

Milyen támadások fenyegetik az egészségügyi adatok bizalmasságát?

Dr. Alexin Zoltán, PhD.
egyetemi adjunktus
SZTE, TTIK, Szoftverfejlesztés Tanszék
e-mail: alexin@inf.u-szeged.hu
<http://www.inf.u-szeged.hu/~alexin>

Tartalom

Mára lényegében megszűnt az egészségügyben a magánszféra és az orvosi titoktartás. Mi ennek az oka?

- Adatéhség (az adat az új olaj)
- Mindent előntő korrupció, személyre szabott törvények
- Hatalmi ambíciók (BM/27090/2023.)
- Inkompetencia. **Mitől lesz valami személyes adat?**

Minden nagy egészségügyi adatállomány veszélyeztetett, pl.:

- Nagy kórházi rendszerek
- Regiszterek
- NEAK elszámolási állományok
- EESZT

Köszönetnyilvánítás

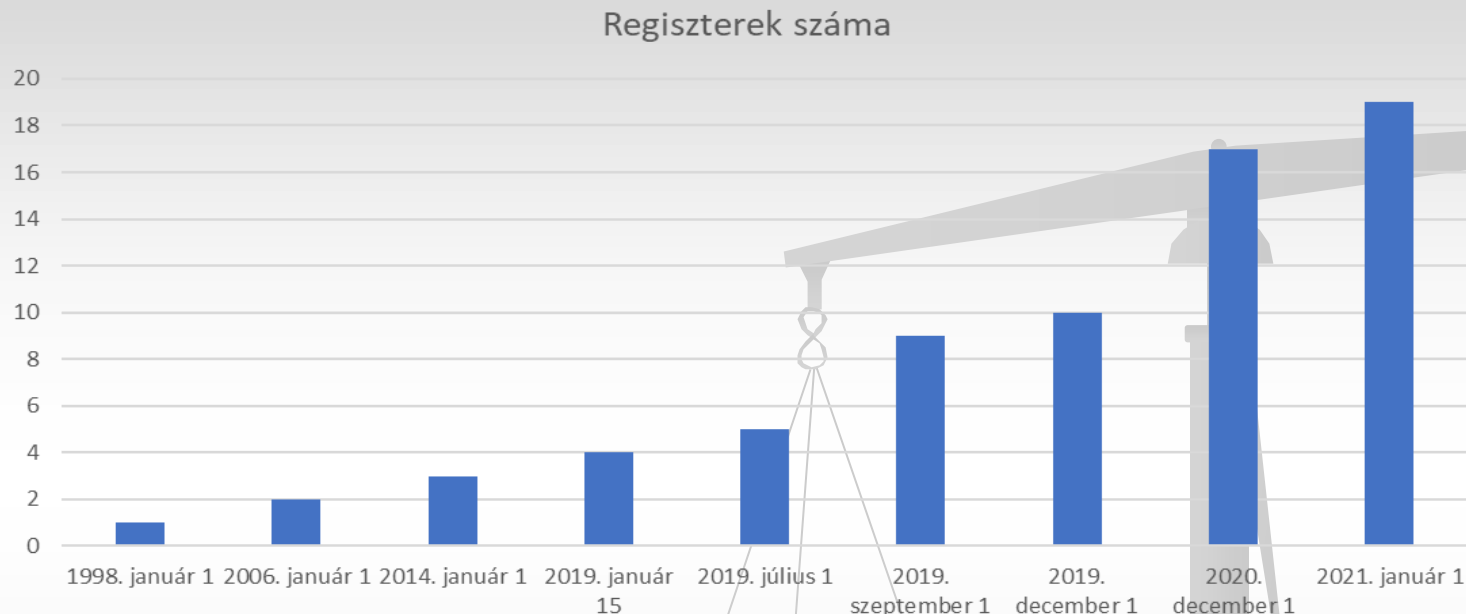
A szerző köszönetét fejezi ki:

- A Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatalának (KEKKH) az állami népszámlálási adatokból ingyenesen biztosított kutatási adatokért.
- Az Állami Egészségügyi Ellátó Központnak (ÁEEK) a járóbeteg ellátási adatokat tartalmazó adatállományból ingyenesen biztosított kutatási adatokért.

Nagy ellátó intézmények

- Kutatási felhasználások (K+F), az ipar hozzájut bármilyen egészségügyi adathoz, az érintettek tudta nélkül (Kheiron Kft. – 500 ezer mammográfia kép)
- A kutatás kötelező adatkezelés! Nincs lehetőség tiltakozásra.
- Amiben nincs név (esetleg TAJ) az anonim, pl. CT (arc) kép [1 mm felbontású], genetikai minta [3000 génre teljes szekvenálás] anonim
- Etikai engedélyek vég nélküli hosszabbítása (pl. 2001-es engedély)
- Etikai bizottságok kevés tudással (az ETT simán ad engedély az EESZT-ből történő adatszerzéshez)
- Az adatvédelmi tisztviselő (felelős) kihagyása a döntésekből
- Ingyen szoftvernek túrók a háta (nyitott DICOM port)
- Személyes kapcsolatokkal is lehet adatokhoz jutni
- Egyéni karrier, a betegek adataira alapozva

Betegségregiszterek száma



- A regiszterek tipikusan egy-egy neves akadémikushoz kötődnek, akiknek a hűségét a kényszerintézkedéssel gyűjtött személyes egészségügyi adatokkal fizeti ki az állam. Fő szempontok: **ne tudjanak az érintettek semmit az adatkezelésről, semmilyen önrendelkezésük ne legyen, a másik oldalon korlátlan szabadság az adatok felhasználásakor, végtelen adatmegőrzési idő.**

Társadalombiztosítás (NEAK)

- A Bőr- és Nemibeteg gondozók „beszántása” 2005
- A házi orvosok jelentési kötelezettsége 2009
- **Megőrzési idő 5 év -> 10 -> 15 -> 30 -> halál+10 év**
- A kényszerintézkedéssel gyűjtött adatok hozzáférhetővé tétele előbb csak a házi orvosok -> kezelő orvosok -> gyógyszerészek számára
- A nem támogatott ellátások adatainak gyűjtése (alkotmányellenesen és törvényellenesen), bármi áron
- A lombik-klinikák „beszántása” 2021
- Eüak. 23. § - a ma már végtelen ideig gyűjtött adatok továbbítása állami erőszakszervezetek felé.
- **Személyazonosításra alkalmatlan adat = nincs benne TAJ**

EESZT

- A kórházakban (rendelőintézetekben) tárolt személyes adatok megszerzése az állam számára
- Megőrzési idő vényeknél 5 -> 30 év, eseménynapló halál + 5 év -> halál + 10 év (egy nappal a határidő előtt)
- Az adatkezelési célok száma: 3 -> 4 -> 27 -> 28
- Nem támogatott ellátások adatainak kényszer gyűjtése
- Az adatok elérhetővé tétele hatósági orvosi ellátások során (NEAK ellenőr, Rehab. hatóság, üzemorvos, ...)
- Átláthatatlan működés, lehetőség van naplózás nélküli adatkezelésre (legalábbis az érintett nem tud róla)
- Hatalmas adatéhség (már semmi sem számít, sem törvény, sem etika) -> 70 millió PDF lementése EESZTGate

Mi számít személyes adatnak?



A data broker reckoning - Proto X Jon Keegan a Twitteren: "Yest X +

https://twitter.com/jonkeegan/status/152224406839410689/photo/1

Placer.ai

Property

Search property name

My-Zone Explore Property Chains Advanced Reports Labs Academy Marketplace

Planned Parenthood

Venues: Planned Parenthood / [redacted]

By: Home Location Metric: Visits Min. Visits: 1 Resolution: Nationwide Visualization: Gradient

of Visits High Low

Map Hybrid

Keyboard shortcuts Map data ©2022 200 m Terms of Use Report a map error

Jon Keegan @jonkeegan

Yesterday we also saw heatmaps showing Planned Parenthood visitors' home locations that in sparsely populated areas revealed single homes. Here (my redactions), the home location of a visitor is shown as a small blob, but the sat view shows one home in the center (my blur). 2/4

du. 4:39 · 2022. máj. 5. · Twitter Web App

3 Retweet 12 Kedvelés

Jon Ke... @jonk... · máj. 5. ...

Válasz @jonkeegan felhasználónak

We also noticed that Placer's "Visitor Journey - Routes" panel appeared to show the homes of Planned Parenthood visitors at the start of their trip to the clinic. Here (I blurred the homes), the route is highlighted and appears to originate at this last house on the street. 3/4

<https://twitter.com/jonkeegan/status/152224406839410689/photo/1>

Fájl Szerkesztés Nézet Előzmények Könyvjelzők Eszközök Súgó

Politikai és médiabotrány lett egy 15 éves fiú

https://444.hu/2023/03/06/politikai-es-mediabotrany-lett-egy-15-es-fiu-ongyilkossagabol-lengyelorszagban

Bevezetés Slack - GoodBrother Slack - NET4AGE Net4Age Core Group f... Más könyvjelzők

!!444!! hirdetés

Politikai és médiabotrány lett egy 15 éves fiú öngyilkosságából Lengyelországban

Miklósi Gábor
KÜLFÖLD - március 06., 13:00

f t in

Péntek este a legnagyobb lengyel ellenzéki párt, a konzervatív-liberális Polgári Platform (PO) szcczini képviselője a Twitteren bejelentette, hogy 15 éves fia február közepén öngyilkos lett. Egyben azt is kérte, hogy a média tartsa tiszteletben a magánéletüket, senki se tudósítson a március 7-i temetésről.

TWITTER MAGDALENA FILIKS

Magdalena Filiks
@FiliksMagdalena · Follow



W dniu 17 lutego odszedł mój syn
Mikołaj Filiks
Miki 8 marca skończyłby 16 lat...

<https://444.hu/2023/03/06/politikai-es-mediabotrany-lett-egy-15-es-fiu-ongyilkossagabol-lengyelorszagban>

Alverad-Bánki Nemzetközi Kiberbiztonsági Konferencia, Budapest, 2023. október 25.



Google and DeepMind face lawsuit over deal with Britain's National Health Service

Sam Shead
@SAM_L_SHEAD

WATCH LIVE

- DeepMind found itself in the spotlight in 2016 when the New Scientist reported that its collaboration with the U.K.'s National Health Service went beyond what was publicly announced.
- British law firm Mishcon de Reya told CNBC Friday it had filed a claim with the High Court on behalf of Andrew Prismall and roughly 1.6 million other individuals whose medical records were obtained by DeepMind.
- DeepMind and the Royal Free London NHS Foundation Trust signed a deal in 2015 that

<https://www.cnbc.com/2021/10/01/google-deepmind-face-lawsuit-over-data-deal-with-britains-nhs.html>

Kvázi-azonosítók

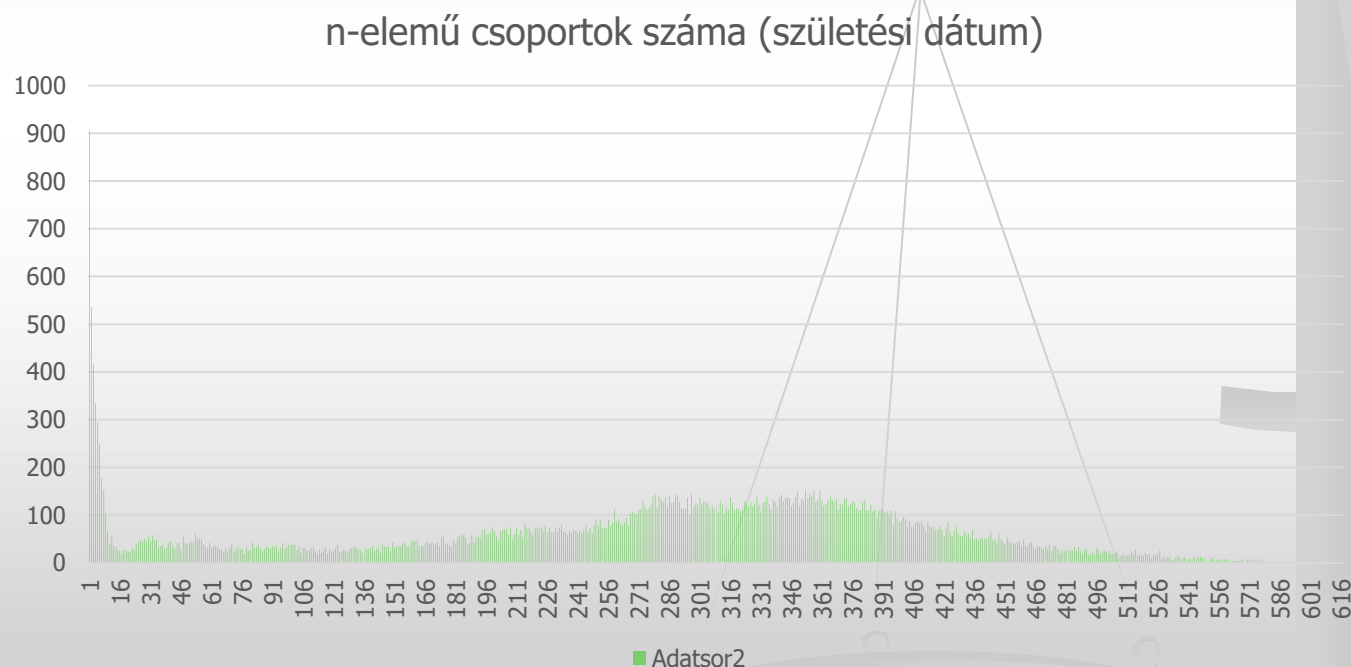
- Kvázi-azonosítók: olyan adatelemek, amelyek magukban nem elegendőek egy személy azonosításához, más kvázi-azonosítókkal kombinálva azonban már azonosíthatnak személyeket.
- A kvázi azonosítók **nem közvetlen személyazonosítók**. Ehelyett olyan adatok, amelyek például a földrajzi helyre utalnak, irányítószám, születési dátum. Sok ember lakóhelyének lehet ugyanaz az irányítószáma, és sok embernek lehet azonos a születési dátuma, azonban nagyon kevés esetben azonos mindkettő.
- Olyan típusú adatokról van szó, amelyeket egy ellenfél (adversary) megszerezhet a természetes személyazonosítókkal együtt mint például a név, anyja neve stb. és ezen a módon feltörheti az álnevesített adatállományt.
- Egy álnevesített adatállomány: $r(q_1, q_2, q_3, q_4, q_5, \dots, q_n, d_1, d_2, \dots, d_m)$.
- Egy ellenfélnek birtokában lehet: $a(q_2, q_3, q_4, q_5, \text{név})$.
- Mik lehetnek kvázi-azonosítók? Születési dátum, irányítószám, nem, munkakör, végzettség, iskolák, betegség, műtét, gyermekszületés, sőt labor eredmény
- Atreya, Ravi V, Joshua C Smith, Allison B McCoy, Bradley Malin and Randolph A Miller, "Reducing patient re-identification risk for laboratory results within research datasets," J Am Med Inform Assoc 2013;20:95–101. doi:10.1136/amiajnl-2012-001026.

„anonim” állományok feltörése

- Adatösszekapcsolás a kvázi-azonosítók segítségével
- Ha van egy olyan állomány, amelyben személyazonosító adatok pl. név szerepel (feltörő állomány) és néhány kvázi azonosító, akkor a kvázi-azonosítók egyeztetésével a két állomány összekapcsolható és azonnal személyes adattá válik.
- Arra kell felkészülni, hogy már **bármilyen kvázi-azonosítókhoz létezik feltörő állomány**, a feltörés ellen nem lehet védekezni.
- Sőt, **bármilyen lehet kvázi-azonosító**, amióta országos kényszer adatgyűjtések vannak.
- Csak az jelenthet védelmet, ha a kvázi-azonosítók, pontosabban a teljes adatállomány információtartalma egy-egy személyre kiszámítva kicsi.

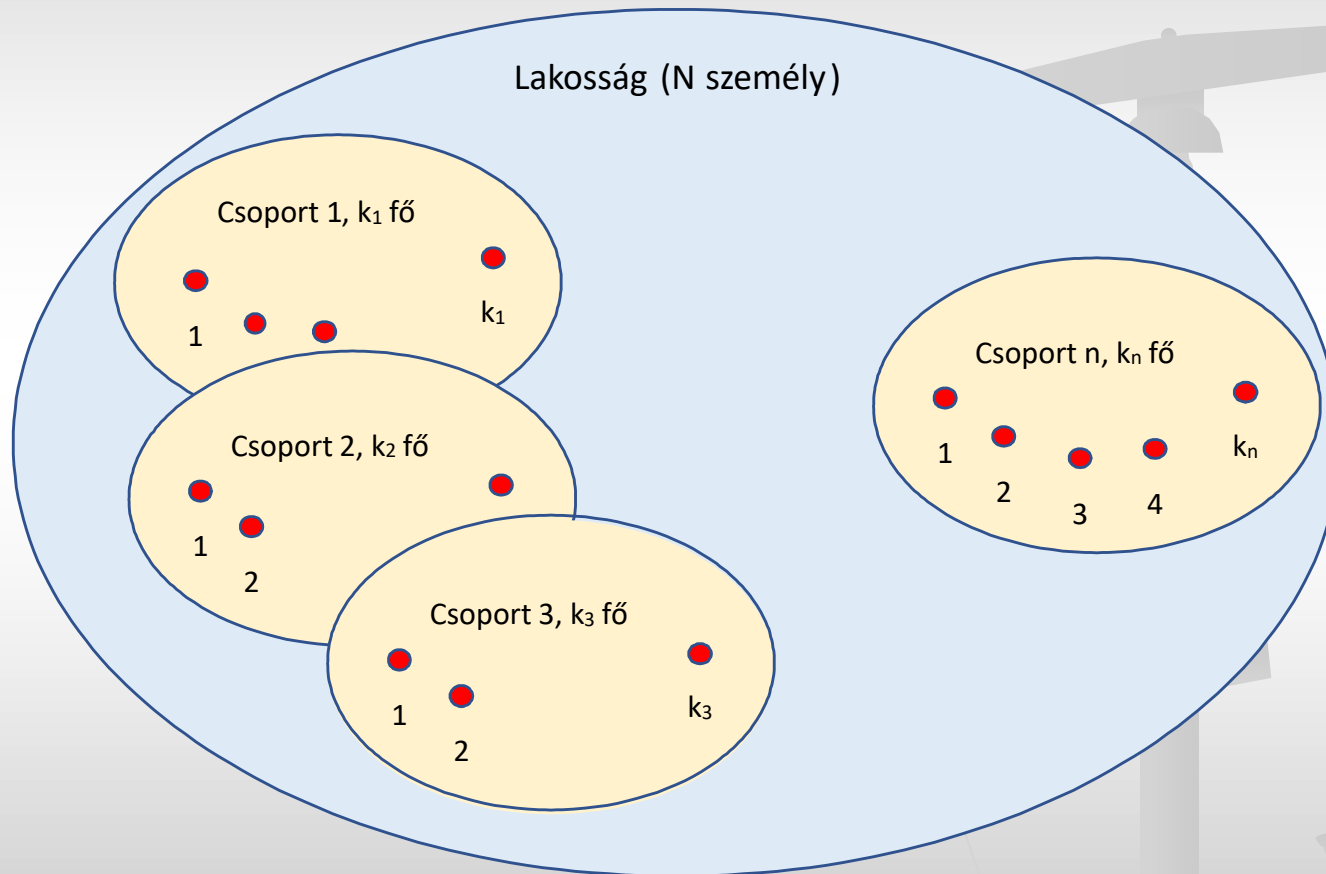
A születési dátum

- Tegyük fel, hogy nem tudjuk megkülönböztetni azokat a személyeket, akik ugyanazon a napon születtek, azaz más adatot nem tudunk róluk.
- Így is van Magyarországon 907 személy, aki egy-elemű csoportot alkot, tehát egyedül a születési dátuma alapján meghatározható.



1	907
2	536
3	418
4	335
5	296
6	249
7	178
8	153
9	103
10	66
...	0
635	1
636	1
640	1
644	1
708	1
718	1

Személyes csoportosítása kvázi-azonosítóik alapján



Mi az entrópia?

0110011101100110011 1010

Csak a csoportot ismerjük,
ahol a keresett személy
található.

$$\log_2(N) - \log_2(k) = \log_2(N/k)$$

k megkülönböztethetetlen
személy, $\log_2(k)$ bit

- Az entrópia egy súlyozott összeg (várható érték, átlag) azon információ mennyiségének átlaga, amelyet az egyes személyekről tudunk.

$$E(D) = \frac{1}{N} \sum_{\text{people}} \log_2 \frac{N}{\#group} = \sum_{\text{people}} -\frac{1}{N} \log_2 \frac{\#group}{N}$$

$$E(D) = \sum_{\text{group}} -\frac{\#group}{N} \log_2 \frac{\#group}{N}$$

Az entrópia tulajdonságai

Az entrópia kapcsolatban áll a k -anonimitással, és teljesül rá, hogy ha egy adatállomány k -anonim, akkor az entrópia: $E(D) < -\log_2(k/N)$. Általános esetben azonban az egyes csoportok elemszáma változó lehet, de ekkor is képezhető a képletből, átrendezéssel egy hozzávetőleges $\hat{k}$ érték.

$$\hat{k} = \frac{N}{2^{E(D)}}$$

Ha pedig az entrópia elég nagy, azaz $\log_2 N - 1$ -nél nagyobb, akkor feltételezve, hogy a csoportok csak egy és kételeműek, akkor az egyelemű csoportok számára van egy becslés. Ez a szám lehet nagyobb is, ha vannak három, négy, ... elemű csoportok is.

$$n_{\text{singletons}} \geq (E(D) - (\log_2(N) - 1)) * N$$

Az irányítószám entrópiája

Irányítószám	Település	Lakosság	Bitek ($\log_2(N/k)$)	Entrópia ($-(k/N) \cdot \log_2(k/N)$)
1011	Budapest I.	3286	11.5719	0.003800
1012	Budapest I.	4446	11.1357	0.004948
1013	Budapest I.	3404	11.5210	0.003920
...				
9982	Apátistvánfalva	589	14.0519	0.000827
9983	Szakonyfalu	769	13.6672	0.001050
9985	Felsőszölnök	589	14.0519	0.000827
Összesen:		10,004,090		10.303428

- A számítás eredménye szerint az entrópia az irányítószám átlagos információtartalma 10,3 bit. Ez azt jelenti, hogy egy véletlenül kiválasztott személy esetén, ha tudjuk a lakóhelye irányítószámát, akkor az átlagosan 10,3 bit információ. Ez 7916-anonimitásnak felel meg.
- A teljes lakosság egyértelmű azonosításához Magyarországon (a lakosság 10 004 090) $\log_2(N) = 23.254$ bitre van szükségünk.

A születési dátum + irányítószám

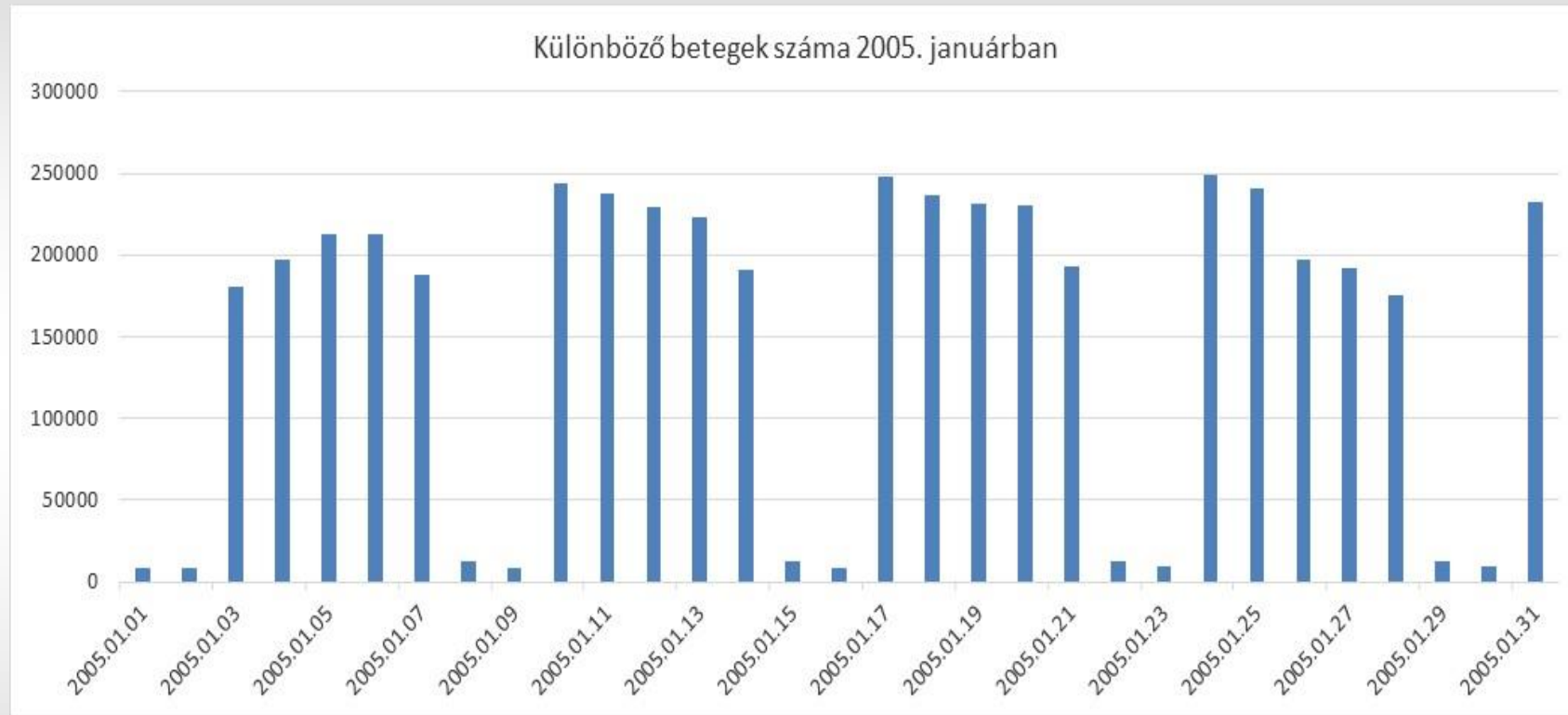
Születési dátum x irányítószám	Lakosság	Bitek ($\log_2(N/k)$)	Entrópia ($-(k/N) \cdot \log_2(k/N)$)
(1894.12.31., 3744)	1	23.254	2.324458e-6
...			
(1975.08.04., 9400)	4	21.254	8.498159e-6
(1975.08.04., 9407)	1	23.254	2.324458e-6
(1975.08.04., 9473)	1	23.254	2.324458e-6
(1975.08.04., 9523)	1	23.254	2.324458e-6
(1975.08.04., 9600)	1	23.254	2.324458e-6
(1975.08.04., 9700)	6	20.669	1.239640e-5
...			
Sum:	10,004,090		22.79385

- Az entrópia 22.7985 bit volt. Ez 1.37-anonimitásnak felel meg. Egy ilyen adatállomány jelentős kockázatot jelent az újra-azonosítás szempontjából.
- Az egy elemű csoportok előrejelzett száma: 54% volt, a valóságban azonban több fordult elő: 6 635 838 személy.

Járóbeteg vizit adatbázis

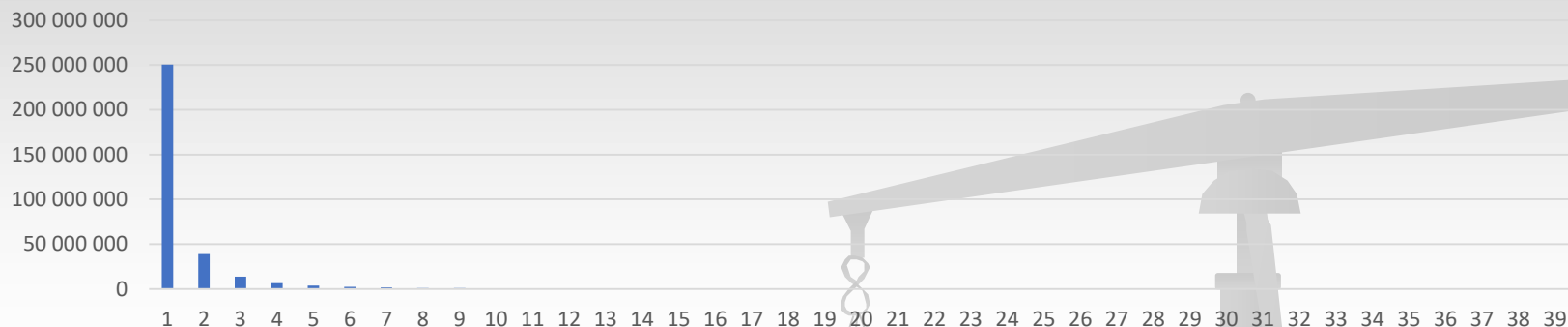
- Az összes NEAK-hoz jelentett járóbeteg ellátás adatát tartalmazza, összesen 721 633 881 vizit adatát. Többségében 2002. január 1. és 2014. december 31-e között.
- Az adatelemek:
 - Kapcsolati kód, amely lehetővé teszi ugyanazon beteg ellátásainak azonosítását;
 - A vizit dátuma;
 - A BNO-kód első betűje kétjegyű számmal kódolva (nem tudom, hogy melyik kód milyen betűt jelent);
 - A rendelőintézet irányítószámának első két számjegye.
 - "G58FG689DDQ3GE1"; "2005.01.27."; "3"; "67"
- Narayanan, A. and Shmatikov, V. (2008): *Robust de-anonymization of large sparse datasets*. In 2008 IEEE Symposium on Security and Privacy (sp 2008) (pp. 111-125). IEEE. (joining the Netflix research database with the IMBD.com)

Járóbeteg vizit adatbázis



Vizit-párok entrópiája

A csoportok méretének eloszlása 2005. január



- A legaktívabb betegek (akik ezerszer vagy többször jelentek meg) törlésre kerültek (1968 beteg, az összes beteg 0,016%-a, 3 348 081 vizit került így ki, amely a 721 millió ellátás 0,46%-a).
- A generált párok nagy méretére tekintettel, egy-egy hónapra eső párokat, először a 2005. januárra eső párokat határoztam meg, ahol a korábbi vizit dátuma eset 2005. januárra. Ez 3 863 139 808 (88.8 GB) párt jelentett.
- A vizit párokból álló csoportok mérete 1 és 3233 közé esett. A legnagyobb csoportban 3233 beteg volt, akik budapesti rendelőintézetekben jártak január 24-én és 26-án (hétfő, szerda). Összesen 539 225 013 különböző vizit pár volt, az egyelemű csoportok száma 250 588 888 volt.

Alexin, Z.: *Entropy based approach to personal data*, In Proceedings of the International Conference on Privacy-friendly and Trustworthy Technology for Society — COST Action CA19121, pp. 18-31. DOI: 10.5281/zenodo.6813377, 28th June 2022. Zagreb, Croatia (2022.)

- EESZTGate: Az EMK (Egészségügyi Menedzserképző Központ) már rendelkezik az EESZT-ből adatokkal [eseménynapló, vény, Covid oltás], amelyhez megkapta a NEAK és az NNK 10 évre visszanyúló idősoros adatait a teljes lakosságról. Semmilyen tájékoztatás nem volt, bár talán(?) a DÖR-t figyelembe vették.

Köszönöm a figyelmet!

