National Czech Programme | eosc

# Anonymizace dat v praxi: jak bezpečně sdílet citlivá data

Tomáš Čížek

Spolufinancováno Evropskou unií

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

CC BY

# Anonymizace podle GDPR (1)

https://eur-lex.europa.eu/CS/legal-content/summary/general-data-protection-regulation-gdpr.html

(26)

Zásady ochrany údajů by se měly uplatňovat na všechny informace týkající se identifikované nebo identifikovatelné fyzické osoby. Osobní údaje, na něž byla uplatněna pseudonymizace a jež by mohly být přiřazeny fyzické osobě na základě dodatečných informací, by měly být považovány za informace o identifikovatelné fyzické osobě. Při určování, zda je fyzická osoba identifikovatelná, by se mělo přihlédnout ke všem prostředkům, jako je například výběr vyčleněním, o nichž lze rozumně předpokládat, že je správce nebo jiná osoba použijí pro přímou či nepřímou identifikaci dané fyzické osoby.

# Anonymizace podle GDPR (2)

- Pokračování recitál 26:

- Ke stanovení toho, zda lze rozumně předpokládat použití prostředků k identifikaci fyzické osoby, by měly být vzaty v úvahu všechny objektivní faktory, jako jsou náklady a čas, které si identifikace vyžádá, s přihlédnutím k technologii dostupné v době zpracování i k technologickému rozvoji. Zásady ochrany osobních údajů by se proto neměly vztahovat na anonymní informace, totiž informace, které se netýkají identifikované či identifikovatelné fyzické osoby, ani na osobní údaje anonymizované tak, že subjekt údajů není nebo již přestal být identifikovatelným. Toto nařízení se tedy netýká zpracování těchto anonymních informací, včetně zpracování pro statistické nebo výzkumné účely.

# Pseudonymizace (1)

„pseudonymizací“ (se rozumí) zpracování osobních údajů tak, že již nemohou být přiřazeny konkrétnímu subjektu údajů bez použití dodatečných informací, pokud jsou tyto dodatečné informace uchovávány odděleně a vztahují se na ně technická a organizační opatření, aby bylo zajištěno, že nebudou přiřazeny identifikované či identifikovatelné fyzické osobě;

(28)

- Použití pseudonymizace osobních údajů může omezit rizika pro dotčené subjekty údajů a napomoci správcům a zpracovatelům splnit jejich povinnosti týkající se ochrany údajů. Výslovné zavedení „pseudonymizace“ v tomto nařízení nemá za cíl předem vyloučit jakákoliv další opatření týkající se ochrany údajů.

# Pseudonymizace (2)

(29)

S cílem vytvořit pobídky pro uplatňování pseudonymizace při zpracování osobních údajů by opatření pseudonymizace při současném umožnění obecné analýzy měla být možná v rámci téhož správce, pokud tento správce přijal technická a organizační opatření nezbytná k zajištění toho, aby bylo v případě daného zpracování provedeno toto nařízení a aby doplňkové informace pro přiřazení osobních údajů konkrétnímu subjektu údajů byly uchovány samostatně. Správce, který zpracovává osobní údaje, by rovněž měl označit oprávněné osoby v rámci téhož správce.

# Osobní údaje

- https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_cs
- Osobní údaje jsou jakékoli informace, které se týkají **identifikované nebo identifikovatelné žijící osoby**. K osobním údajům patří i různé jednotlivé informace, které společně jako celek mohou vést k identifikaci určité osoby.
- Osobní údaje, které sice byly zbaveny informací umožňujících identifikaci, zašifrovány nebo **pseudonymizovány,** ale lze je použít ke zpětné identifikaci osoby, zůstávají osobními údaji a GDPR se na ně i nadále vztahuje.
- Osobní údaje, které byly **anonymizovány** takovým způsobem, že příslušná osoba již není identifikovatelná, již nejsou považovány za osobní údaje. Údaje se pokládají za skutečně anonymizované, pokud je anonymizace nezvratná.

# Citlivá data – zvláštní kategorie citlivých údajů

Následující osobní údaje jsou považovány za „citlivé“ a vztahují se na ně zvláštní podmínky pro zpracování:

- **osobní údaje odhalující rasový či etnický původ, politické názory, náboženské vyznání či filozofické přesvědčení,**
- členství v odborech,
- genetické údaje, biometrické údaje zpracovávané s jediným cílem identifikovat nějakého člověka,
- **údaje související se zdravím,**
- **údaje související se sexuálním životem nebo sexuální orientací.**

Odkazy

- Článek 4 odst. 13, 14 a 15 a článek 9 a odůvodnění 51 až 56 GDPR

# Zvláštní podmínky zpracování?

- 1. **Zakazuje se** zpracování osobních údajů, které vypovídají o rasovém či etnickém původu, politických názorech, náboženském vyznání či filozofickém přesvědčení nebo členství v odborech, a zpracování genetických údajů, biometrických údajů za účelem jedinečné identifikace fyzické osoby a údajů o zdravotním stavu či o sexuálním životě nebo sexuální orientaci fyzické osoby.
- 2. Odstavec 1 **se nepoužije**, pokud jde o některý z těchto případů:
- j) zpracování je nezbytné pro účely archivace ve veřejném zájmu, pro účely vědeckého či historického výzkumu nebo pro statistické účely v souladu s čl. 89 odst. 1 na základě práva Unie nebo členského státu, které je přiměřené sledovanému cíli, dodržuje podstatu práva na ochranu údajů a poskytuje vhodné a konkrétní záruky pro ochranu základních práv a zájmů subjektu údajů.

# Anonymizace – Česká sociologická společnost

- https://ceskasociologicka.org/wp-content/uploads/2021/06/Eticky-kodex-Ceske-sociologicke-spolecnosti_ke-schvaleni-Valnym-shromazdenim.pdf
- Sociologové a sociložky musí chránit **anonymitu** a práva účastníků výzkumu, studujících, zaměstnanců, klientů a dalších osob podle platných právních ustanovení. Výzkumná data musí být chráněna a před použitím, resp. publikací anonymizována. Před získáním souhlasu s účastí na studii musí být lidé potenciálně se účastnící výzkumu písemně nebo ústně informováni o způsobech, jakými bude zajištěna **anonymita a diskrétnost při shromažďování, uchovávání, analýze, archivaci a zveřejňování informací z výzkumu**. Při práci v týmech je vhodné využívat písemných dohod o mlčenlivosti. Sociložky a sociologové mají povinnost nedovolit použití důvěrných informací způsobem, který by mohl ohrozit osoby účastnící se výzkumu, studující nebo jiné osoby. Ochrana dat a informací z výzkumu není garantována v zákonem stanovených důvodech. Pokud se zkoumající dozví o spáchání závažného trestného činu, zejm. je-li tento vyjmenován v patřičném oddíle zákona, je zkoumající povinen nahlásit jej příslušným orgánům

# Způsoby ochrany dat

Anonymizace

Pseudonymizace

Bezpečné uložení

Kontrolovaný přístup například: Data Safe Centrum/TRE

# Anomizace krok za krokem (UK data service)

https://ukdataservice.ac.uk/learning-hub/research-data-management/anonymisation/anonymisation-step-by-step/

- Anonymizace zajišťuje, že riziko identifikace subjektu údajů ve zveřejněných datech je zanedbatelné. Je třeba pečlivě zvážit platné právní předpisy, protože ty ovlivňují definici anonymizovaných dat.
- Techniky anonymizace se mohou lišit pro kvantitativní (např. dotazníková) a kvalitativní (např. přepisy rozhovorů) data. Při přípravě na anonymizaci dat byste měli nejprve zvážit provedení auditu datové situace. Audit datové situace zohledňuje každou situaci, kdy budou data sdílena nebo publikována, od úryvků dat v časopiseckých článcích až po kompletní datové sady archivované v odpovědném repozitáři.

# Krok 1: Vyhledání a posouzení identifikátorů

- Začněte identifikací potenciálních identifikátorů. To by mělo zahrnovat jak přímé identifikátory (informace přímo identifikující subjekty údajů, např. jména, adresy), tak nepřímé identifikátory (informace, které po zkombinování mohou identifikovat subjekty údajů, např. věk, pohlaví, dosažené vzdělání, povolání).
- Vyhodnoťte pravděpodobnost opětovné identifikace (reidentifikace) na základě samotných dat i potenciální dostupnosti externích informací, které by s nimi mohly být propojeny.

# Krok 1: Vyhledání a posouzení identifikátorů Otázky:

- Lze z informací v datovém souboru zjistit identitu účastníka?

- Existuje možnost neúmyslného prozrazení informací nebo způsobení újmy třetí straně na základě informací v datovém souboru?

# Krok 2: Implementace technik anonymizace

- Ujistěte se, že všechny přímé identifikátory byly odstraněny (smazány) nebo pseudonymizovány (nahrazeny fiktivními jmény nebo kódy). Dále se zaměřte na nepřímé identifikátory, které jste určili jako potenciálně vedoucí k identifikaci. Mezi techniky mohou patřit:
- **Tvorba pásem, shlukování a agregace (Banding, binning and aggregation):** Seskupování datových bodů za účelem snížení identifikovatelnosti. Například místo uvedení konkrétního věku jej zařaďte do širších věkových kategorií, jako jsou 20–24, 25–29, 30–34, 35–39 atd.

# Krok 2: Implementace technik anonymizace

- **Generalizace (Zobecnění):** Úprava podrobných informací na obecnější pojmy, aby se zabránilo identifikaci. To je velmi užitečné pro kvalitativní data, jako jsou přepisy, ale i pro dotazníková data obsahující textové proměnné. Například zobecněte „bydlí ve městě Preston v hrabství Lancashire“ na „bydlí na venkově na severozápadě Anglie“.
- **Techniky specifické pro daná data:** Začlenění specializovaných technik, jako je překódování, kódování horních/dolních hodnot (top/bottom coding) nebo statistická kontrola zveřejnění u kvantitativních dat, a metod, jako je rozmazání nebo pozměnění rysů ve vizuálních datech či zkreslení hlasu ve zvukových datech. Metodám jako rozmazání nebo zkreslení hlasu by měla být věnována pečlivá pozornost, protože v závislosti na kontextu a předpokládaném použití dat může dojít k narušení jejich použitelnosti.

# Krok 2: Implementace technik anonymizace Otázky:

- Jak lze data upravit, aby se zabránilo identifikaci a zároveň se zachovala jejich užitečnost pro sekundární analýzu?

- Jsou použité techniky anonymizace dostatečné k ochraně před opětovnou identifikací? Nezapomeňte vzít v úvahu povahu dat, včetně typu a formátu dat, jejich citlivosti a jedinečnosti, stejně jako jejich zamýšlené použití.

# Krok 3: Kontrola dat a přehodnocení případného zbývajícího rizika vyzrazení

- Zajistěte, aby byl proces anonymizace důsledně uplatněn v rámci všech dat.
- Proveďte důkladnou kontrolu, abyste potvrdili, že nezůstává žádné skutečné zbytkové riziko vyzrazení osobních nebo citlivých informací. Pokud existuje nízké zbytkové riziko prozrazení, zvažte přístup k „efektivní anonymizaci“ (PDF), což je koncept zavedený britským ICO. Další informace o klasifikaci dat najdete na naší webové stránce věnované licencím a přístupovému rámci.

# Krok 3: Kontrola dat a přehodnocení případného zbývajícího rizika vyzrazení. Otázky:

•Byly všechny identifikátory, přímé i nepřímé, adekvátně anonymizovány nebo odstraněny?

•Zbývají nějaké informace, které by v kombinaci s dalšími dostupnými daty mohly vést k identifikaci osob?

# CESSDA DMEG – doporučení k anonymizaci. Kvantitativní data

- To může zahrnovat odstranění nebo agregaci proměnných, případně snížení přesnosti či detailního textového významu proměnné;
- Agregujte nebo snižte přesnost proměnné, jako je například věk nebo místo bydliště. Obecným pravidlem je uvádět nejnižší úroveň geografického určení, která potenciálně nenaruší důvěrnost respondentů;
- Zobecněte význam detailní textové proměnné tak, že volné textové odpovědi, které by mohly vést k prozrazení identity, nahradíte obecnějším textem;
- Omezte horní nebo dolní rozsahy spojité proměnné, abyste skryli odlehlé hodnoty (outliery), pokud jsou hodnoty u určitých jednotlivců v rámci širší zkoumané skupiny neobvyklé nebo atypické.

# CESSDA DMEG – doporučení k anonymizaci. Kvalitativní data

•Používejte pseudonymy nebo obecné deskriptory (popisy) k úpravě identifikujících informací, namísto jejich pouhého vymazání (začernění);
•Naplánujte si anonymizaci již v době přepisu nebo prvotního sepisování (výjimkou mohou být longitudinální studie, pokud vztahy mezi jednotlivými vlnami rozhovorů vyžadují zvláštní pozornost kvůli harmonizaci úprav);
•Používejte pseudonymy nebo náhrady, které jsou konzistentní napříč celým výzkumným týmem a projektem. Například používejte stejné pseudonymy v publikacích i v navazujícím výzkumu;
•Funkci „najít a nahradit“ (search and replace) používejte opatrně, abyste neprovedli nechtěné změny a abyste nepřehlédli slova s překlepy;
•Náhrady v textu jasně označte, například pomocí [hranatých závorek] nebo použitím XML tagů, jako je <seg>slovo k anonymizaci</seg>;
•Vytvořte si deník anonymizace (známý také jako deanonymizační klíč) se všemi provedenými náhradami, agregacemi nebo odstraněními a uchovávejte tento deník bezpečně a odděleně od anonymizovaných datových souborů.

# CESSDA DMEG

| Identifier type | Direct identifier | Strong indirect identifier | Indirect identifier | Anonymisation method |
|---|---|---|---|---|
| Personal identification number | x | | | Remove |
| Full name | x | | | Remove/Change |
| Email address | x | x | | Remove |
| Phone number | | x | | Remove |

National Czech Programme | eosc

# Amnesia

## Privacy made simple.
## Data Anonymization.

Transform personal data into open statistical data in just a few simple steps.

Your Research. Your Data. Your Privacy.

Learn more

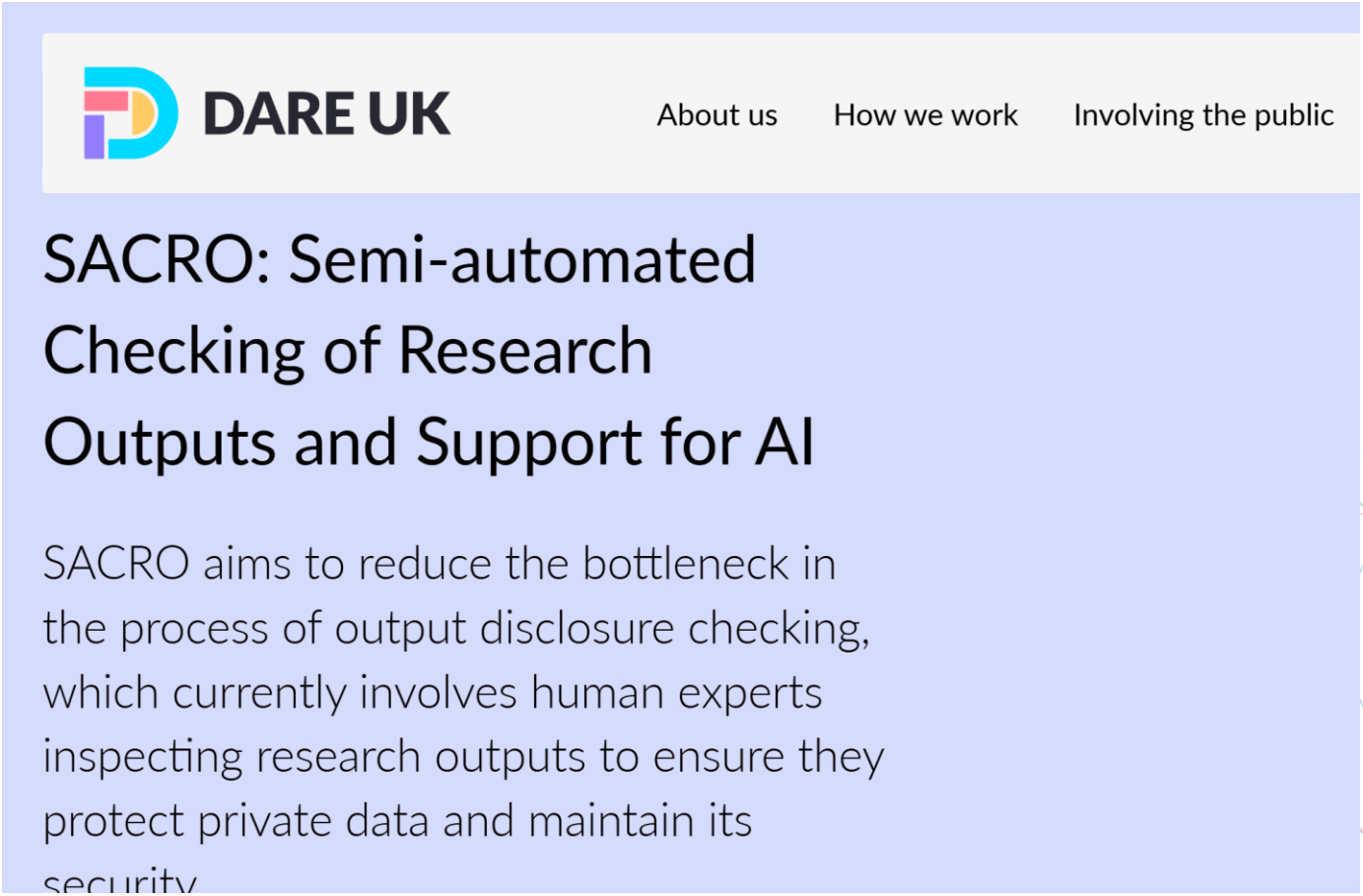
# sdcMicro

## sdcMicro


**sdcMicro** is an R-package to anonymize microdata. Most functionalities of the package are also available via an interactive shiny-based graphical user interface.

The online documentation can also be found at `sdctools.github.io/sdcMicro`.

# SACRO


DARE UK
About us
How we work
Involving the public
SACRO: Semi-automated Checking of Research Outputs and Support for AI
SACRO aims to reduce the bottleneck in the process of output disclosure checking, which currently involves human experts inspecting research outputs to ensure they protect private data and maintain its security

# ARX

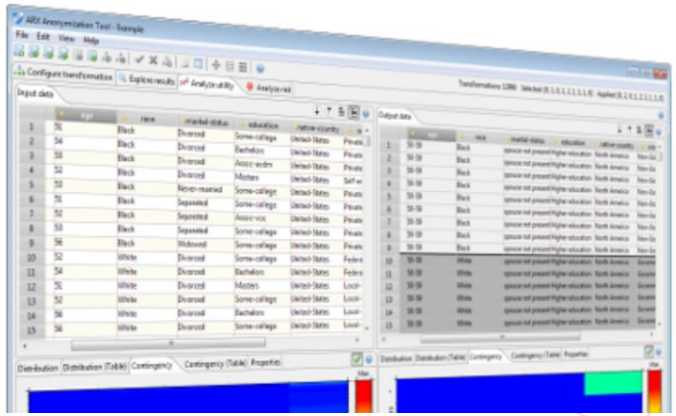
Data Anonymization Tool

Home Overview Anonymization tool Development Publications Downloa

## ARX

### Data Anonymization Tool

ARX is a comprehensive open source software for anonymizing sensitive personal data. It supports a wide variety of (1) privacy and risk models, (2) methods for transforming data and (3) methods for analyzing the usefulness of output data.



National Czech Programme | eosc

# Děkuji za pozornost.

E: tomas.cizek@soc.cas.cz

Spolufinancováno Evropskou unií

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

CC BY