



Citlivé dáta a anonymizácia

Silvia Sofianos

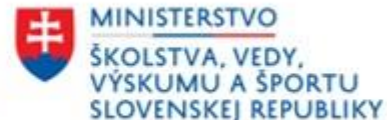
silvia.sofianos@cvtisr.sk



EURÓPSKA ÚNIA
Európske štrukturálne a investičné fondy
OP Integrovaná infraštruktúra 2014 – 2020



MINISTERSTVO
DOPRAVY A VÝSTAVBY
SLOVENSKEJ REPUBLIKY



MINISTERSTVO
ŠKOLSTVA, VEDY,
VÝSKUMU A ŠPORTU
SLOVENSKEJ REPUBLIKY



NISP^{IV}
Investícia do Vašej budúcnosti

OBSAH

- Osobné údaje a citlivé osobné údaje
- Spracovanie citlivých údajov
- Ochrana osobných údajov
- Anonymizácia, pseudonymizácia a enkrypcia
- Spracovanie citlivých údajov vo vedných disciplínach:
 - Biomedicínsky výskum
 - Spoločenské vedy a jazykoveda
 - Výskum životného prostredia a biodiverzita

Dáta alebo údaj?

Osobné údaje

- všetky informácie, ktoré sa týkajú identifikovanej alebo identifikovateľnej fyzickej osoby - jednotlivca.
- Aj informácie, ktoré po ich spojení môžu viesť k identifikácii jednotlivca, takisto predstavujú osobné údaje.
- Nie je potrebné poznať meno osoby, aby sme mohli hovoriť o identifikovaní, stačí že vieme, že je to určitá osoba, ktorú vieme odlíšiť od iných osôb práve na základe daných informácií.

Príklady osobných údajov

- meno a priezvisko,
- adresa bydliska,
- číslo preukazu totožnosti,
- emailová adresa, napr. meno.priezvisko@domena.sk,
- lokalizačné údaje, napr. funkcia na mobilnom telefóne,
- adresa internetového protokolu,
- identifikačné číslo súboru Cookie,
- údaje, ktoré uchováva nemocnica alebo lekár

Ktoré osobné údaje sa považujú za citlivé?

- osobné údaje, ktoré odhaľujú rasový alebo etnický pôvod, politické názory, náboženské alebo filozofické presvedčenie,
- členstvo v odborových zväzoch,
- genetické údaje, biometrické údaje spracúvané len na účely identifikácie ľudskej bytosti,
- zdravotné údaje,
- údaje týkajúce sa sexuálneho života alebo sexuálnej orientácie danej osoby.

Citlivé dáta, ktoré sa netýkajú osôb

- Ekologické dáta
- Komerčné dáta
- Dáta týkajúce sa národnej bezpečnosti

Príklady dát, ktoré sa môžu a nemôžu spracovávať:

- **Citlivé údaje sa môžu spracovávať**

Lekár – databáza – služobné tajomstvo

- **Citlivé údaje sa nemôžu spracovávať**

Predaj šiat – politické preferencie

Zdroj:

https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/under-what-conditions-can-my-companyorganisation-process-sensitive-data_sk

Všeobecné nariadenie o ochrane osobných údajov

- Nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 z 27. apríla 2016 o ochrane fyzických osôb pri spracúvaní osobných údajov a o voľnom pohybe takýchto údajov
- General Data Protection Regulation (GDPR)
- účinné pre každý členský štát EÚ
- v slovenskom právnom poriadku je nariadenie premietnuté aj do [zákona č. 18/2018 Z. z. o ochrane osobných údajov](#)
- chráni základné práva a slobody fyzických osôb so zameraním na právo ochrany osobných údajov

Porušenia GDPR

- Úrad na ochranu osobných údajov
 - orgán štátnej správy s celoslovenskou pôsobnosťou, ktorý sa podieľa na ochrane základných práv fyzických osôb pri spracúvaní osobných údajov
- HR – psychodiagnostické hodnotenie zamestnancov – otázna sloboda udelenia súhlasu
 - Pokuta: 40 000 eur
- Slovenský denník – zverejnené rodné čísla hercov
 - Pokuta: 1500 eur
- Google – nedostatočné informovanie o spracovaní osobných údajov
 - Pokuta: 50 mil. eur

Zdroj:

<https://www.havelpartners.sk/4-roky-s-gdpr-za-ake-porusenia-udeloval-uouu-najcastejsie-pokuty/>

<https://euractiv.sk/section/digitalizacia/news/za-nedodrziavanie-gdpr-dostal-google-vo-francuzku-pokutu-50-milionov-eur/>

Citlivé výskumné vs. osobné údaje

Citlivé výskumné údaje - príklady

- Osobné údaje – identifikačné číslo, údaje o fyzickom alebo mentálnom zdraví, kultúrne, ekonomické alebo sociálne charakteristiky, údaje o polohe z mobilných telefónov..
- Tajné údaje – obchodné tajomstvá, údaje z rôznych vyšetrovaní, heslá, finančné informácie, údaje týkajúce sa národnej bezpečnosti, vojenské informácie
- Kombinácia rôznych datasetov, ktoré možno spojiť do citlivých alebo osobných údajov
- Biologické údaje – ohrozené (rastlinné alebo živočíšne) druhy, ktorých prežitie závisí od ochrany údajov o ich polohe
- Osobné a citlivé metadáta

Citlivé výskumné údaje

- Pri obavách o práva duševného vlastníctva, ochrany osobných údajov a dôvernosti sa treba riadiť
 - „tak otvorené, ako je len možné a tak uzavreté, ako je potrebné“

Spracovanie citlivých údajov vo vedných disciplínach

Biomedicínsky výskum

- Zvyšujúci sa počet repozitárov s citlivými dátami
- Hlavne dátové inštitucionálne repozitáre bez otvoreného prístupu
- Výskumníci však musia byť schopní efektívne pristupovať do týchto databáz a registrov pacientov pomocou ochrany súkromia a oprávnených záujmov pacientov ako dotknutých osôb
- Ochrana údajov vs. bezpečnosť údajov
- Dohody o používaní údajov, etické kódexy správania, špeciálne usmernenia pre genetické/genomické údaje
 - Musí byť zahrnuté v Politikách pre využívanie dát

Spoločenské vedy a jazykoveda

- Kvantitatívne (prieskumy) a kvalitatívne (rozhovory) súbory dát
- Nahrávky osvojovania si jazyka deťmi, záznamy údajov z ceremónii domorodých kmeňov, politické názory
- Príklady repozitárov: DRYAD, DANS
- Manažment citlivých údajov je využívaný na skrytie identity respondentov pri výskumoch na ozbrojené konflikty, liečebné procesy, hendikapy..

Výskum životného prostredia a biodiverzity

- Zhromažďovanie dát pomocou crowdsourcingu (údaje zbierané dobrovoľníkmi a informácie o ich polohe poskytovaných mobilmi)
- Príklad: projekt Life+ Respira
 - Meranie znečistenie ovzdušia senzormi pripevnenými na bicykli a fyziologických parametrov senzormi pripojenými na ich telo
 - Zbierané dáta: o životnom prostredí, osobné údaje, citlivé zdravotné údaje

Ako pripraviť citlivé dáta na uchovávanie alebo zdieľanie?

- Mať dobrú bezpečnostnú stratégiu
- Citlivé dáta sa stále môžu riadiť FAIR princípmi

- 3 spôsoby ako ochrániť citlivé dáta
 - Anonymizácia
 - Pseudonymizácia
 - Enkrypcia
- Čo ak sa ani jeden z týchto spôsobov nedá aplikovať?

Anonymizácia

- Anonymizované údaje sú informácie, ktoré sa nevzťahujú na identifikovanú alebo identifikovateľnú fyzickú osobu
- Proces, ktorým sa osobné údaje upravujú do formy, že dotknutá osoba už nie je identifikovateľná
- Aby boli dáta naozaj anonymizované, proces je nezvratný

SSN	Name	Gender	Profession	Age	Date of Visit	Diagnosis
x	Alice Smith	F	Waiter	25	22. 3. 2019	Heart attack
x	Richard Johnson	M	Lawyer	54	13. 10. 2019	Cholesterol
x	Jane Brown	F	Unemployed	40	7. 2. 2020	Heart attack
x	Ralf Taylor	M	Doctor	38	30. 4. 2020	Heart attack

SSN	Name	Gender	Profession	Age	Date of Visit	Diagnosis
x	A*** S***	F	Waiter	20<Age<=25	3/2019	Heart attack
x	R*** Jo***	M	Lawyer	50<Age<=55	10/2019	Cholesterol
x	J*** Br***	F	Unemployed	35<Age<=40	2/2020	Heart attack
x	R*** T***	M	Doctor	35<Age<=40	4/2020	Heart attack

Zdroj:

<https://mostly.ai/blog/3-reasons-to-drop-classic-anonymization-and-upgrade-to-synthetic-data/>

Amnesia



- Nástroj na anonymizáciu dát
- Postup:

Krok 1: Import originálnych dát



Krok 2: Vytvorenie hierarchií



Krok 3: Anonymizácia



Krok 4: Výber riešení



Krok 5: Uloženie dát



Pseudonymizácia

- Pseudonymizované údaje sú osobné údaje, ktoré sú spracúvané takým spôsobom, aby osobné údaje už nebolo možné priradiť konkrétnej dotknutej osobe bez použitia dodatočných informácií
- Pseudonymizované údaje naďalej spadajú pod ochranu osobných údajov, keďže sa na nich naďalej hľadí ako na osobné údaje.



Personal sensitive data

This is the full data including personal and special data.

Name	John Briggs
Date of birth	14.04.87
Email	jb89@mail.com
User ID	john_briggs_89
Health	type 1 diabetes



Pseudonymous data

IDs are replaced with pseudonyms. Sensitive data is encrypted.

Names	User-78463
Date of birth	14.04.87
Email	[REDACTED]
User ID	[REDACTED]
Health	type 1 diabetes



Anonymous data

IDs removed & sensitive data randomised/generalised.

Sex	Male
Age	30-49
Health	type 1 diabetes

Zdroj:

<https://www.chino.io/compliance/pseudonymization-for-health-applications>

Aký bol pseudonym Jamesa Bonda?

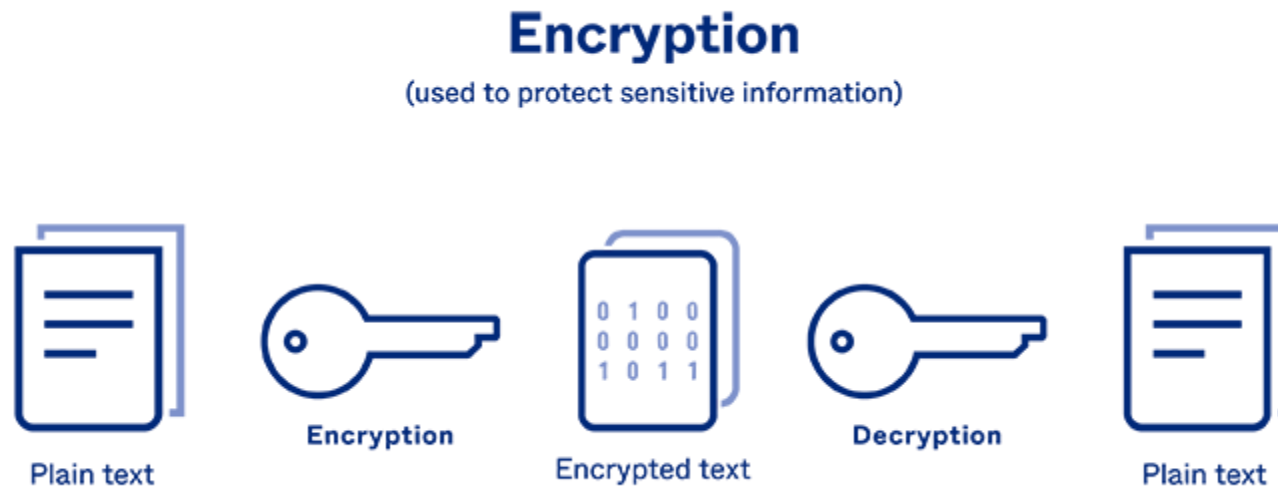


Aký pseudonym mala Judi Dench?



Enkrypcia

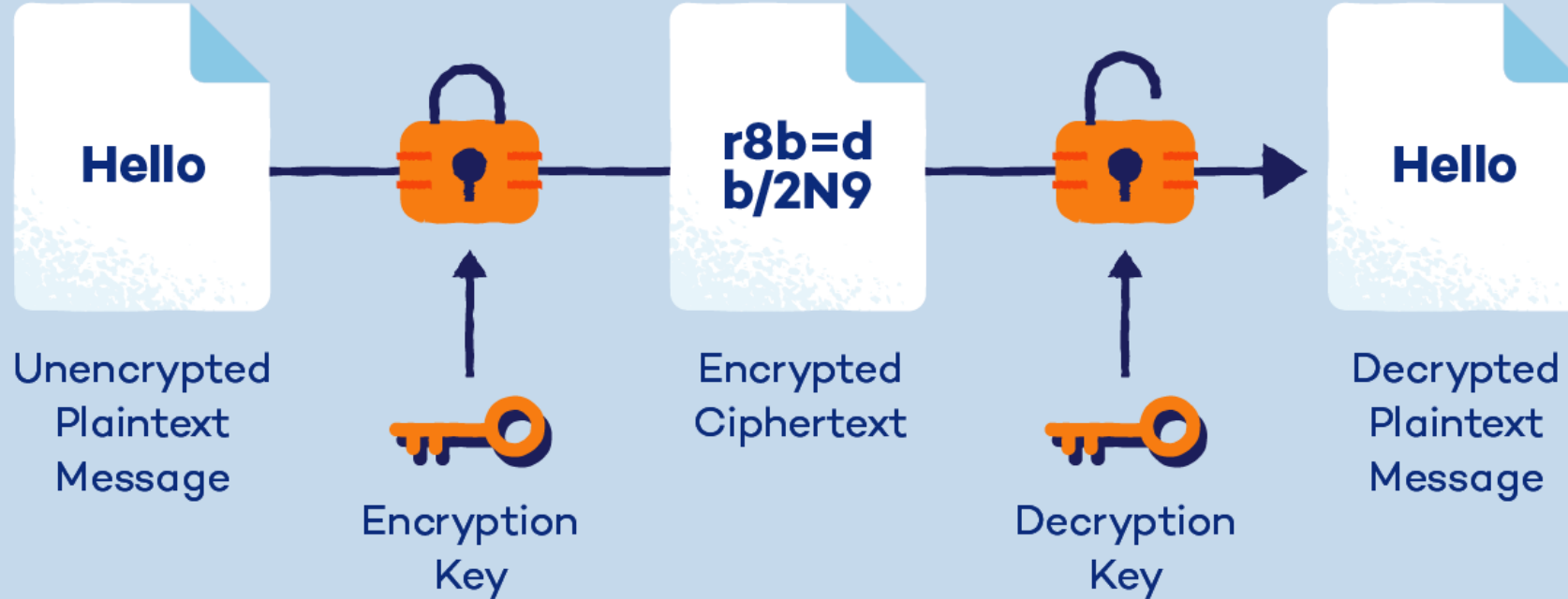
- Proces zakódovania informácií
- Používa sa náhodný šifrovací kľúč generovaný algoritmom



Zdroj:

<https://www.okta.com/identity-101/password-encryption/>

How Encryption Works



Znovu-využitie citlivých dát

- Nemôžete publikovať dáta, ktoré ste nezbierali, pokým nemajú licenciu, ktorá umožňuje redistribúciu
- Znovu-využitie dát musí byť zodpovedné a právo na súkromie účastníkov výskumu musí byť dodržané
- Použitie takýchto dát treba správne odcitovať
 - Podmienky podľa použitej licencie
 - Ak dáta nie sú licencované, treba kontaktovať vlastníka dát

Zdroje:

- Citlivé údaje https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data_sk
- Consolidated text: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504>
- Čo (nie) je osobný údaj? Je ním fotografia, telefónne číslo alebo názov firmy? <https://www.podnikajte.sk/zakonne-povinnosti-podnikatela/co-je-osobny-udaj>
- Úrad na ochranu osobných údajov – FAQ <https://dataprotection.gov.sk/uouu/sk/content/casto-kladene-otazky-k-nariadeniu-zakonu-c-182018-z-z-2>
- How can e-infrastructures deal with the sensitive data challenge? Results of the EUDAT Sensitive Data Group workshop <https://b2share.eudat.eu/records/3d1dfb9b889c4022ae7b308df009fcc9>
- OpenAIRE Sensitive Guide <https://www.openaire.eu/sensitive-data-guide>
- Publishing Sensitive Data <https://zenodo.org/record/7259742>

Ďakujem za pozornosť.

silvia.sofianos@cvtisr.sk