

---

# Sdílení výzkumných dat

---

Sdílení výzkumných dat patří mezi základní stavební kameny otevřené vědy (open science). Díky sdílení dat je snazší ověřit výsledky výzkumu a publikovaný výzkum se tak stává transparentnější a důvěryhodnější. Jsou-li data publikovaná pod **otevřenou licenci**, mohou je pak využít další výzkumníci a také širší veřejnost. Navzdory rozšířeným představám, sdílení výzkumných dat neznamená, že všechna vaše data musí být volně přístupná komukoli. Obecně platí, že data by měla být co nejotevřenější, ale uzavřená dle potřeby – poučka v angličtině známá jako:

**As open as possible, as closed as necessary.**

V této sekci najdete obecný úvod ke sdílení dat (jaká data sdílet, proč a kdy), dále informace o možných způsobech sdílení výzkumných dat, tipy pro sdílení a také doporučení pro citování výzkumných dat a mýty o sdílení.

## Proč data sdílet?

V současné době již **řada vydavatelů a poskytovatelů financí požaduje sdílení podkladových dat** k publikovaným studiím. I pokud to však není podmínkou grantu či publikování v časopisu, existuje mnoho dobrých důvodů, proč by výzkumná data měla být co nejotevřenější – můžete z toho mít prospěch jak vy, tak širší vědecká komunita.

- Otevřením výzkumných dat zajistíte robustnost vašeho výzkumu, protože umožníte ostatním replikovat výsledky vaší studie
- Vylepšení reputace jakožto čestného a pečlivého vědce může zvýšit dopad vaší práce
- Vyšší citovanost – vaše data mohou být také citována! Výzkum navíc naznačuje, že zveřejnění dat může vést k vyšší citovanosti související publikace (např. [Piwowar et al. 2007](#), [Colavizza et al. 2020](#))
- Splnění podmínek poskytovatelů financí nebo vydavatelů
- Efektivní využití zdrojů – stejná data se nemusí vytvářet opakovaně
- Sdílení dat může urychlit výzkumný proces
- Možnost navázání nové spolupráce mezi autory dat a jejich uživateli
- Kombinování výzkumných dat z různých zdrojů může vést k novým poznatkům
- Podpora občanské vědy
- Omezení podvodů ve vědě

## Jaká data sdílet?

Obecně platí, že dobrou praxí je zpřístupnit **všechna data, která jsou potřeba k replikaci publikované studie**. Kromě samotných nasbíraných údajů to může být také počítačový skript nebo software použitý k jejich zpracování či analýze, laboratorní deník, terénní poznámky, číselníky a další materiály. Zároveň s daty je třeba sdílet také jejich **dokumentaci** (metadata), která poskytne základní informace o datové sadě (kdo je autorem, kdy vznikla, jakým způsobem byla data sbírána apod.) a umožní uživatelům data interpretovat.

Kromě podkladových dat k publikaci je také možné sdílet **samostatné datové sady**, obzvláště pokud se jedná o rozsáhlou datovou sadu, která se nevztahuje pouze k jedné publikaci a která by mohla být cenná pro další výzkumníky v daném oboru i mimo něj.

Podle oborových zvyklostí, kontextu a typu sbíraných dat se může být vhodnější sdílet buď tzv. **raw data**, tedy hrubá, nezpracovaná data, nebo **zpracovaná data**, která už byla nějakým způsobem např. seřazena, vyčištěna nebo anotována, případně oba typy. V obou případech by měl být sdílený také skript a/nebo popis metody, jakou byla data zpracována.

Je třeba pamatovat, že **ne všechna data je možné sdílet**. Sdílení může být omezeno například v případech, kdy data obsahují osobní údaje, jedná se o data spojená s (národní) bezpečností, jedná se o data třetích stran, k jejichž zveřejnění nemáte oprávnění, v případech kdy by zveřejnění dat ohrozilo dokončení výzkumného projektu nebo pokud chcete výstup svého výzkumu patentovat apod. I v případě, že není možné data sdílet, je dobré taková data uložit (např. do repozitáře) z důvodu dlouhodobé ochrany, a sdílet alespoň metadatový záznam.

## Kdy data sdílet?

Výzkumná data je možné sdílet v **jakékoli fázi výzkumného procesu**. Běžnou praxí je **sdílet dat zároveň s publikací**, ke které se vážou.

Někteří výzkumníci se však mohou rozhodnout otevřít celý výzkumný proces a sdílet data ještě před vydáním vlastní publikace, např. prostřednictvím platformy [Open Science Framework](#) (OSF). Tato platforma umožňuje také provést [předregistraci](#) výzkumného projektu, pokud máte naplánované další studie, které na zveřejňovaných datech chcete provést. Výhodou tohoto postupu je jednak transparentnost (a z toho plynoucí vyšší důvěryhodnost daného výzkumu), a jednak příležitost odhalit možné chyby ve sběru či analýze dat ještě před odesláním článku do recenzního řízení.

## Užitečné zdroje

Colavizza, Giovanni et al. 2020. The citation advantage of linking publications to research data. *PLOS ONE* 15(4). e0230416. <https://doi.org/10.1371/journal.pone.0230416>

Drachen, Thea M., Olle Ellegaard, Asger V. Larsen & Søren B. F. Dorch. 2016. Sharing data increases citations. *LIBER Quarterly* 26(2). 67–82. <http://doi.org/10.18352/lq.10149>

Piwowar, Heather A., Roger S. Day & Douglas B. Fridsma. 2007. Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2(3): e308. <https://doi.org/10.1371/journal.pone.0000308>

Rygelová, Pavla & Tereza Simandlová. Osobní údaje a politika otevřených výzkumných dat EK: Jak může pomoci OpenAIRE?: informační materiál pro výzkumníky, administrátory a koordinátory projektů: prosinec 2016. 2016. Dostupné z: <http://hdl.handle.net/10084/116555>

Open Data Institute: [Open Data Essentials](#) - e-learning course