

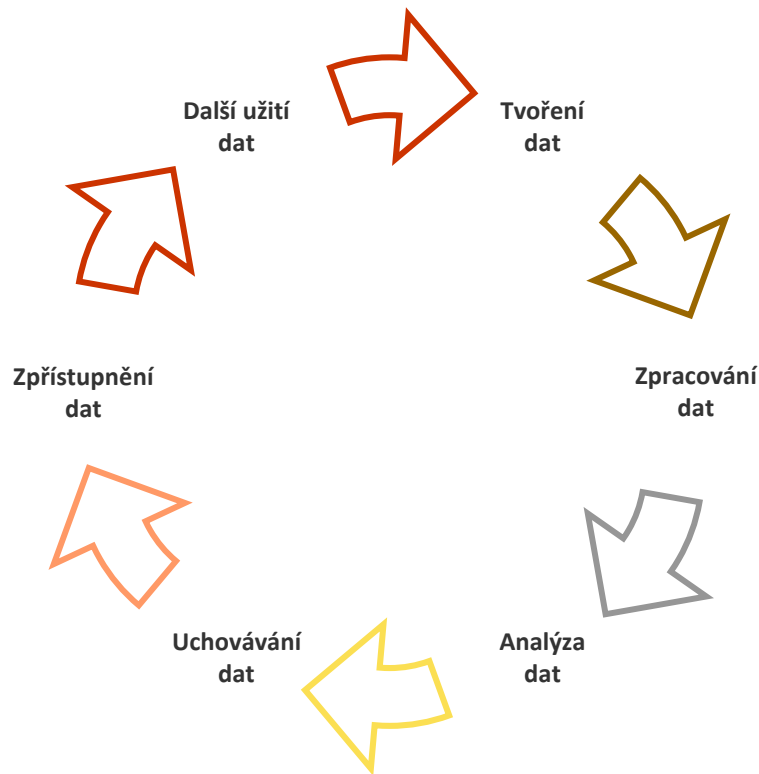
Výzkumná data na Univerzitě Karlově v Praze

Eliška Pavlásková
Ústřední knihovna UK

Výzkumná data

- Jakákoli informace uložená v digitální formě, a to včetně textu, čísel, obrázků, videa nebo filmu, zvuku, softwaru, algoritmů, rovnic, animací, modelů, simulací, atd. (National Science Board, 2005)
- Dělení dle způsobu vzniku dat (NSB, 2005; Borgmanová, 2012, 2015; Hendl 2006)
 - **Experimentální** - závislá proměnná je cíleně ovlivňována
 - **Observační** - objekty sledovány a jsou měřeny proměnné, ale zároveň proměnné nejsou ovlivňovány
 - **Komputační** - výsledek počítačového modelování, simulací či workflow
 - **Záznamy, dokumenty (records)** - téměř jakýkoli fenomén či lidská aktivita, která může být využita jako data pro výzkum (Borgmanová, 2015)

Životní cyklus (UK Data Archive)



- **Tvorba dat**

plán výzkumu, správy dat, sběr dat, metadat

- **Zpracování dat**

zadání, digitalizace, transkripce, překlad; kontrola, validace, čištění, anonymizace

- **Analýza dat**

interpretace, publikační výstup, příprava na uchovávání

- **Dlouhodobé uchování dat**

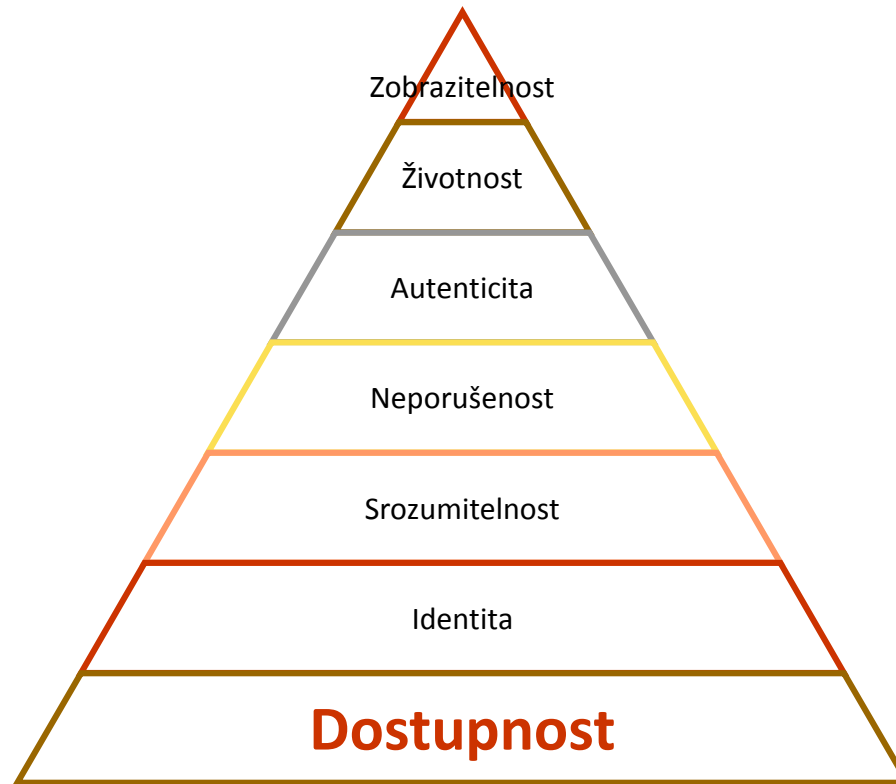
- **Zpřístupnění dat**

- **Další využití dat**

Motivace výzkumu

- Stanovení význačných charakteristik výzkumných dat
- **Zmapování využívání výzkumných dat na Universitě Karlově v Praze**
- Předběžná identifikace faktorů ovlivňujících dlouhodobé uložení výzkumných dat

Pyramida dlouhodobého uložení (Caplan, 2008)



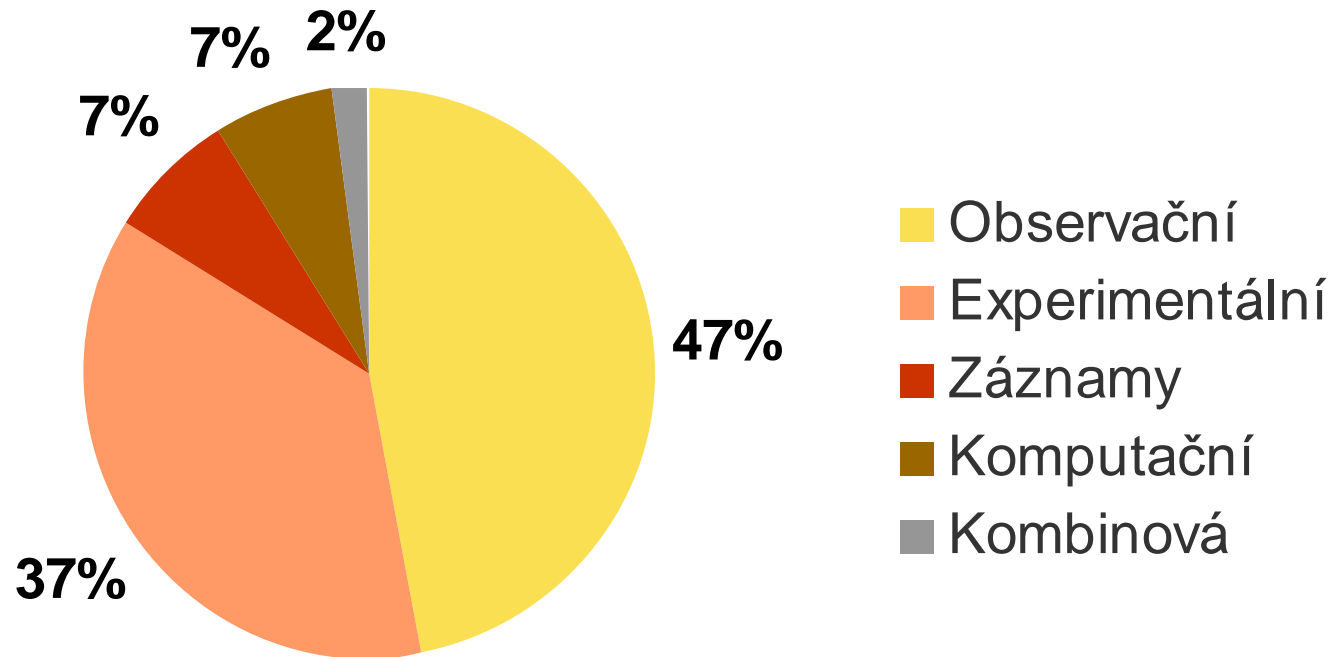
Zdroj výzkumných dat

- Disertační práce University Karlovy v Praze
 - Průřez všemi vědeckými obory
 - Jednoznačně splňují požadavky na vědecký publikační výstup
 - Nezkresleno publikační praxí
 - Pragmatické důvody
 - Možnosti rozšíření výzkumu
- Vzorek
 - 1.1.2011 - 18.6.2015
 - jazykové omezení
 - 400 prací z 2928
 - Kvótní výběr, v rámci fakulty náhodný výběr

Obsahová analýza

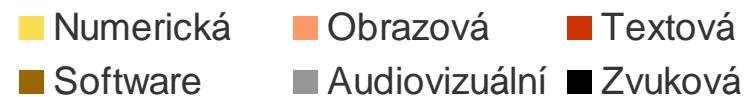
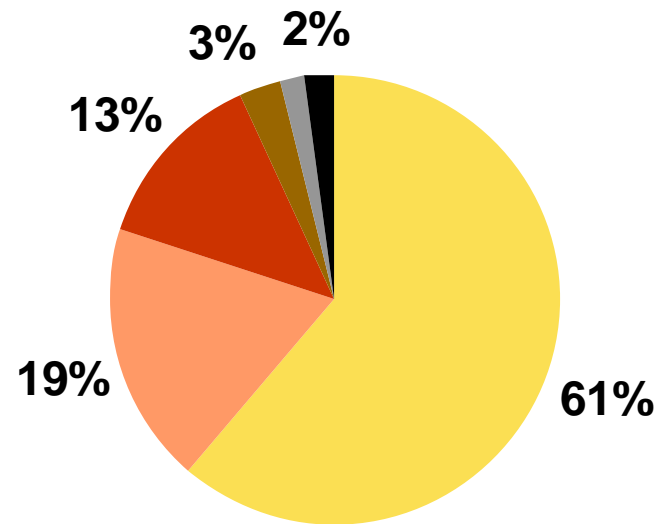
- Záznamová jednotka = kvalifikační práce, případně dataset
- Původ dat – interní/externí
- Charakter dat – zařazení do formátové skupiny
 - Obrazová
 - Zvuková
 - Audiovizuální
 - Textová
 - Numerická
 - Software
- Zdroj dat
 - **Experimentální**
 - **Observační**
 - **Komputační**
 - **Dokumenty, záznamy (records)**

Rozložení interních dat (67%) ve vzorku



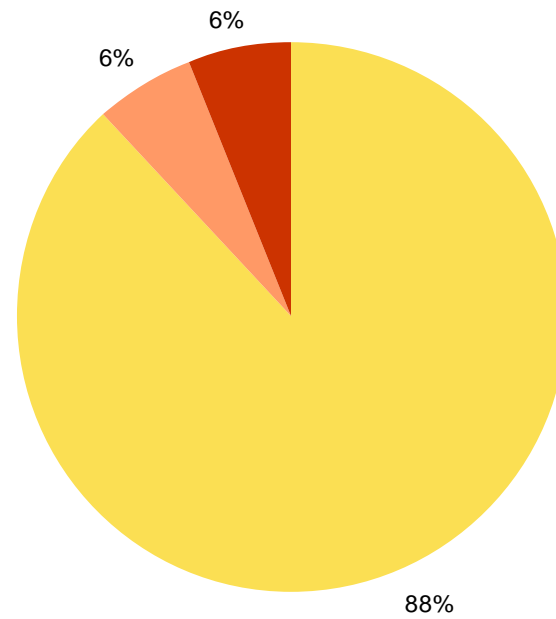
Dělení dle formátových skupin

- Pouze částečně využitelné
- Určena převládající formátová skupina
- Zpřesnění možné v případě, že budou zachyceny všechny intelektuální entity vědeckých dat účastníků se výzkumu.
- Rozlišení problematické i v případě konkrétního datasetu.
- Závěry pouze přibližné



Humanitní vědy

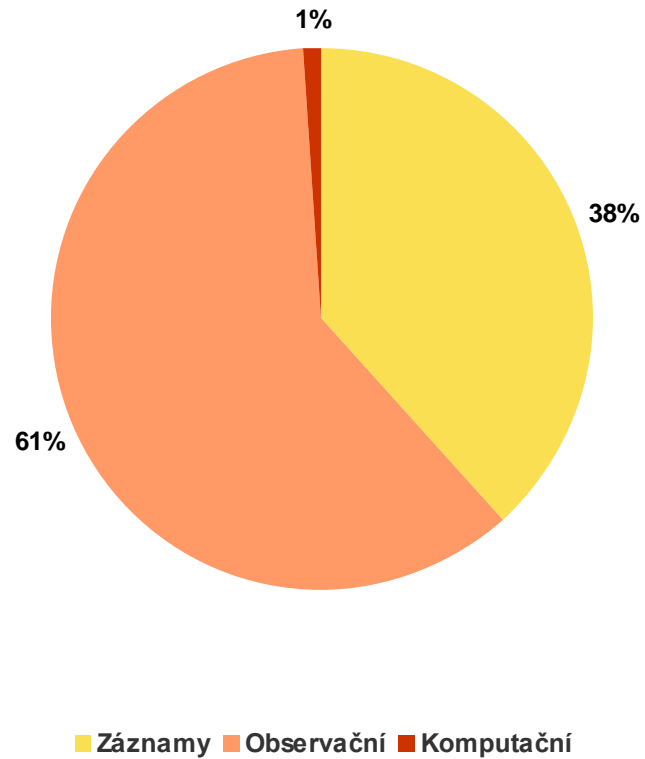
kompletní množina (51)



■ Záznamy ■ Komputační ■ Bez dat

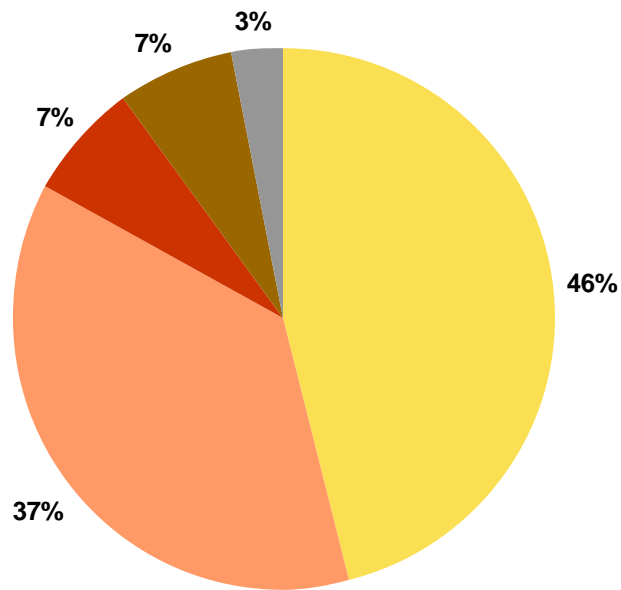
Společenské vědy

množina digitálních dat (91)



Přírodní vědy

kompletní množina (215)



■ Experimentální ■ Observační ■ Komputační ■ Bez dat ■ Kombinovaná

Shrnutí

- Dělení není aplikovatelné na medicínské obory
- Záznamy (dokumenty)
 - Převážně externí
 - Interní unikátní
- Observační
 - Unikátní
 - Tendence ke zveřejnění/sdílení
- Experimentální
 - Výrazný vliv fází životního cyklu dat
 - Problematická identifikace formátové skupiny
- Komputační data
 - Data i software pro zpracování

eliska.pavlaskova@ruk.cuni.cz

DĚKUJI ZA POZORNOST!