Etická úskalí a řešení při práci s citlivými daty ve výzkumu

Helena Klimusová

Psychologický ústav FF MU

Co jsou to citlivá data?

- zpřístupnění mimo skupinu oprávněných osob způsobí škodu velkého rozsahu se závažnými následky
- samotná **povaha dat** (data chráněná ze zákona, na základě smlouvy)
  - citlivé osobní údaje
  - genetické údaje, biometrické údaje zpracovávané výhradně za účelem identifikace člověka
  - velmi cenná výzkumná data
  - přístupové údaje k důležitým systémům a diskrétním/citlivým datům
- charakteristiky účastníka výzkumu
  - vyplývající z **cíle/zaměření výzkumu**
- další okolnosti výzkumu
  - **pasivní sběr dat**, bez informovaného souhlasu

Citlivé osobní údaje podle GDPR

- **speciální kategorie osobních údajů**
  - o rasovém či etnickém původu
  - politických názorech
  - náboženském nebo filozofickém vyznání
  - členství v odborech
  - o zdravotním stavu
  - sexuální orientaci
  - trestních deliktech či pravomocném odsouzení osob

- mají **způsobilost uškodit** v zaměstnaní, společnosti, škole atd.
- podmínky pro jejich zpracování: **souhlas**, **veřejný zájem**, nebo nezbytnost zpracování pro **vědecký výzkum**

MUNI

Jak zacházet s citlivými daty ve výzkumu?

- informovaný souhlas
- anonymizace
- bezpečné ukládání dat
- omezení přístupu
- zveřejnění citlivých dat

Informovaný souhlas účastníků výzkumu

- **svobodné** vyjádření souhlasu na základě **přiměřených** informací

- každý účastník, od kterého získáváme data, ale s principem **přiměřenosti**
- kdy **není nutné** souhlas získat?
- **kdo** ho může udělit?
- je souhlas skutečně **svobodný**?
- jakou má mít **formu**?
- **pilotní test** souhlasu

Co by měl obsahovat informovaný souhlas?

- cíl projektu
- typ shromažďovaných dat (explicitní souhlas s příp. nahráváním)
- metoda získání dat
- důvěrnost a anonymita informací
- časové závazky účastníků
- právo odmítnout sdílet informace
- právo odstoupit kdykoli ze studie (bez následků)
- právo nechat odstranit údaje z databáze (do kdy je možné)
- údaje o všech rizicích spojených s účastí
- údaje o kompenzaci vynaložené námahy a času
- kontakty na výzkumníky a kontrolní orgány
- údaje o debriefingu
- údaje o využití dat a výsledků
- potencionální přínos výzkumu
- zprostředkování výsledků výzkumu účastníkům
- příp. souhlas se zpracováním osobních údajů

Modely informovaného souhlasu

- různé modely zdůrazňují **etické principy v rozdílné míře**
  - autonomie účastníků
  - veřejný zájem
  - spravedlnost
  - důvěra
- modely souhlasu
  - specifický
  - široký
  - vrstvený/stupňovitý
  - dynamický
  - meta consent

- Wiertz, S., Boldt, J. Evaluating models of consent in changing health research environments. *Med Health Care and Philos* **25**, 269–280 (2022). https://doi.org/10.1007/s11019-022-10074-3

Specifický informovaný souhlas

- specific informed consent

- tradiční model
- souhlas pro každou jednotlivou studii
- chrání autonomii účastníků výzkumu
- náročný a méně praktický pro velké výzkumné projekty, biobanky atd.

Široký informovaný souhlas

- broad informed consent

- účastník souhlasí, že jeho data mohou být využita v různých **budoucích studiích**
- v době souhlasu **nejsou detailně specifikovány**
- vhodný pro rozsáhlé projekty, biobanky
- účastníci nemusí mít dostatečnou **kontrolu nad využitím svých dat**

Stupňovitý informovaný souhlas

- tiered informed consent

- kombinace specifického a širokého souhlasu
- účastník se **na začátku rozhodne**, jak široký bude jeho souhlas (pro konkrétní studie, nebo obecnější souhlas pro více typů výzkumů)
- **větší autonomie** účastníků výzkumu
- **náročnější na správu dat** a záznamů o udělených souhlasech

Dynamický informovaný souhlas

- dynamic informed consent

- **online platforma**
- účastník může **průběžně spravovat**, měnit a přidávat své souhlasy pro různé výzkumy
- flexibilní
- **větší kontrola účastníků** nad jejich daty
- **vyšší důvěra/ response rate** vs. **riziko „únavy z poskytování souhlasu“**
- náročný na implementaci i údržbu

MUNI

Příklad využití dynamického souhlasu: RUDY studie

- Rare UK Diseases of bone, joints, and blood vessels
- informace o kvalitě života pacientů, klinických událostech a zkušenostech s nemocí
- účastníci mohou **poskytovat a upravovat svůj souhlas prostřednictvím online platformy kdykoli během studie**
- rozhodnout, jakým způsobem mohou být použity jejich lékařské údaje, biologické vzorky a odpovědi na dotazníky, a určit, pro které typy výzkumných projektů mohou být jejich data zpřístupněna (+vybrat organizace, národní či mezinárodní, které mohou s jejich daty pracovat)
- když jsou do platformy RUDY přidávány **nové podstudie**, účastníci jsou **automaticky informováni** e-mailem a požádáni, aby znovu **zhodnotili a potvrdili svůj souhlas s daným výzkumem** (např. podstudie, která vyžadovala propojení dat pacientů se vzácnými onemocněními s daty jejich zdravých rodinných příslušníků - prostřednictvím dynamického souhlasu mohli pacienti i jejich rodinní příslušníci v reálném čase udělit souhlas pro propojení dat).
- pacienti mají **aktivnější** roli ve stanovení **priorit výzkumu**
- zlepšení **rekrutace a udržení** účastníků ve studii
- Teare, H., Hogg, J., Kaye, J. *et al.* The RUDY study: using digital technologies to enable a research partnership. *Eur J Hum Genet* **25**, 816–822 (2017). https://doi.org/10.1038/ejhg.2017.57

Meta informovaný souhlas

- meta consent

- kombinace stupňovitého a dynamického souhlasu
- postaven na online platformě, **vysoce flexibilní**
- **definuje preference pro budoucí komunikaci a rozhodování**
- účastník si volí, zda chce být dotazován na každou studii zvlášť, nebo zda chce udělit široký souhlas pro některé oblasti výzkumu, zatímco pro jiné bude chtít specifický souhlas
- může řešit problém „únavy z poskytování souhlasu“

Anonymizace dat

- klíčový nástroj při práci s citlivými daty

anonymizace vs. pseudonymizace:

- **anonymizace**
  - **nevratné** odstranění nebo změna osobních údajů tak, aby jedinec **nemohl být identifikován přímo ani nepřímo** (propojením s dalšími údaji)
  - anonymní datová sada **nepodléhá GDPR**
- **pseudonymizace**
  - **nahrazení identifikovatelných údajů** pseudonymy (např. číselnými kódy)
  - **původní identifikátory** jsou odděleně a bezpečně **uchovávány** (převodní tabulka, šifrovací klíč)
  - data jsou stále **považována za osobní údaje podle GDPR**

Metody anonymizace

- **odstranění** identifikátorů
  - nemusí být postačující (nepřímé identifikátory kombinací údajů)
- **maskování**
  - nahrazení citlivých údajů fiktivními/neurčitými hodnotami
- **kombinace** odstranění přímých identifikátorů a maskování nepřímých
- **generalizace** a **agregace**
  - specifické údaje na obecnější (datum narození -> ročník)
  - sloučení dat do větších skupin (individuální příjem-> příjem v daném regionu)

MUNI

Metody anonymizace

- **perturbace**
  - přidání šumu
- **swapping**
  - výměna hodnot mezi záznamy v datové sadě
- **anonymizace s využitím pokročilejších statistických metod** (k-anonymita, l-diverzita, T-blízkost)
- **generování umělých dat**

MUNI

Volba metody anonymizace

- důležitá hlediska
  - typ dat
  - úroveň rizika
  - zachování užitečnosti dat
  - právní aspekty

- **kombinace metod**
- rovnováha mezi ochranou soukromí a možností smysluplných analýz
- nastavení prahu rizika re-identifikace

- **softwarové nástroje**: ARX Data Anonymization Tool, sdcMicro v R

Rizika de-anonymizace

- reidentifikace pomocí **externích dat**
- útoky založené na **unikátních kombinacích atributů**
- **pokroky** v datové analýze (AI)
- **chyby** v procesu anonymizace

- příklady z praxe
  - Netflix Prize dataset
  - identifikace zdravotních záznamů guvernéra státu Massachusetts

Bezpečné ukládaní dat

- **technická opatření**
  - **šifrování** při uložení
  - šifrování při přenosu
  - zabezpečené **datové úložiště** (lokální servery, datová centra, cloudová úložiště)
  - pravidelné **zálohování** a **plán obnovy**
  - **doba uchovávání**
  - bezpečné **odstranění** dat

- **organizační postupy**
  - data management plan
  - interní směrnice pro práci s citlivými daty
  - školení personálu

Omezení přístupu k citlivým datům

- důvody
  - etické, právní, ochrana instituce
  - roste počet kolaborativních projektů
  - práce na dálku
- metody
  - autentizace a autorizace
  - role a oprávnění
  - definování pravidel
  - monitorování přístupu
  - využití repozitářů pro sdílení citlivých dat

tipy pro praxi

- princip **nejmenšího oprávnění**
- 2FA, silná hesla, správci hesel
- **šifrování přenosných zařízení**
- pravidelné **revize práv**
- bezpečnostní **audity**, testování bezpečnosti
- pravidelná **školení** zaměstnanců
- při spolupráci více pracovišť/institucí **smlouvy o sdílení dat**

MUNI

Zveřejnění citlivých dat a Open Science

- transparentnost vs. ochrana soukromí
- alternativy ke sdílení citlivých dat
  - sdílení anonymizovaných dat (agregovaných, uměle generovaných...)
  - sdílení metadat
  - kontrolovaný přístup (zabezpečené prostředí, individuální žádost)
- využití repozitářů pro sdílení citlivých dat

- příklad s daty ze seznamky OKCupid

MUNI