Zdroje datových souborů jsou převážně instituce a organizace, kterým se podrobněji věnovala předchozí kapitola. V rámci této analýzy samotných datových souborů je nutné upozornit, že samotné datové soubory většinou obsahují celou řadu proměnných, které jsou na rozdílné úrovni agregace, rozdílné úrovni měření (nominální, ordinální, škálová). Datové soubory zpravidla obsahují velký počet proměnných. Analyzovat jednotlivé proměnné datových souborů nedává praktický smysl a výsledná analytická tabulka by obsahovala desetitisíce řádků jednotlivých proměnných. Z tohoto důvodu jsou analyzovány datové soubory jako celek. Analýza upotřebitelnosti datových souborů je uvedena v následující tabulce.

¹⁹ V tabulce nejsou zahrnuty ty datové zdroje, které sice mohou být teoreticky využity k analýze výsledků ve vzdělávání, ale svým charakterem se nejedná o data přímo spojená s oblastí vzdělávání (například volební výsledky, složení krajských rad, složení městských rad a zastupitelstev, proměnné sociálního kapitálu, mapa exekucí a celá řada dalších veřejně dostupných dat). Zahrnuty nejsou zdroje, jejichž získání jen pro účely této analýzy by bylo komplikované (např. komerční zdroje).