HADMÉRNÖK

VIII. Évfolyam 1. szám - 2013. március

Fleiner Rita
fleiner.rita@nik.uni-obuda.hu

KÖZIGAZGATÁSI ADATBÁZISOK ÖSSZEKAPCSOLÁSÁNAK TECHNIKAI MEGVALÓSÍTÁSA

Absztrakt

Az államigazgatási szerveknek kötelessége a gazdaságra és a társadalmi folyamatokra kiható döntéseiket előzetesen mérlegelni és eredményeiket utólagosan vizsgálni. A költségvetési szervek által létrehozott adatbázisok nem csak adminisztratív célokra, hanem elemzések, hatásvizsgálatok és döntés előkészítés céljára is alkalmazandók. A vizsgálatokhoz szükség lehet a különböző adatbázisokból származó adatok együttes kezelésére, az adatbázisok összekapcsolására. A publikációban a közigazgatási adatbázisok összekapcsolásának technikai megvalósítását vizsgáljuk. Ennek keretében elemezzük az összekapcsolás alapját képező anonim kapcsolati kód fogalmát és képzési módját, feltárjuk az adatbázisok összekapcsolásának biztonsági elvárásait, követelményeit és bemutatjuk az összekapcsolás folyamatát.

It is the duty of the government to consider their decisions influencing the economy and the social processes beforehand and to examine their results posteriorly. Databases created by government agencies can be used not only for administrative purposes, but also for analyses, impact assessments and decision making processes. These examinations may need joint data from different databases, so the connection of databases must be performed. In the publication we examine the technical implementation aspects of connecting administrative databases. In this framework we analyse the concept and construction of the anonym connection code providing the basis of the connection, we reveal the security requirements of the database connection, and we present the steps of the connection process.

Kulcsszavak: *adatbázis, adatbázis összekapcsolás, elektronikus közigazgatás, biztonsági elvárás, hash függvény ~ database, connecting databases, electronic government, security requirement, hash function*

BEVEZETÉS

A közigazgatás elektronikus nyilvántartásai az elsődleges adminisztratív célok mellett adat alapú döntéshozás, statisztikai számítások, társadalomkutatási feladatok számára is felhasználhatók. *Közigazgatási adatbázison* a továbbiakban olyan, a magyar közigazgatásban fellelhető adatbázist értünk, melyet valamely államigazgatási szerv a saját feladatai ellátása során hozott létre. A közigazgatási adatvagyon felhasználásának előmozdítása érdekében született meg az Európai Parlament és a Tanács 2003/98/EK számú irányelve [1]. Az irányelv minimum szabályokat állapít meg a tagállamok közigazgatási szervei birtokában lévő dokumentumok (ideértve az adatokat is) további felhasználására.

Magyarországon az irányelv céljainak elérését a döntés előkészítéshez szükséges adatok hozzáférhetőségének biztosításáról szóló 2007. évi CI. törvény [2] teszi lehetővé. A törvény kimondja, hogy az államnak és az államigazgatási szerveknek kötelessége, hogy a gazdaságra és a társadalmi folyamatokra kiható döntéseiket előzetesen mérlegeljék, eredményüket utólagosan vizsgálják, és ehhez adminisztratív céllal gyűjtött adatokat is felhasználjanak. Ez a törvény szabályozza a közigazgatási adatok összekapcsolásának lehetőségeit is. A szabályozásra az európai uniós irányelv mellett jelentősen hatottak a hazai adatvédelmi előírások, melyek európai szinten kifejezetten szigorúnak számítanak.

A kormányzati döntés-előkészítéshez és hatásvizsgálatához szükség lehet természetes személyekre vagy gazdasági szervezetekre vonatkozó mikroszintű adatokra. A mikroszintű adatok kezelésének feltételeit különböző jogszabályok - méghozzá az adatvédelmi törvény [3], az adózás rendjéről szóló törvény [4] és a statisztikáról szóló törvény [5] - írják le. A személyes adatokat is tartalmazó adatbázisok közérdekű, vagy más jogszerű célra való felhasználásának jogi feltétele az *adatok személyességének megszüntetése*, az ún. *anonimizálás*. Az anonimizálás egy olyan eljárás, melynek során az egyes személyeket beazonosító adatokat törölni vagy visszafordíthatatlan módon módosítani kell úgy, hogy a tárolt adatok az érintettel a továbbiakban ne legyenek kapcsolatba hozhatók.

Az elemzések során felmerülhetnek olyan kérdések, melyek megválaszolásához szükség lehet több adatbázis összetartozó sorainak összekapcsolására. *Adatbázisok összekapcsolása* technikai értelemben azt jelenti, hogy különböző adatbázisok adataiból – fizikailag, vagy virtuálisan – egy új adatbázist hozunk létre. Relációs adatbázisok esetében ez különböző adattáblák sorainak összekapcsolását jelenti, amely megfelelő oszlopokban szereplő adatértékek egyezésére épül. Az adatbáziskezelés fogalomrendszerében ez ún. elsődleges kulcsok és idegen kulcsok segítségével történhet.

A közigazgatási adatbázisok összekapcsolásának vizsgálatát a [6] publikációban leírtakra alapozva végezzük el. A publikáció

- elemzi az összekapcsolás alapját képező anonim kapcsolati kód fogalmát és képzési módját;
- feltárja az adatbázisok összekapcsolásának biztonsági elvárásait, követelményeit;
- bemutatja az összekapcsolás folyamatát.

ANONIM KAPCSOLATI KÓD FOGALMA ÉS KÉPZÉSE

A kutatásokban, elemzésekben szükség lehet a különböző adatbázisok sorainak összekapcsolására, ami az adatbázis azonosítók (egyedi kulcsok) egymáshoz rendelésén keresztül valósítható meg. Ma Magyarországon az adatvédelmi törvény nem teszi lehetővé a különböző nyilvántartásokban található személyes adatok összekapcsolását. Személyes adatnak minősülnek a természetes személlyel kapcsolatba hozható adatok, mint például név, születési év, lakcím, életkor. Külön törvény [7] véd egyes mesterséges személyes azonosítókat, mint adószám, TAJ szám, személyi szám, ezeket még korlátozottabban lehet

kiadni, összekapcsolásra felhasználni. Az adatvédelmi törvény szerint személyes adatokat akkor lehet kezelni, továbbadni, összekapcsolni, ha ehhez az érintett hozzájárul, vagy törvény megengedi.

Az univerzális azonosításra alkalmas személyazonosító jel használatát Magyarországon az Alkotmánybíróság 1991-es döntése [8] a személyes adatok védelme érdekében megtiltotta. Az univerzális személyazonosító helyett ma Magyarországon természetes személyazonosítóként a (név, anyja neve, születési hely és idő) hármast, illetve a személyiadat- és lakcímnyilvántartásban a személyazonosító jelet, az adózással kapcsolatos nyilvántartásokban az adóazonosítót, a társadalombiztosítási rendszerben pedig a társadalombiztosítási azonosító jelet (TAJ) használjuk. Ez utóbbi három mesterséges azonosítót a személyazonosító jel helyébe lépő azonosítási módokról és az azonosító kódok használatáról szóló 1996. évi XX. törvény [7] hozta létre, továbbá rendelkezett arról, hogy ezek az azonosítók személyes adatok, valamint arról, hogy mely szervek és hogyan használhatják ezeket az azonosítókat.

Az adatbázisok összekapcsolásának folyamata kapcsán adatkérő szervnek nevezzük azt, aki az adatok átadását igényeli, vagyis aki számára az összekapcsolt adatok átadásra kerülnek. Az adatkezelő szerv vezeti az adott nyilvántartást, adatbázist. Az adatbázis létrehozásáért felelős szerv végzi el a különböző adatkezelőktől megkapott adatok összekapcsolását és adja át az összekapcsolt adatokat az adatkérő szerv számára.

Magyarországon a kormány az adatbázis létrehozásáért felelős szervként a Neumann János Digitális Könyvtár és Multimédia Központ Kht.-t (továbbiakban: NKHT) jelölte ki. A különböző adatbázisok összetartozó sorainak összekapcsolására a 2007. évi CI. törvény az anonim kapcsolati kód használatát írja elő. A törvény szerint *anonim kapcsolati kódnak* nevezzük az ugyanazon személyre vonatkozó személyazonosító adatokból véletlenszerű elemet is tartalmazó módszerrel képzett karaktersort, amellyel ugyanazokból az adatokból mindig ugyanaz a karaktersor jön létre, de amely eredményeképpen létrejött karaktersorból a személyazonosító adatok nem állíthatók helyre. Az adatok összekapcsolását az adatbázis létrehozásáért felelős szerv végzi úgy, hogy a különböző adatkezelőktől azonos módszerrel képzett anonim kapcsolati kóddal veszi át és kapcsolja össze a különböző forrású adatokat. A törvény előírja, hogy az adatbázis létrehozásáért felelős szerv feladata az anonim kapcsolati kód képzésének módszerének meghatározása. A törvény szerint a kódképzés módszerének a következő kriptográfiai biztonsággal is kapcsolatos tulajdonságokkal kell bírnia:

- olyan személyazonosító adatok alapján kell képezni, melyek kezelésére valamennyi érintett adatkezelő jogosult,
- a kódképzés alapját nem képezhetik olyan személyazonosító adatok, amelyek kezelésére az adatkérő vagy az adatbázis létrehozásáért felelős szerv jogosult,
- az adatkezelőnek a kezelésében lévő személyes adatról, adótitokról, vagy egyedi statisztikai adatról készített másolatot visszafordíthatatlan módon kell módosítania úgy, hogy az az érintettel a továbbiakban ne legyen kapcsolatba hozható,
- a kódképzés konkrét módszere tartalmazzon egyedi, véletlenszerűen megállapított elemet,
- egy adott adatkérés esetén ugyanazokból az adatokból mindig ugyanaz a karaktersor jöjjön létre,
- az összekapcsolást anonim kapcsolati kód segítségével kell végrehajtani.

Az anonim kapcsolati kód képzésének alapjául az NKHT hash függvény használatát választotta. A hash függvények fő tulajdonsága, hogy tetszőleges hosszúságú bemenetet (bitsorozatot) egy véges hosszúságú kimenetre képeznek le. Hash függvény esetén a bemenetet gyakran üzenetnek, a kimenetet pedig lenyomatnak vagy hash értéknek nevezzük. A leképezés több-az-egyhez típusú, azaz egy hash értéket sokféle lehetséges bemenet előállíthat.

Hash függvény definíciója a következő: Legyen X a lehetséges bemenetek halmaza, Y pedig a lehetséges kimenetek véges halmaza és teljesüljön rájuk, hogy $|X| \geq |Y|$. Ekkor a $h:X \rightarrow Y$ függvényt hash függvénynek nevezzük.

A hash függvények szoros kapcsolatban állnak az egyirányú függvényekkel. Egy $f: X \rightarrow Y$ függvény egyirányú, ha a minden lehetséges bemenethez tartozó függvényérték könnyen kiszámítható (legfeljebb polinom idejű algoritmussal), azonban a kimenetek inverzeit nehéz kiszámítani (legalább exponenciális idejű algoritmust igényel). Egy hash függvénnyel szemben a következő biztonsági kritériumokat szokták megfogalmazni. Nagyon nehéz (gyakorlatilag lehetetlen):

- adott h(m) hash értékhez egy megfelelő m bemenetet találni (egyirányúság, őskép-ellenállás),
- adott m-hez olyan m'-t találni, hogy h(m) = h(m') (gyenge ütközésmentesség, 2. őskép-ellenállás),
- két olyan tetszőleges m-et és m'-t találni, hogy h(m) = h(m') (erős ütközésmentesség),
- illetve a bemenet bármely bitjének megváltozása a kimenet bitjeinek mindig átlagosan felét változtassa meg (lavinatulajdonság).

A hash függvények fontos alkotórészei különböző kriptográfiai probléma megoldásának. A kriptográfiai hash függvények egyik elsődleges felhasználási területe az adatintegritás biztosítása. Tetszőleges hosszúságú adatról fix hosszúságú lenyomatot, ellenőrző összeget készítenek hash függvénnyel és azt tárolják el. Később az adat változatlanságáról úgy győződhetnek meg, hogy ismét elkészítik az adat lenyomatát, és ha ez nem egyezik az eredetivel, akkor biztos, hogy változás történt. A kriptográfiai hash függvények fontos szerepet játszanak a digitális aláírás folyamatában. A digitális aláíró algoritmusok először lenyomatot készítenek az aláírandó adatról, és csak a lenyomatot írják alá. Szinten gyakori, hogy a jelszavas azonosítást használó rendszerek nem a jelszavakat tároljak, hanem a jelszavak lenyomatait és a jelszavak ellenőrzése is a lenyomatok alapján történik.

A kapcsolati kód képzéséhez ki kell jelölni azon adatbázis attribútumokat, melyek alapján történik a lenyomat képzése, másrészt meg kell adni a lenyomatképző algoritmust, amely egy, az adott adatkérésre vélelenszerű elemet is felhasznál. Az anonim kapcsolati kód képzésére az NKHT a Kripto Kft. által készített és kereskedelmi forgalomba hozott, Codefish nevű hash függvényre [9] alapuló algoritmust választotta. A Codefish hash függvény kriptoanalízisét vizsgálták tudományos kutatások [10], [11], melyek kimutatták, hogy a függvény gyenge és erős ütközésmentesség tulajdonságai támadhatóak, a függvény ezeket biztonsági kritériumokat nem teljesíti.

ADATBÁZISOK ÖSSZEKAPCSOLÁSÁNAK BIZTONSÁGI ELVÁRÁSAI

A következőkben a törvényből és a kormányrendeletből levezethető, az összekapcsolás rendszerére vonatkozó biztonsági elvárásokat vizsgáljuk. A biztonsági elvárások pontos megfogalmazása segíti a megvalósítandó rendszer tervezését, érékelését és ellenőrzését. Mivel az összekapcsolást informatikai rendszer valósítja meg, továbbiakban az összekapcsolás informatikai rendszerének biztonsági elvárásait keressük.

A megvalósított feladat alapján az adatbázisok összekapcsolását végző informatikai rendszert fokozott biztonsági kategóriába sorolták be [12]. A [13] dokumentumban ismertetett biztonsági kategorizálás módszere szerint az informatikai rendszer által kezelt adatok sérülésének következményét alacsony, fokozott és kiemelt osztályokba sorolják be. Az adatbázisok összekapcsolását végző rendszer esetében az adatok bizalmasságának és sértetlenségének sérülését fokozott szintbe sorolták, míg a rendelkezésre állását alacsony kategóriába. A biztonsági kategóriát a három biztonsági célra (bizalmasság, sértetlenség,

rendelkezésre állás) kapott legnagyobb szint határozza meg, ez pedig az adatbázisok összekapcsolása esetén a fokozott.

A fokozott biztonsági kategória meghatározza a rendszer által felvállalt általános informatikai biztonsági követelményeket. Ezek vonatkoznak például az azonosítás és hitelesítés, hozzáférés ellenőrzés, rendszer és kommunikáció védelem, naplózás, konfiguráció kezelés folyamataira. Azonban az összekapcsolást szabályozó törvényből és kormányrendeletből levezethetőek a rendszerre vonatkozó speciális biztonsági elvárások is. Ezeket az elvárásokat meg kell fogalmazni az informatikai biztonság fogalomrendszerével, ami a rendszer helyes megvalósítása, értékelése és ellenőrzése kapcsán elengedhetetlen. Ez konkrétan azt jelenti, hogy meg kell határozni a rendszer specifikus biztonsági célokat és ezek teljesüléséhez szükséges biztonsági követelményeket.

A törvényből és a kormányrendeletből egyértelműen levezethető a következő két speciális biztonsági cél:

1. Szükséges az összekapcsolt adatok alanyának védelme. Biztosítani kell, hogy az összekapcsolt adatokból ne tudjon senki következtetni az adatok alanyára (sem kívülálló, sem adatkérő, sem adatkezelő, sem az adatbázis létrehozásáért felelős szerv).
2. Szükséges a különböző összekapcsolások ismételt összekapcsolásának megakadályozása. Ez azt jelenti, hogy a különböző adat összekapcsolásokból származó eredményeket ismételten ne tudja összekapcsolni senki. Ez abból következik, hogy adatbázis összekapcsolást csakis konkrét, előre definiált célhoz kötötten lehet végrehajtani, amit miniszter vagy kormányhivatal vezetője rendelhet el, és ennek kapcsán az átadandó adatok körét és az átadás konkrét célját meg kell határozni.

A törvény előírja, hogy az összekapcsolást anonim kapcsolati kód képzésén keresztül kell megvalósítani. Az anonim kapcsolati kód képzését úgy kell meghatározni, hogy ugyanazokból az adatokból mindig ugyanannak a kimenetnek, azaz karaktersornak kell előállnia (egy adott lekérdezés esetén a különböző adatbázisok), de a létrejött karaktersorból a személyazonosító adatok ne legyenek előállíthatók. Ez a követelmény a kriptográfiai algoritmusok közül az egyirányú függvények csoportjára jellemző. Az adatbázis létrehozásáért felelős szerv a kódképző algoritmus alapjául hash függvényt választott. A következőkben azokat a kriptográfiai tulajdonságokat (követelményeket) soroljuk fel, melyekkel a kódképző szolgáló algoritmusnak, azon belül hash függvénynek rendelkeznie kell ahhoz, hogy a törvény előírásai teljesüljenek.

- A hash függvénynek egyirányúnak kell lennie
- E tulajdonság teljesülése szükséges ahhoz, hogy a kommunikációban átadott hash értékekből a védendő információk ne legyenek visszaállíthatóak, az anonimitás biztosítható legyen.
- A hash függvénynek rendelkeznie kell lavina-hatással
- E tulajdonság teljesítésülése szükséges ahhoz, hogy néhány karakter, vagy bit becslése nem vigyen közelebb a védendő információhoz. Azaz egy bit eltérés az inputban teljesen véletlenszerű változást okozzon a hash képben. Szintén ez a tulajdonság szükséges, hogy közeli bemenő adatok (például csak kevés karakterben eltérő adószámok) ne legyenek felismerhetők a hash értékek sokaságában.
- A hash függvénynek erősen ütközésmentesnek kell lenni
- E tulajdonság teljesítése szükséges ahhoz, hogy ne lehessen reális időben találni két azonos védendő adatot, melynek ugyanaz a hash képe. Az ütközés (két bemenő adat hash képének egyezése) zavart okozhatna a különböző adatbázisokból származó adatok párosításában.
- Rövid bemenő adatsor biztonsága

- A hash függvény előzőekben felsorolt kriptográfiai tulajdonságainak teljesülnie kell a feladatkörből adódó rövid bemenő adatsorra is, mivel Magyarországon a statisztikai minta kiindulási alapja eléggé kicsi. Rövid bemenő adatsor mellett is garantálni kell a megfelelő biztonságot.
- A kódképző algoritmus véletlenszerűségét biztosítani kell
- A törvény előírja, hogy az eljárásnak tartalmazni kell véletlenszerű (azaz tikos) kulcselemeket. Ezen elemek kellő szabadsági fokú véletlenszerűségét garantálni kell azért, hogy az eljárás ismeretében se legyenek kipróbálhatóak, megszüntetve ezáltal a véletlenszerűséget. A szabadsági foknak 100 bitnek kell lennie.

A különböző összekapcsolások ismételt összekapcsolásának megakadályozása teljesüléséhez közvetlenül hozzájárul a véletlenített hash függvény azáltal, hogy az anonim kapcsolati kód képzési módszerében véletlenszerű elemeket is használnak, így több adatkérést nem lehet ugyanazzal a kóddal megvalósítani, ezáltal ezek eredményeit még az adatkérő és az adatkezelő sem képes összevonni.

A Codefish hash függvény nem ütközésmentes és a gyakorlati támadások különösen érvényesek rövid bemenő adatsorra [10], ezért ennek kiküszöbölésére az összekapcsolást elvégző algoritmust megvalósító programot a következő tulajdonságokkal ruházták fel:

- Egy véletlenszerű, 1024 bites S, illetve egy véletlenszerű 3*1024 bites V paramétert generál, melyeket a kódképzési módszer adatkérésenkénti módosítására használnak. Ezt az S és V paramétert állítja elő az NKHT minden adatkéréskor egyszer, és küldi el az érintett adatkezelőknek. Az S paramétert modulusnak, a V paramétert maszk-vektornak hívják.
- A program a célkörnyezetben képes arra, hogy egy adott, 1024 bites bemenő adatra, valamint az adott modulusra és maszk vektorra válaszul egy 256 bites, hexadecimális számrendszerben ábrázolt számsorozatot adjon. Ez lesz a kapott anonim kapcsolati kód.

Az NKHT kérésére a Hunguard kft. elvégezte az adatbázisok anonimizált összekapcsolását megvalósító rendszer biztonsági értékelését [14]. A minősítők a tanúsítványban kijelentik, hogy az összekapcsolást elvégző algoritmus és program megfelel a rendszer biztonsági előirányzatában [12] megfogalmazott funkcionális és garanciális biztonsági követelményeknek.

AZ ÖSSZEKAPCSOLÁS FOLYAMATA

Az összekapcsolás folyamatának lépéseit az NKHT eljárásrendjét leíró dokumentum alapján tárgyaljuk [15]. Adatbázisok összekapcsolását miniszter vagy kormányhivatal vezetője rendelheti el. Az adatbázisok összekapcsolásának folyamatában az adatbázis létrehozásáért felelős szervnek az adatkezelőkkel egyeztetve ki kell választania az anonim kapcsolati kód képzésére felhasználható személyi azonosító mezőket. Ezeket a törvény [2] szerint úgy kell meghatározni, hogy kezelésükre minden érintett adatkezelő jogosult legyen, ugyanakkor sem az adatkérő, sem az adatbázis létrehozásáért felelős szerv nem. Ezután az adatbázis létrehozásáért felelős szervnek elő kell előállítania az anonim kapcsolati kód képzéséhez szükséges véletlenszerű paramétert és továbbítania kell a programot, a futtatásához szükséges információkat és a paramétert a mintaképzésért felelős adatkezelőhöz és a kapcsolódó adatkezelőkhöz. A sikeres továbbítás után az adatbázis létrehozásáért felelős szerv törli az anonim kapcsolati kód képzésére szolgáló véletlen paramétert.

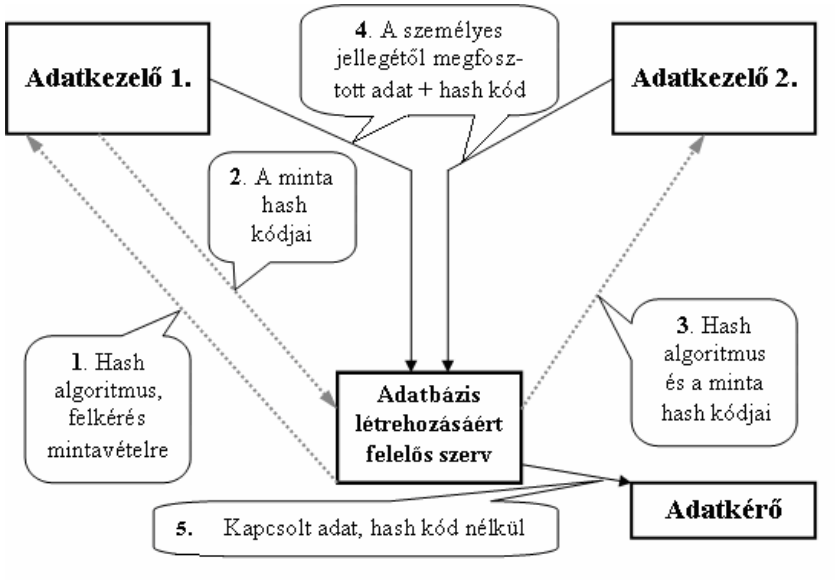
Az adatbázis létrehozásáért felelős szerv kérésére az egyik adatkezelő szerv (a továbbiakban mintaképzésért felelős adatkezelő) mintát vesz az általa kezelt adatbázisból. Mintaképzésnek nevezzük azt a folyamatot, amikor a mintaképzésért felelős adatkezelő az adatkérés paramétereinek megfelelően kialakítja az igényelt mintát, azaz leválogatja az

adatbázisaiból a kiválasztási és mintaképzési szempontoknak megfelelő tételeket. Ezután a minta tételeiben a kiválasztott azonosító adatokra alkalmazza a kapott anonim kapcsolati kód előállító programot. Végül az előállított kapcsolati kódokat és a hozzájuk tartozó, igényelt adatokat az adatkezelő átadja az NKHT-nak.

Az adatbázis létrehozásáért felelős szerv az anonim kapcsolati kódokat, az ehhez tartozó adatok nélkül átadja az adatkérésben megjelölt többi adatkezelőnek. Az érintett adatkezelők saját adatbázisaikból leválogatják az eredeti kérés kiválasztási szempontjainak megfelelő tételeket, alkalmazzák rájuk a kapott anonim kapcsolati kód előállító programot, és kiválasztják azokat, amelyek megfelelnek a kapott kapcsolati kódoknak. Ezután elvégzik az anonimizálást, majd az adatkezelők átadják az NKHT-nak az eredményül kapott adatokat a kapcsolati kódokkal. Az adatkezelő szervek a kódképzéshez kapott véletlenszerűséget biztosító paramétert és a kapcsolati kódokat az adatok átadása után törlik.

Az összekapcsolás folyamatában az NKHT átveszi a leválogatott és anonimizált adatokat tartalmazó állományokat és a kapcsolati kódok segítségével összekapcsolja ezeket. Amennyiben ismétlődő tételek is előfordulnak vagy keletkeznek, és az adatkérő igényli az azonos alanyhoz tartozó adatok jelzését, a kapcsolati kódokat az összekapcsolás után futó sorszámmal helyettesíti úgy, hogy azonos kódhoz tartozók azonos sorszámot kapjanak. Ellenőrzi - szükség esetén az adatkezelők bevonásával -, hogy az eredmény továbbra is megfelel az anonimizálás feltételeinek. Törli a kapcsolati kódokat az eredményből és átadja az adatkérő kapcsolattartójának a leválogatott és összekapcsolt adatokat.

A döntés-előkészítéshez szükséges adatok hozzáférhetőségének biztosításáról szóló törvény indoklásában [16] megtalálható ábra jól szemlélteti az összekapcsolás folyamatát:



1. ábra. Az összekapcsolás folyamata [16]

Előfordulhat, hogy az eredményezett minta túl kevés adatot tartalmaz, vagy a reprezentativitási mutatói az igényeknek nem felelnek meg (például olyan esetekben, amikor a mintaképző adatkezelő nem rendelkezett minden szükséges adattal). Ebben az esetben a kapcsolódó adatkezelő jelezheti a hibát, és kérheti a mintaképzésért felelős adatkezelőt, hogy készítsen bővebb, több tételből álló mintát. Ilyenkor az első lépéstől kezdve ismétlődik ez a folyamat. Amennyiben a minta megfelelő, az adatkezelők elkezdhetik a leválogatást és anonimizálást.

Egyének esetén a kódolás legtöbbször a személyek természetes azonosítóin alapul. Ekkor a lenyomatképzés felveti az adatok különböző adatkezelőknél lévő szinkronizáltságának kérdését. Elvileg ez a probléma rendezett, mert a természetes azonosító adatok tekintetében a

személyiadat-nyilvántartás a hitelesnek elfogadott, és [17] törvény rendelkezik a személyiadat- és lakcímnyilvántartásból történő rendszeres adatátadásról. Ugyanakkor a gyakorlat szerint jelentős eltérések vannak az egyes adatkezelőknél tárolt természetes azonosító adatok tekintetében. Ezért ilyen esetben meg kell vizsgálni az adatkezelőknél tárolt természetes azonosító adatok szinkronizáltságnak mértékét és meg kell határozni ezek közös formátumát. Ki kell térni a több elemből álló azonosító helyes összefűzésére, a megfelelő karakterkódolásra, az eltérő formátumokból adódó egyéb konverziós lépésekre.

Az adatbázis létrehozásáért felelős szervnek az adatok átadása előtt a polgárok lakcímére vonatkozó adatokat úgy kell módosítania, hogy azokból az érintettek lakóhelye a kistérségnél pontosabban ne legyen megállapítható, továbbá adatkezelőnként legalább száz főt eredményező mintát kell venni. A teljes népességre vonatkozó adatbázisok esetében a minta nem haladhatja meg a teljes sokaság 50%-át. A [2] törvényben foglaltak nem alkalmazhatóak a külön törvény által szabályozott minősített adatokra, illetve azokra a közérdekű adatokra, amelyeknek megismerhetőségét nemzetbiztonsági érdekből szintén külön törvény korlátozza.

KÖVETKEZTETÉSEK

A közigazgatási adatvagyon felhasználásának segítése céljából született meg az Európai Parlament és a Tanács 2003/98/EK számú irányelve. Az irányelv minimum szabályokat állapít meg a tagállamok közigazgatási szervei birtokában lévő dokumentumok (ideértve az adatokat is) további felhasználására. Az irányelv céljainak elérését Magyarországon a döntés előkészítéshez szükséges adatok hozzáférhetőségének biztosításáról szóló 2007. évi CI. törvény teszi lehetővé. Ez a törvény és a hozzá kapcsolódó kormányrendelet [18] szabályozza a közigazgatási adatok összekapcsolásának lehetőségeit is. Az adatbázis összekapcsolás folyamatának bonyolultsága mögött – mely az adatkezelőktől független újabb szereplő bevonását és erős kriptográfiai algoritmusok alkalmazását is igényli – elsősorban az európai szinten is kifejezetten szigorúnak számító hazai adatvédelmi előírások állnak.

Hazánkban az adatbázisok összekapcsolására vonatkozó jogi korlátozások a személyes adatok védelméhez kapcsolódnak. Magyarországon a személyes adatok védelmét információs önrendelkezési jogként értelmezik, eszerint személyes adatot felvenni és felhasználni csakis az érintett beleegyezésével szabad. Személyes adat kötelező kiszolgáltatását és felhasználását csak kivételesen és törvényben lehet elrendelni. A személyes adatok védelmére érvényes a célhoz kötöttség elve, miszerint személyes adatot feldolgozni csak pontosan meghatározott és jogszerű célból szabad. Személyes adatot harmadik személy számára hozzáférhetővé tenni csak akkor szabad, ha az erre konkrét törvényi felhatalmazással rendelkezik, vagy az érintett ehhez hozzájárulását adta.

Az adatbázisok összekapcsolása a kért adatok más számára történő rendelkezésre bocsátását jelenti. Az összekapcsolás lehetősége kizárná az érintettet az adatáramlásból, korlátozná abban, hogy adatai útját és felhasználását ellenőrizze. A személyes adatokat is tartalmazó adatbázisok összekapcsolásának jogi feltétele az adatok személyességének megszüntetése (anonimizálása). Az anonimizálás egy olyan eljárás, amelynek eredményeként az egyes adatsorok már nem kapcsolhatóak adott személyhez.

A magyarországi szabályozás az adatkezelőket (még ha kormányzati szervezetekről van is szó!) nem megbízható szereplőnek tekinti. Az összekapcsolás folyamatára megfogalmazott két biztonsági cél (1. az összekapcsolt adatok alanyának védelme, 2. összekapcsolások ismételt összekapcsolásának megakadályozása) megvalósítása érdekében Magyarországon az adatkezelők a tárolt adataikon kívül más adatkezelőktől nem kérhetnek adatokat összekapcsolás céljára. A hazai szabályozás nem engedélyezi azt a szenáriót, miszerint egy adatkezelő a saját személyes adatait összekapcsolja egy másik féltől kapott adatokkal, majd anonimizálja a kapott adathalmazt, amellett, hogy törli az adatbázisából az „illegális”

személyes adatokat. A jelenleginél egyszerűbb - azaz plusz szereplő bevonását nem igénylő rendszer - úgy lenne elképzelhető, ha a személyes adatok védelme tekintetében a közigazgatási szervek megbízhatóbb szereplőnek számítanának (nem az lenne a feltételezés, hogy az elképzelhető legrosszabbat fogják megtenni) és a visszaélés elkerülését, a törvénytelen használatot és a szabályos működést rájuk vonatkozó jogi előírásokkal és megfelelő ellenőrzéssel biztosítanák.

Felhasznált irodalom

[1] Az Európai Parlament és a Tanács 2003/98/EK irányelve (2003. november 17.) a közszféra információinak további felhasználásáról.

[2] 2007. évi CI. törvény a döntés-előkészítéshez szükséges adatok hozzáférhetőségének biztosításáról

[3] 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról

[4] 2003. évi XCII. törvény az adózás rendjéről

[5] 1993. évi XLVI. törvény a statisztikáról

[6] Fleiner Rita – Munk Sándor: Közigazgatási adatbázisok összekapcsolásának biztonsági kérdései. Hadmérnök, VII. Évfolyam 4. szám - 2012. december

[7] 1996. évi XX. törvény a személyazonosító jel helyébe lépő azonosítási módokról és az azonosító kódok használatáról

[8] 15/1991. (IV. 13.) AB határozat.

[9] Bérczes A., Ködmön J., and Pethő A. A one-way function based on norm form equations. Periodica Mathematica Hungarica 49, 1 (2004), 1–13.

[10] Jean-Philippe Aumasson: Cryptanalysis of a hash function based on norm form equations. Cryptologia, 33. évf. (2009. Január), 12–15. p. ISSN 0161-1194.

[11] Folláth János: Kriptográfiai hash függvények és álvéletlenszám generátorok. Egyetemi doktori (PhD) értekezés. Debreceni Egyetem, 2011.

[12] Adatbázisok anonimizált összekapcsolását megvalósító rendszer (DBCS) Rendszer biztonsági előirányzat 2009.01.23. www.hunguard.hu/fileok/m003_bizt_eloiranyzat.pdf (2012.11.30.)

[13] Útmutató az IT biztonsági szintek meghatározásához BME IK, 2008.

[14] Adatbázisok anonimizált összekapcsolását megvalósító rendszer (DBCS v1.0) biztonsági minősítése http://www.hunguard.hu/fileok/m003.pdf (2012.11.30.)

[15] Eljárásrend a döntés-előkészítéshez szükséges adatok hozzáférhetőségének biztosítására Verzió: 1.00 Készítette: NKHT 2009.03.10.

[16] MAGYAR KÖZTÁRSASÁG KORMÁNYA T/3029. számú törvényjavaslat a döntés előkészítéshez szükséges adatok hozzáférhetőségének biztosításáról www.parlament.hu/irom38/03029/03029.pdf (2012.11.30.)

[17] 1992. évi LXVI. Törvény a polgárok személyi adatainak és lakcímének nyilvántartásáról

[18] 335/2007. (XII. 13.) Korm. Rendelete a döntés előkészítéshez szükséges adatok hozzáférhetőségének biztosításáról szóló 2007. évi CI. törvény végrehajtásáról