
Uchovávání dat

Proč vybírat data k uchování?

Není možné a ani vhodné navždy uchovávat veškerá digitální data. Dlouhodobé uchovávání a správa dat s sebou nesou **další výdaje**, které by se musely vynakládat i po ukončení výzkumného projektu. Kromě nákladů souvisejících s uchováváním dat je třeba také zvážit, že při stále rostoucím objemu digitálních dat je pak **těžší** v těchto datech **najít informace**, které skutečně hledáme.

Pět kroků, které vám pomohou vybrat data pro dlouhodobé uchování

Whyte, Angus. [DCC: Five steps to decide what data to keep](#) (CC-BY)

Pět kroků, které byste měli zvážit, když se rozhodujete, jaká data by měla být uchována	
1.	Zvažte možný účel opětovného využití – k čemu by data mohla sloužit?
2.	Zkontrolujte, je-li nutné data uchovávat na základě smluvních či zákonných podmínek .
3.	Určete, která data by měla být uchována, protože mohou mít dlouhodobou hodnotu .
4.	Zvažte náklady – jaké náklady na správu dat již byly vynaloženy – a tím přispívají jejich hodnotě – a jaké další náklady jsou plánované? Odkud pochází finanční prostředky na úhradu těchto nákladů? Bude možné je čerpat i po ukončení projektu?
5.	Dokončete hodnocení vašich dat – hodnocení by mělo obsahovat seznam toho, jaká data musí, měla by nebo by mohla být uchována. Posudek by měl také shrnout kroky potřebné k přípravě dat pro dlouhodobé uchovávání, nebo zdůvodnit, proč ta která data nebudou uchovávána.

1. Identifikujte účel, který by data mohla splnit

Zvažte, který typ dat je vhodný pro opakované využití

- Zdrojová data: posbíraná nebo vytvořená data, která byla použita pro výzkum
- Sestavené datasety: Data extrahovaná nebo odvozená z (a.)
- Odkazovaná data: data přepracovávající podskupinu (b.) s cílem data dále analyzovat nebo vyvozovat závěry

Datové kolekce mohou zahrnovat i další objekty, které jsou nezbytné pro interpretaci nebo přístup k výše uvedeným datům, např. software, laboratorní deník, poznámky atd.

? je možné, že bude třeba vytvořit více datových sbírek, protože mohou existovat různá přístupová oprávnění nebo licenční podmínky

Zvažte, k jakému účelu by data mohla být využita nad rámec výzkumného projektu

- Ověření:** umožní ostatním replikovat proces, který vedl k publikovaným zjištěním
- Další analýza:** příležitost pro další analýzu získaných údajů
- Budování akademické reputace:** data, která jsou snadno objevitelná mohou zvýšit citovanost publikovaných zjištění
- Vývoj komunitních zdrojů:** publikování dat, která jsou hodnotná pro určitou skupinu uživatelů
- Další publikace:** publikování datového článku přispěje k vědecké komunikaci
- Výuka:** využití dat k interaktivní výuce
- Soukromé využití:** v budoucnu data snáze naleznete, abyste je mohli využít pro další účely

2. Ověřte, která data musí být uchována

Rozhodnutí o tom, jaká data se uchovají bude záviset také na **legislativních a smluvních podmínkách**. Zkontrolujte, zda existují nějaké institucionální zásady nebo pravidla poskytovatelů financí, která definují, jaká data se musí uchovávat – většinou se jedná o tzv. “data v pozadí grafu” - a zda by data měla být volně dostupná, nebo mít omezený přístup.

3. Identifikujte, která data by měla být uchována

Obecně platí, že data by měla být uchována, pokud to vyplývá ze smluvních podmínek (viz. krok 2) nebo pokud odpovíte “ano” alespoň na jednu otázku ze dvou různých skupin uvedených níže.

Jsou data dostatečně kvalitní?

- **Popis:** Existuje dostatek informací o tom, co je to za data a jak a proč byla sbírána
- **Kvalita:** Jsou data dostatečně kvalitní, pokud jde o úplnost, velikost vzorku, přesnost, spolehlivost, reprezentativnost nebo další kritéria, která jsou pro daný obor relevantní?

Je pravděpodobné, že po datech bude poptávka?

- **Známí uživatelé:** Jsou nějakí známí uživatelé, kteří na tato data čekají, nebo existuje nějaký důkaz o poptávce v minulosti?
- **Doporučení:** Doporučuje poskytovatel financí, instituce nebo autority v oboru sdílení tohoto typu dat?
- **Integrační potenciál:** Jsou data popsána tak, aby to odpovídalo zavedené terminologii v daném oboru?
- **Reputace:** Byla data vytvořena výzkumnou skupinou, která je známá originálním výzkumem a kvalitními výstupy? Posílí zpřístupnění dat reputaci dané výzkumné skupiny?
- **Dosah:** Mohla by data mít široký dosah? (např. vztahují se k nějakému významnému objevu?)

Jsou data snadno replikovatelná?

- **Nereplikovatelnost:** Bylo by reprodukování dat obtížné/drahé?

Nereplikovatelnost: Bylo by reprodukování dat obtížné/drahé?

- **Schválení:** Jsou data klasifikována podle jejich citlivosti a bez etických nebo licenčních podmínek omezujících přístup veřejnosti k datům a k jejich opětovnému využití?
- **Otevřený formát:** Jsou data ve formátu, který nevyžaduje placení licencí nebo využití proprietárního softwaru/hardware
- **Nezávislost:** pokud je ke znovuvyužití dat potřeba specializovaný software/hardware, je tento dostupný a v oboru běžně využíván?

Je to jediná kopie?

- **Jedinečnost:** Je to jediná a nejúplnější kopie dat?
- **Ohrožená data:** Jsou data uložena někde, kde není zaručeno jejich dlouhodobé uchování?

4. Zvažte náklady

Zvažte **ekonomické důvody** pro uchovávání dat: Kolik se již investovalo času, peněz do vybaven a poplatků za služby či licence a kolik ještě bude třeba investovat? Jsou k dispozici finanční prostředky na přípravu dat k archivaci a na úhradu poplatků za skladování a správu dat po dokončení projektu?

5. Dokončete hodnocení dat

Zvažte hodnotu dat a veškeré náklady, které bude ještě třeba vynaložit, s ohledem na dlouhodobé cíle, čas a prostředky, které již byly investovány, a nutnost připravit data, která musí být uchována (viz Krok 2)

Užitečné zdroje

Whyte, Angus & Andrew Wilson. 2010. [How to Appraise and Select Research Data](#) . Edinburgh: Digital Curation Centre

DCC. 2014. [Five steps to decide what data to keep: a checklist for appraising research data v.1](#) . Edinburgh: Digital Curation Centre.