

A Semmelweis Egyetem Adatstratégiája

Készítette:

Klinikai Adatszolgáltató Intézet

Dr. Bagyura Zsolt

Dr. Kiss Loretta

Tartalom

Bevezető.....	3
1. Adatgyűjtés és adattárolás	3
1.1. Klinikai adatok másodlagos felhasználása.....	3
Adatáramlás és álnevesítés.....	4
1.2. Strukturált adatgyűjtés leletezéshez, prospektív kutatásokhoz	4
1.3. Radiológiai képi adatok	4
1.4 Egyéb adatforrások	5
2. Adatok hasznosítása	5
2.1 Adattérkép, adatkatalógus	5
Simmelweis Dataverzum	5
2.2. Adatminőség.....	6
2.3. Az adathasznosítás elvi megfontolásai	7
2.3. Az adathasznosítás gyakorlata.....	9
Adatigénylés	9
Anonimizálás, álnevesítés.....	11

Bevezető

A Semmelweis Egyetem Magyarország legnagyobb egészségügyi szolgáltatója és ennél fogva a legnagyobb egészségügyi adatvagyon kezelője. Az Egyetem által kezelt adatvagyon a kutatási és oktatási cél mellett gazdasági jelentőséggel is bír, figyelemmel az Európai Unió adatok és másodlagos adatok felhasználásának stratégiájára.

A 2025. március 26-án közzétett, európai egészségi adatterről szóló rendelet, mint a közös uniós adattér eleme, biztosítja az egészségügyi adatok kutatási, innovációs, szakpolitikai és szabályozási tevékenységekhez való biztonságos és megbízható felhasználását (az adatok másodlagos felhasználása); és előmozdítja az elektronikus egészségügyi dokumentációs rendszerek egységes piacának létrehozását, amely az elsődleges és a másodlagos felhasználást egyaránt támogatja.

A Semmelweis Egyetem – ehhez kapcsolódóan – a kezelésében lévő adatvagyon egységes hasznosításának céljára a jelen stratégiát alkotta meg, melyhez a Szabályzat szerinti és rendkívüli esetben attól eltérő rendelkezéseit minden, az Egyetemmel kapcsolatban álló, vagy kapcsolatot létesíteni kívánó szervezetnek igazodnia kell.

A dokumentum célja összefoglalni a Semmelweis Egyetemen a betegellátás és kutatás során keletkezett adatok másodlagos felhasználásának folyamatát. Ez a folyamat két fő részből tevődik össze:

- 1) adatgyűjtés és adattárolás irányelvei, szempontjai;
- 2) adatigénylés és adatszolgáltatás folyamata.

1. Adatgyűjtés és adattárolás

1.1. Klinikai adatok másodlagos felhasználása

Az Egyetem klinikáin zajló betegellátás során keletkezett és a kórházi informatikai rendszerben rögzített adatokat a másodlagos felhasználás érdekében egy standardizált, pszeudonimizált adattárházban (Semmelweis Egyetem Klinikai Adatok) gyűjti és tárolja.

Ebben az adattárházban az adatok a forrásrendszerhez hasonló felbontásban, de már tisztított, standardizált, strukturált formában érhetőek el, illetve az adattárház több forrásból származó adatokat is tud fogadni¹. Fő célja, hogy egyetlen adatforrást biztosítson kutatási és egyéb célokra optimalizálva, valamint, hogy egyértelmű és szabályozott legyen az adatokhoz való hozzáférés. Jelenleg az adatbázis a Kórházi Információs Rendszerből (MedSolution) származó adatokat tartalmaz, de a jövőben más forrásokból származó adatok (pl.: regiszterek) kezelését is tervezzük. Az adatbázis az

¹ A későbbiekben leírtak szerint az adattárházba már anonimizált adatok kerülnek.

OMOP adatmodellt követi (<https://www.ohdsi.org/data-standardization/>). Az OMOP Common Data Model (CDM) egy egységes rendszer, ami az adatok nemzetközileg szabványosított kódrendszerek szerinti és meghatározott struktúrában való tárolását jelenti. Ily módon az adattárház interoperabilissá válik, mely jelentősen növeli az értékét. Ez a megoldás teszi lehetővé, hogy a Semmelweis Egyetem által kezelt adatvagyon hazai és nemzetközi együttműködésekben valóban felhasználható legyen, tehát biztosítja az adatok egymásnak való megfeleltetését, együttműködési képességét és a kutatások reprodukálhatóságát.

Adatáramlás és álnevesítés

Az adatok forrása az Egyetem egészségügyi informatikai rendszere, a Medsolution rendszer. Az adatok Szabályzat szerinti lekérdezését követően egy ideiglenes adatbázisba (data warehouse, DWH) kerülnek továbbításra, amely továbbra is tartalmazza a betegazonosítókat, és a SE belső hálózatában található. Adattisztítási lépések után álnevesítés következik, majd az adatok átalakítása az OMOP CDM szabványra. Az így keletkező adatbázis szintén helyben, a belső hálózaton, de egy másik szerveren tárolódik. Az álnevesítés miatt önmagában már nem tartalmaz közvetlenül személyes adatokat. (A szabad szövegek ilyen jellegű adattartalmával külön fejezetben foglalkozunk.)

1.2. Strukturált adatgyűjtés leletezéshez, prospektív kutatásokhoz

Az Egyetemen a prospektív adatgyűjtések támogatására egy saját fejlesztésű, űrlap-rendszerű elektronikus esetjelentő lap (eCRF-szerű) regiszter rendszer áll rendelkezésre, amely támogatja, kiszolgálja a szakmákhoz kötött, magasabb granularitású adatgyűjtéseket. A rendszer - felépítéséből adódóan - biztonságos helyi adattárolást biztosít, könnyen használható kutatási célokra, ugyanakkor alkalmas szöveges leletek előállítására is a strukturált adatokat felhasználva, így az Egyetem számos klinikáján a mindennapi ellátás része. Ezen adatbázisok kezelését, a felhasználók támogatását, valamint új űrlapok, adatbázisok létrehozását és fejlesztését a Klinikai Adatszolgáltató Intézet végzi.

1.3. Radiológiai képi adatok

Az Egyetem egységes PACS rendszert üzemeltet, melybe minden képalkotó berendezés továbbítja az adatokat a betegellátás támogatására és az archiválás biztosítására. Ez hatalmas mennyiségű képi - DICOM formátumú - adatot jelent az elmúlt évtizedek tekintetében. Ezen adatok kutatási célból használhatóak, és nem szükséges ezeket az adattárházba áttölteni, mivel a klinikai PACS rendszerből egyszerűen és szabványosan kinyerhetőek igény esetén. A kutatási adatok további feldolgozásával (álnevesítés vagy anonimizálás) külön fejezet foglalkozik.

1.4 Egyéb adatforrások

A betegellátás és kutatás során számtalan egyéb, jellemzően lokálisan használt informatikai rendszer van használatában, melyeken az OMOP adattárházba való becsatornázást el kívánjuk végezni, majd Klinikai Adatszolgáltató Intézetben (KLASZ) keresztül, egységesen történik az adatok kezelése és az adatszolgáltatás. Minden szervezeti egység törekszik az OMOP tárházba való csatlakozásra az adatok kezelésére vonatkozó irányelvek egységes érvényesítése mellett.

2. Adatok hasznosítása

2.1 Adattérkép, adatkatalógus

Az adatok hasznosításának alapvető feltétele, hogy egy átlátható, kereshető, egységes és könnyen elérhető lista (ún. metaadat katalógus) legyen az Egyetemen épülő adatbázisainkról. Ezt teszi lehetővé a KLASZ és a Központi Könyvtár által közösen megálmodott és a megvalósulás utolsó lépcsőfokán tartó projekt, a Semmelweis Dataverzum létrehozása.

Semmelweis Dataverzum

Az Open Science korszakában igény van a kutatások alapját képező adatbázisok közzétételére a FAIR alapelvek (Findable, Accessible, Interoperable, and Reusable) mentén. Az egészségügy területén ugyanakkor, mivel az adatok – főszabály szerint – nem tehetők közzé, így erre kevés törekvés van, noha minden esetben lenne lehetőség az adatbázisok meglétének és metaadatainak rögzítésére.

A Semmelweis Dataverzum az Egyetemen a betegellátás során vagy célzottan kutatási tevékenység keretén belül összegyűjtött, és adatbázisokban tárolt adatokat (át)láthatóvá, kereshetővé, és hivatkozhatóvá teszi az adatbázisok leírásának, metaadatainak megjelenítésével egy arra alkalmas „dataverzum”-ban, amelyben személyazonosításra alkalmas adatok nem szerepelnek.

A Semmelweis Dataverzumba kerülnek tárolásra azon adatbázisok, amelyeket szellemi alkotás részeként az Egyetemi Innovációs Bizottság befogadott. Ezen adatbázisok hozzáférése szigorúan ellenőrzött.

A KLASZ és a Központi Könyvtár együttműködve az alábbi lépésekben valósította meg a fenti célt.

1. Tájékozódás a meglévő külföldi és hazai dataverzumokról
2. Előnyök/Hátrányok mérlegelésével kiválasztásra került a nekünk megfelelő formátum kiemelet figyelemmel a védendő adatokkal kapcsolatos adatkezelési és információbiztonsági követelményekre
3. Összegyűjtésre, listázásra kerültek az itt megjeleníteni kívánt adatbázisok

4. Kialakítottuk a logikai felépítést, csoportosítási, fontossági szempontokat
5. Megtörtént az elsődleges adatbetöltés
6. Közzététel az Egyetem honlapján (könyvtár).

Belső előnyök

A projekt megvalósulásával az egyetemi kutatók, klinikusok könnyedén gyors és pontos információt kaphatnak

- a) egyrészt az adattárházban már meglévő, létrehozott adatok köréről,
- b) másrészt a folyamatban lévő kutatások során keletkező adatgyűjtések tartalmáról, egy-egy témában zajló adatgyűjtésről.

Ezáltal lehetőség nyílik szélesebb körű, multidiscplináris együttműködésekre, valamint elkerülhető, hogy több munkacsoport különböző módon és helyre gyűjtsön hasonló adatokat, valamint új projektbe fogjon, meglévő adatok felhasználása helyett, ezzel munkát, időt lehet megtakarítani.

Külső előnyök

Az egyes adatbázisok egyedi elektronikus azonosítót kapnak, ezáltal a tudományos közleményekre vonatkozó szabályok szerint publikációkban hivatkozhatóak lesznek. Növeljük a láthatóságot nemzetközi kutatások, valamint cégek számára is, mely együttműködési lehetőségeket teremthet. Az Egyetem magas színvonalú és szerteágazó kutatási és gyógyító tevékenységét ily módon is meg tudjuk jeleníteni.

The screenshot shows the website of the Központi Könyvtár (Central Library) of Semmelweis University. The header includes the university's logo, the library's name in Hungarian and English, and a search bar with the text 'Mit keres?' and a magnifying glass icon. Below the header, there is a navigation bar with links to 'INFORMÁCIÓFORRÁSOK', 'SZOLGÁLTATÁSAINK', 'KUTATÁSTÁMOGATÁS', 'OPEN SCIENCE', and 'KÖNYVTÁRRÓL'. The main content area features a star icon and the text '_DATAVERSE' in large, bold letters. Below this, there is a search bar with the text 'Keresés (cím, leírás, kulcsszavak):' and a magnifying glass icon. To the right of the search bar is a link for 'Összetett keresés'. Below the search bar, there is a blue button labeled 'Keresés'. The search results show a dataset with the ID '1 DOI: 10.42.53et-s' and the title 'Szív CT vizsgálatok standardizált leletezése kliniumban (plakk és eltérés alapú leletezés) (aktív)'. The description of the dataset is: 'Szakmai irányelvek szerinti standardizált szív CT leletezés és adatgyűjtés kialakítása. Plakk morfológia vizsgálata. Képkötő eredmények elemzése klinikummal. Szív CT felvételezés optimális technikai beállítása. AI-alapú digitális fejlesztéshez strukturált adatbázis kialakítása.' The dataset is associated with 'Semmelweis Egyetem Városmajori Szív- és Érgyógyászati Klinika'.

A Semmelweis Dataverzum tesztfelülete 2025.12.02-án.

2.2. Adatminőség

Az adatok hasznosíthatóságát nagy mértékben befolyásolja, hogy milyen minőségűek az adatok. A klinikai ellátás során a kórházi informatikai rendszerbe való adatbevitelre nincs

ráhatásunk, sok tényezőtől függ, és mindig szem előtt kell tartanunk, hogy az adatgyűjtés eredeti célja a betegellátás/ egészségügyi finanszírozás ebben az esetben.

A Semmelweis Egyetem Klinikai Adattárház kialakítása során a szabványosítási folyamatba építetten vannak adattisztítási lépések, melyek standard analitikai módszerekkel folyamatosan ellenőrizhetők, majd javíthatók az adatminőség karbantartásának részeként.

A prospektív adatgyűjtések, tehát a strukturált leletező és egyéb űrlapos adatbázisok előnye, hogy az adatbevitel szabályozottabb, így kevesebb a hiányzó vagy hibás adatpont.

2.3. Az adathasznosítás elvi megfontolásai

Az Egyetem szervezeti egységei közül a klinikai adatok másodlagos hasznosításával a KLASZ foglalkozik az SE TTC Zrt. közreműködésével.

Az adatok hasznosításából származó profit sokrétű, ide értendő a társadalmi, közegészségügyi haszon, a szellemi tulajdon (IP) keletkezése, termék licence, szerzői jog (royalty) vagy anyagi haszon.

Az alábbi táblázat összefoglalva mutatja be az adatok hasznosulásának módjait különböző szempontok szerint:

Adatfelhasználás célja	Tudományos/oktatási		Piaci (termékfejlesztés)
Adatkérő	SE belső	Akadémiai külső	Ipari külső
Kért adatok típusa	egyszerű/tabuláris	Nyers digitális, képi	Digitális, annotált
Adatkérés komplexitása	Egyszerű (demográfia, anamnézis)	Közepesen összetett (ellátások, teljesítések, utánkövetés)	Összetett (pl. nyers digitális, genetikai, stb.)
Adatok átadási formája	aggregált	tételes	
Személyes azonosítók	anonimizált adat	pszeudonimizált adat	
Adatkérés rendszeressége	egyszeri	többszöri	
Felhasználás időtartama	Korlátozott		Korlátlan

Adatátadás módja	Virtuális kutatószoba		Megosztott mappa	Egyéb	
Konzultáció igény	Nincs		Max 1 munkanap	Több, mint 1 munkanap	
Előfeldolgozás, adattisztítás, statisztikai elemzés KLASZ részéről szükséges?	Nem		Csak adattisztítás szükséges	Statisztikai elemzés is szükséges	
Szellemi tulajdon	Nem képződik	Képződik, SE kizárólagos tulajdona marad	Képződik, SE és adatkérő közös tulajdonába kerül	Képződik, adat kérő tulajdonába kerül	
Ellenszolgáltatás	Nincs	Létrehozott termék licensze/publikáció	Szerzői jogdíj (royalty)	Díjtétel (HUF)	

Nemzetközi hálózatok- példák jelenleg is folyó adathasznosításra

Az Egyetem a fent említett OMOP CDM-re standardizált adatbázisával több nemzetközi kutatási hálózatban is részt vesz. Korábban az EHDEN, majd Evidence Network tagja, mely elsősorban akadémiai jellegű kutatásokat végez, epidemiológiai és real-world evidence vizsgálatokat folytat.

A DARWIN EU® pedig olyan adatelemzési és valós világbeli lekérdezési hálózat, Data Analysis and Real World Interrogation Network (DARWIN EU®), amelyet az Európai Gyógyszerügynökség (EMA) és az Európai Gyógyszerügyi Szabályozási Hálózat hozott létre annak érdekében, hogy az Európai Unió (EU) egészségügyi adatbázisaiból megbízható bizonyítékokat szolgáltatasson az emberi felhasználásra szánt gyógyszerek - beleértve a vakcinákat is - használatáról, biztonságosságáról és hatékonyságáról.

A bizottság meghatároz olyan kérdéseket, melyek befolyásolhatják a döntéshozatalt. Ezután konkrét kutatási kérdést fogalmaznak meg, és beadnak egy study kérelmet. Egy koordinációs központ megvizsgálja a kivitelezhetőséget, összeköti az adatpartnereket, elkészíti a protokollt és a statisztikai kódot, és később felügyeli a study kivitelezését. Az adatpartnerek megkapják és lefuttatják a kódot, átnézik az eredményeket, és visszaküldik a koordinációs központnak. A központ megkapja az aggregált eredményeket, ellenőrzi és analizálja azokat, és elkészíti a study jelentést. A bizottság megosztja az eredményeket. A bizottság integrálja az eredményeket a döntéshozatali folyamatokba.

Az Egyetem – mint adatpartner – keretszerződés alapján vesz részt ebben, és anyagi ellentételezést kap az egyes vizsgálatok lefuttatásáért.

A klinikai adatok másodlagos, további hasznosításáról a Felek az Egyetem és a TTC közötti külön megállapodásban rendelkeznek.

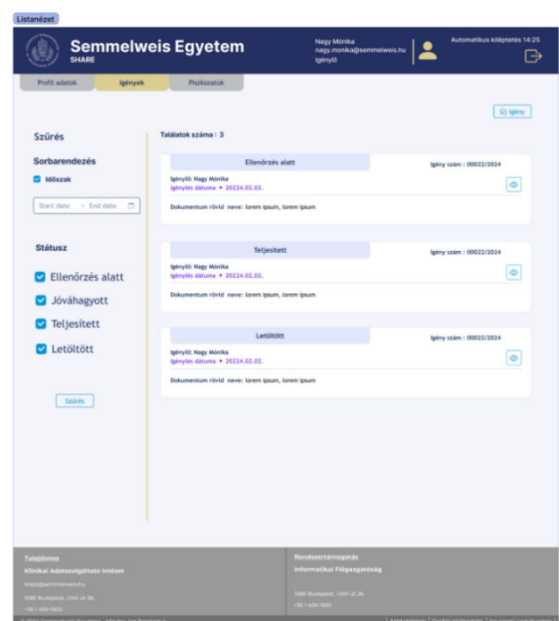
Az adatok átadásán túl, ahhoz kapcsolódóan az Egyetem adatelemzési és kutatószoba szolgáltatást is nyújt.

- Az **adatelemzési szolgáltatás** keretében anonimizált vagy álnevesített, aggregált adat vagy elemzési eredmény kerülhet átadásra.
- A (virtuális) **kutatószoba szolgáltatás** keretében a kutatószobában meghatározott időtartamra szóló hozzáférést biztosítunk a megjelölt klinikai adatok anonimizált adatbázisához. A virtuális kutatószobák informatikai szempontból az egyetem belső informatikai hálózatától szeparáltak.

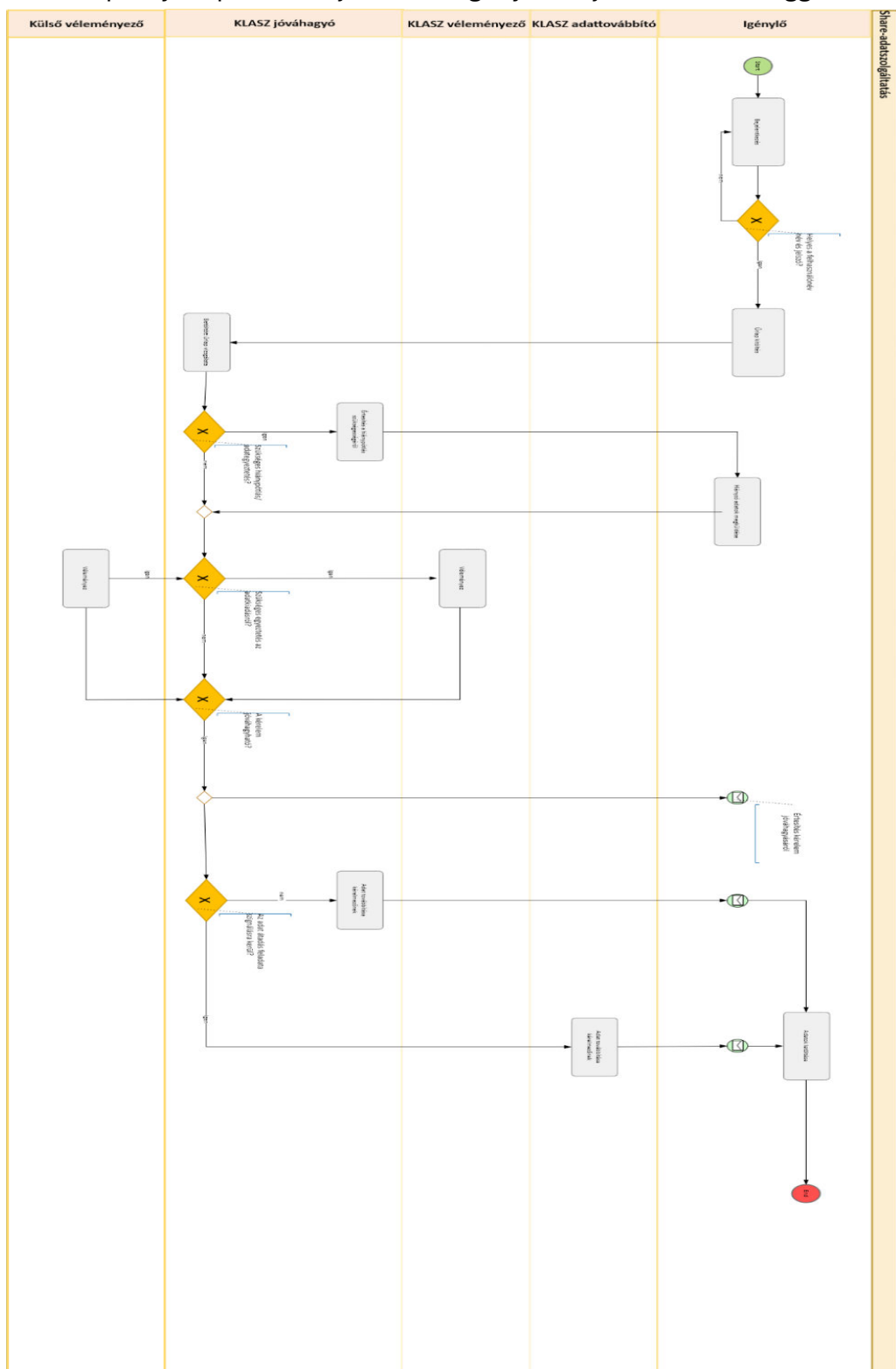
2.3. Az adathasznosítás gyakorlata

Adatigénylés

Az adatigénylési kéréseket a Klinikai Adatszolgáltató Intézet kezeli. Az igények jobb áttekinthetősége, a folyamat egységesítése, gyorsítása és követhetővé tétele érdekében minden kérés egy dedikált informatikai rendszeren keresztül kerül feldolgozásra. A SHARE rendszerhez az egyetemi SSO, tehát O365 bejelentkezés után, az Egyetem belső hálózatáról lehet hozzáférni. Az adatkérés egy jóváhagyási folyamaton megy keresztül; a



folyamat és az adatátadás ténye is naplózásra kerül. A rendszer működését az alábbi folyamatábra és képernyőkép illusztrálja. Az adatigénylési eljárással összefüggő részletes



szabályokat az SE vonatkozó, az adatok másodlagos felhasználására irányuló eljárást rendező szabályzat tartalmazza.

Anonimizálás, álnevesítés

A klinikai adatok anonimizált formában adhatók át, illetve tehetők hozzáférhetővé. Az anonimizált adatok olyan adatok, amelyeket oly módon változtattak meg, hogy az adatalany nem, vagy már nem azonosítható olyan módon, amely észszerűen valószínűsíthetően felhasználható. Az azonosíthatóság tekintetében figyelembe kell venni minden korszerű és aktuálisan felhasználható módszert, különös figyelmet fordítva arra, hogy a technológia jelenlegi állása szerint milyen módszer, eszköz vált az utóbbi időben elérhetővé és alkalmazhatóvá. Annak meghatározásakor, hogy mely eszközökről feltételezhető észszerűen, hogy egy adott természetes személy azonosítására fogják felhasználni, az összes objektív tényezőt figyelembe kell venni, így például az azonosítás költségeit és időigényét, számításba véve az adatkezeléskor rendelkezésre álló technológiákat, és a technológia fejlődését. Ennek megfelelően, az anonimizálási technikák alkalmazása során fel kell mérni valamennyi olyan körülményt, amely a megfelelő anonimizációs technika kiválasztását befolyásolhatja. Az azonosítás kapcsán figyelembe kell venni a kiválasztás, összekapcsolhatóság és következtetés által történő lehetséges azonosíthatóságot is. Ügyelni kell továbbá arra, hogy kis létszámú érintetthez vonatkozó adat az érintett azonosíthatóságához vezethet.

Álnevesítés alkalmazandó, amennyiben a célhoz az anonimizálás nem elegendő (pl: utánkövetéses vizsgálatok), azaz az adatkezelés célja anonimizálás mellett nem elérhető. A megfelelő álnevesítési technika meghatározása és kivitelezése a személyes adatok oly módon történő átalakítását jelenti, hogy további információk felhasználása nélkül már nem állapítható meg, hogy a személyes adat mely konkrét személyre vonatkozik, feltéve, hogy az ilyen további információkat külön tárolják, és technikai és szervezési intézkedések megtételével biztosított, hogy ezt a személyes adatot nem lehet egyénekhez kapcsolni. A gyakorlatban ez jelentheti a személyes adatok helyettesítését egy adatkészletben közvetett azonosító adatokkal (más néven, sorszám stb.). Az álnevesített adat továbbra is személyes adat, kezelése a GDPR hatálya alá tartozik. Valamely természetes személy azonosíthatóságának meghatározásakor minden olyan módszert figyelembe kell venni, amelyről észszerűen feltételezhető, hogy a természetes személy közvetlen vagy közvetett azonosítására felhasználható.

Az azonosítás, illetve újra azonosítás kockázatát mindkét esetben folyamatosan értékelni kell. Ezek ismeretében – ha szükséges – a technikai és szervezési intézkedéseket újra kell értékelni.

Egyes adattípusok

Strukturált formátumú adatok anonimizálása a leginkább magától értetődő folyamat. Az adatsorokból eltávolításra kerülnek a személyes adatok, mint a név, lakcím, telefonszám, azonosítószámok, mint a TAJ-szám, esetszám, naplósám. A születési dátum csak hónap pontossággal marad meg. Aggregált adatok esetén cellatitkosítás alkalmazandó, az 5 elemnél kisebb tartalmú cellák esetében.

Szabad szöveges adatok (leletek, zárójelentések, ambuláns lapok) anonimizálhatósága bonyolultabb megoldásokkal lehetséges csak. Ennek keretrendszere jelenleg nem megoldott, így jelenleg manuális ellenőrzéssel biztosítható.

Orvosi képi adatok esetén a DICOM file-ok anonimizálása az alábbi szabvány szerint történik: Basic Profile of PS 3.15-2021b Table E.1-1.