

Személyes adatokat tartalmaz-e a Tételes Egészségügyi Adattár?

Dr. Alexin Zoltán, PhD.

Szegedi Tudományegyetem, TTIK,
Szoftverfejlesztés Tanszék
H-6720 Szeged Árpád tér 2.

e-mail: alexin@inf.u-szeged.hu

<http://www.inf.u-szeged.hu/~alexin>

A Tételes Egészségügyi Adattár

- Az OEP által összegyűjtött elszámolási adatokból származik.
- A vények, járó- és fekvőbeteg adatok (+ CT-MR, művese, külön keretes),
- Az OEP negyedévente küldi az új adatokat TEA adattárba, amelyből egy-egy példány jelenleg a GYEMSZI-nél és az OTH-nál található,
- Az OEP kicseréli a TAJ azonosítókat egy pszeudo-TAJ azonosítóra, így az egy személyre vonatkozó ellátások adatai összekapcsolhatók. **Lényegében minden magyar állampolgár megtalálható az adattárban, sőt a halottak adatai is megmaradnak.**
- A megőrzési idő nincs meghatározva, élethosszig tart.
- **Tartalmazza a páciens lakóhelyének az irányítószámát, a páciens nemét és születési dátumát is.**
- Számtalan más kvázi-azonosító is van benne: dátumok, ellátás kódok, intézmények, orvos pecsétkódok, *pro familia* jelzés, betegségek
- A nem támogatott vények adatai is szerepelnek benne (legalábbis 2009-ig).
- Nincs független adatvédelmi és orvosi etikai felügyelet.

Előzmények

- A TEA (Tételes Egészségügyi Adattár) létrehozása 2004-ben a 76/2004. (VIII. 19.) számú ESzCsM rendelettel
- Az adatvédelmi biztos 1301/A/2006-9. számú állásfoglalása (négy egészségügyi adatvédelmi problémára hívta fel a figyelmet – de egyiket sem fogadta meg a tárca)
- Az Alkotmánybírósághoz benyújtott 937/B/2006 indítványom ügyében hozott elutasító határozat (az OEP csak személyazonosításra alkalmatlan adatokat továbbíthat – vizsgálat nélkül hozta meg az AB)
- Valóban anonim-e? Tényleg személyazonosításra alkalmatlan adatokat tartalmaz-e?

Mit jelent az anonimizálás?

- Az anonim a görög *ἀνωνυμία* (anonymia) szóból ered, jelentése név nélkül.
- A cél, hogy az érintett személyisége rejtve maradjon, ne fedje fel a kilétét.
- Az infokommunikáció korában az **anonimizálás** azt jelenti, hogy minden olyan adatot el kell távolítani, amely ahhoz vezethet, hogy az érintett azonosítani lehessen.
- A személynév eltávolítása nem elegendő.
- Az emberek azonosíthatók nem csak a nevük alapján, hanem pl. munkahely, munkakör, munkahelyi vezető; vagy pontos lakóhelyük; vagy születési dátum, iskola, osztályfőnök; kórház, orvos, beavatkozás, dátum adatok alapján is.

Azonosító	Lakóhely	Születési dátum	Betegség
10784343	Kiskörös	1965. május 3.	HIV+
13453453	Mohács	1946. június 2.	rák
53353534	Bonyhád	1964. augusztus 17.	szifilisz+

<http://www.parlament.hu>

Név	Bizottsági tagság	Születési dátum	Lakóhely
Kis Péter	gazdasági	1964. augusztus 17.	Bonyhád
Nagy János	-	1965. május 3.	Kiskörös
Tóth Viktor	mentelmi	1946. június 2.	Mohács

A gondatlan anonimizálás veszélyei

- Az anonimizálás káros hatásait nem lehet meg nem történné tenni.
- A már közzé tett adatokat nem lehet visszavonni.
- Olyan károkat okoz, amelyeket nem lehet jóvátenni, orvosolni.
- Egy jövőbeli kockázat (az újra azonosítás kockázata folyamatosan fenyegeti az érintetteket).
- **Nem tekinthető tisztességes adatkezelésnek.**

Előnyök és hátrányok

■ Előnyök:

- Az adatok feldolgozhatók, anélkül, hogy az érintetteket sértenék
- Az adatvédelmi törvényt nem kell alkalmazni
- Nem merülnek fel etikai kérdések
- Az adatok megoszthatók, eladhatók

■ Kétségek:

- Az emberek és cégek (munkahelyek) egyre több információt tesznek fel magukról az Internetre
- Nem tudjuk megjósolni a jövőt. Legközelebb milyen információt fognak nyilvánosságra hozni.
- Egyes cégek leszűretelik a nyilvános információkat a webről (neveket, fényképeket, születési dátumokat, lakóhelyet, iskolákat stb.)
- Ez megteremti az iparszerű újraazonosítás lehetőségét.

Statisztikai adatbázis a kockázat elemzéshez

- Az állami népességnyilvántartásból származó statisztikai adatok.
- Tartalmazza az irányítószámot, nemet, születési dátumot minden magyar lakóhellyel rendelkező állampolgárról (10 004 090 fő).
- P-ikrek (pseudo ikrek): olyan személyek, akik ugyanabban az irányítószám körzetben laknak, azonos neműek, és azonos napon születettek. Ha egyéb adat nem áll rendelkezésre, akkor megkülönböztethetetlenek.
- A legnagyobb klón 11-ikrekből áll (1 klón, 1975), majd további 12 klón tartalmaz 10-ikert, stb.

1011;1989.01.23.;N;2

1011;1989.02.01.;N;1

1011;1989.03.11.;N;1

(8 million lines)

...

Falvak és városok

Lakosság	Irányítószám körzetek száma	Teljes lakosság	P ₁	P ₂
n < 1000	1339	725628	98,218%	99,973%
1000 ≤ n < 5000	1296	2800312	94,798%	99,811%
5000 ≤ n < 20 000	402	3883348	82,026%	97,839%
20 000 ≤ n	73	2594802	49,838%	80,315%
Összesen:	3110	10004090	78,426%	94,001%

Egy személy akkor egyértelműen azonosítható, ha nincs P-ikertestvére.

P_1 = az azonosítás valószínűsége = A P-egykek száma osztva az összes személy számával.

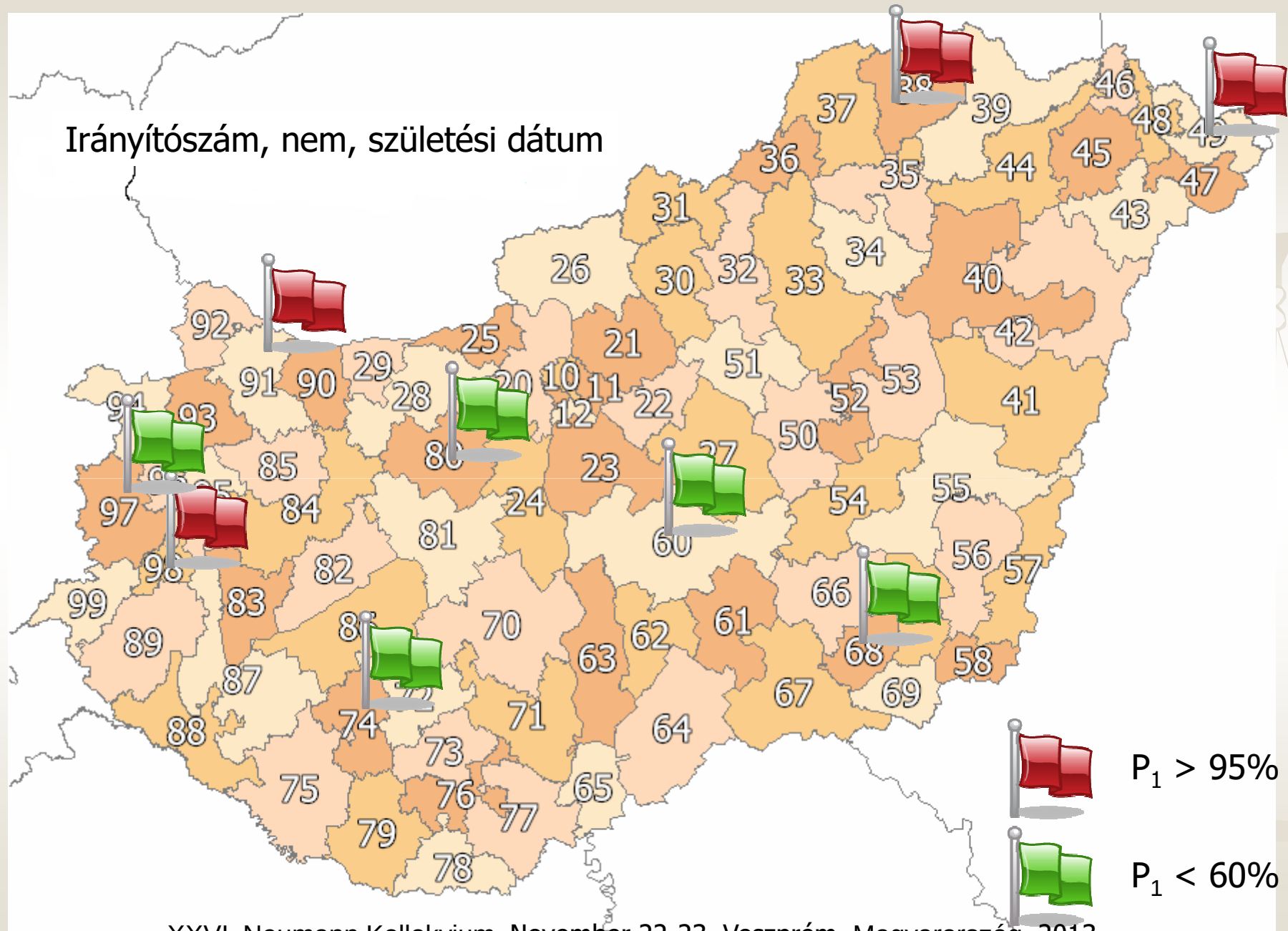
Ha két személy közül mindig ki tudjuk számítani azt az egyet, amelyet azonosítani akarunk (egyéb információ alapján).

P_2 = (a P-egykek száma + P-kettes ikrek száma) / összes személy száma.

P-ikrek eloszlása

ZIP	Régió	1	2	3	4	P ₁	P ₂
1xxx	Budapest	1311381	147157	16027	1815	78,902%	96,610%
2xxx	Middle	1364579	154293	27018	5560	76,460%	93,751%
3xxx	N-East	978924	69693	9255	1555	84,835%	96,915%
4xxx	East	897630	80463	13854	3622	79,942%	94,274%
5xxx	M-East	589923	63741	11594	2604	76,957%	93,588%
6xxx	S-East	690776	83418	18849	5607	72,585%	90,116%
7xxx	S-West	686907	53665	8645	1795	82,811%	95,750%
8xxx	M-West	780474	75089	19937	6205	75,736%	90,309%
9xxx	West	545256	51508	11789	3142	77,632%	92,300%
Sum:		7845850	779027	136968	31905	78,4264%	94,001%

Irányítószám, nem, születési dátum



XXVI. Neumann Kollokvium, November 22-23, Veszprém, Magyarország, 2013.

Az azonosítás kockázatának csökkentése

Általánosítás	1-különböző	5-különböző (pesszimista)	5-különböző (realista)
3-jegyű irányítószám	57.86%	98.71%	74.88%
2-jegyű irányítószám	14.81%	71.29%	34.54%
Születési év, hónap	15.0%	50.96%	27.65%
Születési év	0.59%	6.24%	2.29%
3-jegyű irányítószám, születési év, hónap	1.85%	18.18%	6.83%
3-jegyű irányítószám, születési év	0.037%	0.27%	0.11%
2-jegyű irányítószám, születési év, hónap	0.056%	0.39%	0.15%
2-jegyű irányítószám, születési év	0.0037%	0.022%	0.0096%
HIPAA Privacy rule	1 személy	7 személy	3 személy

Összefoglalás

- Egyes személyeket a törvény kötelez arra, hogy közzé tegyék életrajzukat és vagyonnyilatkozatukat. Ebben szerepel a születési dátum, lakóhely (tudósok, politikusok).
- Az üvegzséb törvény előírja a parlamenti képviselők számára, hogy tegyék közzé életrajzukat és vagyonnyilatkozatukat.
- Híres színészek, miniszterek lakóhelye (csak a város) sokszor elhangzik egy TV műsorban.
- Budapest kis körzetekre van osztva, úgy viselkedik mint egy nagyobb falu vagy kisváros.
- Orbán Viktor születésnapja, lakóhelyének irányítószáma az interneten megtalálható – egy iskolás gyerek is megtalálja 2-3 percen belül – megszerezhető a teljes egészségügyi élettörténete.
- **Az adatkezelők is azonosítani akarják az egyes pácienseket (erről a jogukról egyáltalán nem akarnak lemondani).**

Lehetséges megoldások

- Az adatállomány 2-3 éves időtartamra korlátozása (élettartam meghatározása) és
- Etikai felügyelet alá helyezés teljes nyilvánosság mellett (lásd Helsinki Nyilatkozat: 2013. október 19.) és
- A kiszivárgott adatok visszavonása, törlése és
- **Hatékony anonimizálás!**



Köszönöm a figyelmet!

A nem, a születési dátum és a lakóhely irányítószáma

Philip Golle képlete:

$$f_n(i) = \binom{n}{i} \cdot b^{1-n} \cdot (b-1)^{n-i}$$

- $f_n(i)$ – azoknak a napoknak a száma, amelyeken pontosan i személynek van születésnapja. Az i egy paraméter.
 - n – a lakosok száma egy irányítószám körzetben
 - b – a napok száma az adott évben
- Akkor is meghatározható, ha csak népszámlálási adatok vannak, pl $f_{113}(1)$ – azoknak a napoknak a száma, amelyeken pontosan 1 személynek van a születésnapja.

HIPAA Privacy Rule (USA)

- P. Golle eredményei nyomán:
 - Anonim egészségügyi adatállományból el kell távolítani legalább a HIPAA törvényben felsorolt 18 adatkategóriát és általánosítani kell az irányítószámokat és a dátumokat.
 - A módszer csak az 5-jegyű irányítószám első három jegyét engedi megtartani. Ha 20 ezernél kevesebben laknak ott, akkor egy 000 kategóriában kell egyesíteni az adatokat.
 - Minden dátumot, amely egy érintetthez vonatkozik törölni kell, kivéve az évet, pl. születési dátum, felvételi dátum, elbocsátási dátum, halálozás napja, stb.
 - Minden 89 feletti életkorra utaló esetben az évet is törölni kell, vagy egyetlen „90 vagy feletti” kategóriában egyesíteni az életkort.
 - Törölni kell a: név, cím, bármilyen személyes azonosítószám, e-mail, telefonszám, fax szám, berendezések típusa, gyári száma, gépkocsi rendszám, arcot ábrázoló fénykép, biometrikus azonosító, stb.

ZIP	Lakosság	1-twins	2-twins	3-twins	4-twins	5-twins	6-twins
6500	32660	18705	4997	1072	155	25	
4060	17795	12945	2065	213	19	1	
6237	8829	7473	613	38	4		
6635	4699	4331	181	2			
8248	2969	2792	84	3			
8096	1306	1272	17				
7381	817	807	5				

A népszégnylvántartás adatai

ZIP	Lakosság	1-twins	2-twins	3-twins	4-twins	5-twins	6-twins
6500	32660	19039.98	5029.45	957.79	143.76	17.95	1.93
4060	17795	13240.45	1936.18	202.29	16.57	1.12	
6237	8829	7629.82	554.96	28.51	1.14	0.04	
6635	4699	4347.56	166.39	4.43	0.09		
8248	2969	2828.17	67.25	1.09			
8096	1306	1271.14	12.32	0.08			
7381	817	806.98	4.49				

Golle képletével számított adatok