



Sociologický ústav
Akademie věd ČR



Centrum
pro výzkum
veřejného
mínění

Shrnutí vývoje projektu ze strany SOÚ AV ČR

Paulína Tabery

Workshop „Data webových archivů a možnosti jejich využití“

18. 10. 2022

Výzvy projektu - kapacita

- Kapacitní/lidské výzvy
 - datoví analytici, kteří jsou schopní pracovat s velkými daty
 - schopnost pochopit omezení dat, ale také být schopen psát skripty a programovat kvůli přípravě dat na analýzu
- Kapacitní/technické výzvy
 - infrastruktura a výpočty

Výzvy projektu – povaha dat

- **Specifický typ velkých dat**
 - data z webového archivu – sociologové/sociální vědci obecně spíše pracují s daty ze sociálních médií
- **Pochopení struktury dat, potřeb projektu a možností partnerů**
 - tedy jak jsou data sklížena, co vše je ukládáno a jaké jsou možnosti vytěžování dat pro různé typy analýz, textová analýza/analýza obrázků
- **Standardizace dat**
 - dokumentace postupů pro transparentnost ve výzkumu

Výzvy projektu – právní aspekt

- Právní prostředí a přístup k datům
 - přípravné analýzy se odehrávaly pouze na veřejně dostupné /zasmluvněné části Webarchivu

Děkuji za pozornost

paulina.tabery@soc.cas.cz



Sociologický ústav
Akademie věd ČR



Centrum
pro výzkum
veřejného
mínění