147. / 621–634.

MAGYAR ÁLLATORVOSOK LAPJA | 2025. OKTÓBER

https://doi.org/10.56385/magyallorv.2025.10.621-634

Establishing a data infrastructure system for veterinary public health and food chain safety data through the development of a repositioning platform

Zs. Farkas[1*]
Sz. Csorba[1]
K. Vribék[1]
M. Süth[2]
Zs. Czudor[3]
Á. Tényi[3]
Á. Jóźwiak[1]

1. Állatorvostudományi Egyetem, Élelmiszerlánc-tudományi Intézet, Digitális Élelmiszertudományi Tanszék, H-1078 Budapest, István u. 2.

2. Állatorvostudományi Egyetem, Élelmiszerlánc-tudományi Intézet, Budapest

3. E-Group ICT Software Zrt., Budapest

**e-mail: farkas.zsuzsa@univet.hu*

ÉLELMISZER-HIGIÉNIA

Állatorvosi közegészségügyi és élelmiszerlánc-biztonsági adatok adatinfrastruktúra rendszerének megalapozása repozíciós platform kialakításával

Farkas Zsuzsa[1*], Csorba Szilveszter[1], Vribék Krisztián[1], Süth Miklós[2], Czudor Zsófia[3], Tényi Ákos[3], Jóźwiak Ákos[1]

ÖSSZEFOGLALÁS

A tanulmányban a szerzők bemutatják egy egységes adatinfrastruktúra-rendszer kidolgozását az állatorvosi közegészségügyi és élelmiszerlánc-biztonsági adatok kezelésére, amelyet egy, az antimikrobiális rezisztencia elleni küzdelmet célzó projekt során készített repozíciós platform létrehozása alapoz meg. Az adatmodell lehetővé tette a releváns adatforrások adatbázisba integrálását, ezáltal az antibiotikumrezisztencia csökkentésére alkalmazható hatóanyagok adatvezérelt azonosítását. A jövőbeni cél egy adatinfrastruktúra rendszer létrehozása az állategészségügyi és élelmiszerlánc-biztonsági adatok integrálására átfogó, modern adatelemzési módszereken alapuló kutatásokhoz.

SUMMARY

Background: The complexity of veterinary public health and food chain safety is influenced by various factors, including multiple stakeholders and rapid dynamics. The emergence of antibiotic resistance poses a significant global health challenge, with projections indicating that antibiotic resistance could become the leading cause of death by 2050. This highlights the need to manage and analyze veterinary public health data effectively.

Objectives: To establish a data infrastructure system for veterinary public health and food chain safety data through the development of a repositioning platform. This platform aims to integrate relevant data sources, support the identification of antibiotic resistance countermeasures, and facilitate the creation of a future food safety data ecosystem.

Materials and Methods: The methods involved identifying and mapping various Hungarian, EU and international data sources related to animal health and food chain safety. Key tasks included prioritizing and preparing data sources for integration into a veterinary public health data lake. A logical data model was developed to define entities and attributes, and relationships between data elements were mapped to create a structured database. The data lake construction utilized the ETL process to extract, transform and load data, addressing the complexities of integrating diverse data types and file formats to enhance research and repositioning infrastructure.

Results and Discussion: An integrated database, a veterinary drug repositioning data lake was created, aiding in repositioning active ingredients for reducing antimicrobial resistance and facilitating data-driven decision-making. The findings indicate significant need for standardized and reliable data systems in veterinary public health. The established repositioning platform will support future research and innovations in antibiotic resistance management and food chain safety, providing a model for other stakeholders and contributing to a comprehensive understanding of the interconnectedness of health systems.

ÁLLATORVOSI KÖZEGÉSZSÉGÜGYI ÉS ÉLELMISZERLÁNC-BIZTONSÁGI ADATOK ADATINFRASTRUKTÚRA RENDSZERÉNEK MEGALAPOZÁSA REPOZÍCIÓS PLATFORM KIALAKÍTÁSÁVAL

Az állatorvosi közegészségügy és az élelmiszerlánc biztonsága rendkívül összetett rendszert érintő közegészségügyi és gazdasági tudományterület. Ennek jellemzői:

- magas szintű komplexitás (sok szereplő, sok terméket mozgat sok kapcsolaton keresztül);
- rendkívül gyors időbeli dinamika (az élelmiszerrendszerek tranzakciói egységnyi idő alatt nagyszámúak és időben változékonyak);
- sokrétű visszacsatolási mechanizmus és adaptáció (az egyik szereplő és/vagy termék hatással van a többi szereplőre, amelyek alkalmazkodnak a változó környezethez; a biológiai ágensek folyamatosan adaptálódnak a kockázatkezelési intézkedésekhez);
- sok külső hatásnak kitett (az élelmiszerrendszerek egy önmagában is rendkívül összetett külső környezetben működnek, amelyekre olyan tényezők hatnak, mint pl. gazdasági, jogi és politikai folyamatok, kutatás-fejlesztés-innováció, fogyasztói szokások, demográfiai és társadalmi változások, klímaváltozás stb.).

Az állatorvosi közegészségügy és az élelmiszerlánc biztonsága rendkívül összetett tudományterület

Ez a komplexitás a tudomány és a technológia jelenlegi fejlettségi szintjén nagyon nehézzé teszik a rendszer viselkedésének modellezését és előrejelzését. Ezért az elmúlt évtizedekben különböző egyszerűsítő, közelítő megoldások születtek. Korábban nem volt cél a (teljes) előrejelzés, inkább a háttérmechanizmusok megértésére törekedett a tudományterület, és ezek alapján valószínűségi megközelítéssel meghatározta, hogy mely veszélyeket kezelünk, és melyeket nem. Így jelenleg minden élelmiszerlánc-biztonsági felügyeleti rendszer kockázatalapon működik: azaz a rendszert érő különböző ágensek (kémiai, fizikai, biológiai veszélyek) hatása mellett azok – közelítő – valószínűségével is számol. Ez a paradigma áthatja a jelenlegi nemzetközi, EU-s és hazai jogi környezetet és a mindennapi ipari és hatósági döntéshozatali folyamatokat is, viszont az adatelemzési módszerek fejlődésével, az elemzési kapacitás rohamos fejlődésével és a hozzáférhető adatok mennyiségének növekedésével tér nyílt a fejlett előrejelzési kapacitások fejlesztése felé. Az adatalapú döntéshozatal növekvő jelentősége az állategészségügyben és az állatorvosi közegészségügyben egyre nagyobb szükségletet teremt a szabványosított és megbízható adat- és információs rendszerek iránt [1].

REPOZÍCIÓS PLATFORM MINT ELSŐ LÉPÉS AZ ÁLLATORVOSI KÖZEGÉSZSÉGÜGYI ÉS ÉLELMISZERLÁNC-BIZTONSÁGI ADATINFRASTRUKTÚRA-RENDSZEREK KIALAKÍTÁSÁHOZ

Az antibiotikumokkal szembeni rezisztencia napjaink egyik legnagyobb egészségügyi kihívása

Az antibiotikumokkal szembeni rezisztencia napjaink egyik legnagyobb egészségügyi kihívása. Az előrejelzések szerint 2050-re a multirezisztens fertőzések lesznek a vezető halálokok [2]. Az antibiotikumok felelőtlen használata mind az emberi, mind az állatgyógyászatban jelentősen hozzájárul a rezisztencia kialakulásához, miközben az új antibiotikumok fejlesztése lassú és költséges folyamat [3]. Az ECDC/EFSA/EMA (European Centre for Disease Prevention and Control/European Food Safety Authority/European Medicines Agency) jelentések [4] alapján szoros összefüggés van az élelmiszertermelő állatokban használt antibiotikumok és az emberi kórházakban található rezisztens mikróbák között, ami különösen fontossá teszi az „Egy egészség" (*One Health*) koncepciót [5].

E kihívásokra válaszul jött létre az „Antibakteriális rezisztencia kialakulását gátló termék fejlesztése az állatgyógyászatban, támogató repozíciós platform fejlesztése" című, Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal által támogatott (2020-1.1.2-PIACI-KFI-2021-00246) projektkezdeményezés, amelynek egyik célja egy innovatív „állategészségügyi repozíciós platform" létrehozása, ami lehetővé teszi az állatgyógyászati antibiotikumrezisztencia elleni hatóanyagok azonosítását.

A rendszer célja, hogy integrálja a releváns állategészségügyi, élelmiszer-biztonsági és humánegészségügyi adatokat, és alapot szolgáltasson egy későbbi

élelmiszerlánc-biztonsági adattó létrehozásához. A projekt kezdete idején az állategészségügyi területen sem Magyarországon, sem Közép-Európában nem létezett hasonló megoldás.

A projekt célja gyógyszer-repozíciós hatóanyagjelöltek heurisztikus és a rendszer által támogatott adatvezérelt azonosításán túl ezen jelöltek tesztelése *in vitro* és *in vivo* körülmények között. Cél továbbá szabadalomképes termék kifejlesztése, amely antibiotikumokkal kombinálva növelik azok hatékonyságát, és lassítják a rezisztencia kialakulását az állategészségügyben.

CÉLOK/SAJÁT KUTATÁS

Ez a tanulmány a bemutatott projekthez kapcsolódó, az állategészségügyi közegészségügyi adatok egységes adatinfrastruktúra rendszerének kialakítását mutatja be. A kialakított rendszer elsődleges feladata az antibiotikumrezisztencia elleni védelmet szolgáló hatóanyagjelöltek generálásának támogatása, ugyanakkor a kialakított adattárolási stratégia lehetőséget biztosít az állategészségügy és élelmiszerlánc-biztonsági kockázatainak integrált elemzésére, és modellként szolgálhat más országok és a közös európai állatorvosi közegészségügy számára.

A szerzők által létrehozott adatinfrastruktúra-rendszer elsődleges feladata az antibiotikumrezisztencia elleni védelmet szolgáló hatóanyagjelöltek generálásának támogatása

Emellett a platform alapjául szolgálhat egy későbbi integrált élelmiszerlánc-biztonsági adattónak, amely integrálja az állategészségügyi, közegészségügyi és élelmiszer-biztonsági adatokat, valamint helyettesítheti az eseti, heterogén adatgyűjtéseket.

Az Állatorvosi Közegészségügyi Adattó adatközpontban különböző publikus farmakológiai, kémiai, omikai (pl. transzkriptomikai, mikrobiomikai) és fenotípusos adatbázisok, valamint nemzeti, EU-s és nemzetközi takarmányozási, állatgyógyászati és laboratóriumi adatbázisok kerülnek integrálásra. Ezek az adatok támogatják az antibiotikumrezisztencia statisztikai és epidemiológiai elemzéseit, és segítenek az elemzéseket nemzeti kontextusba helyezni. Egy olyan környezet kerül kialakításra, amely demokratizálja a legmodernebb elemzési módszereket, segíti az adatok értékelését és repozíciós jelöltek generálását. Ez lehetővé teszi a kutatók számára, hogy egyszerűbben és gyorsabban találjanak potenciális repozíciós jelölteket [6].

ANYAG ÉS MÓDSZER

A projekt során azonosításra és feltérképezésre kerültek a magyar, EU-s és nemzetközi statisztikai, hatósági és egyéb releváns publikus és nem publikus adatforrások. Adatgyűjtéseken (mérések, megfigyelések) alapuló állategészségügyi és élelmiszer-biztonsággal kapcsolatos adatokon (pl. EU-engedélyezett takarmányadalékanyag-regiszter) túl bioinformatikai, genomikai és molekuláris kémiai adatbázisok (pl. ChEMBL) [7], valamint természetes hatóanyagok adatbázisai (pl. NPASS) is górcső alá kerültek. Vizsgáltuk az elérhető adatok mennyiségét, formátumát, frissítési gyakoriságát és struktúráját, valamint az adatok közötti kapcsolatokat. Nem publikus adatokat is bevontunk, pl. a NÉBIH-től kapott állatgyógyászati készítmények adatait.

ADATFORRÁSOK PRIORIZÁLÁSA, RELEVÁNS ADATFORRÁSOK KIVÁLASZTÁSA, ADATELŐKÉSZÍTÉS

A legfontosabb adatkörök az állatgyógyászati készítményekre, valamint a takarmány-adalékanyagokra vonatkozó hazai és EU-s adatbázisok

Az adatforrásokat priorizáltuk a repozíciós platformba való betöltés és az Állatorvosi Közegészségügyi Adattó hosszú távú stratégiai fontossága alapján. Az adatforrásokban az adatmezők pontos definiálása volt a betöltésük előkészítésének első lépése. A legfontosabb adatkörök az állatgyógyászati készítményekre, valamint a takarmány-adalékanyagokra vonatkozó hazai és EU-s adatbázisok. Ezek az adatok jelenleg rosszul vagy egyáltalán nem strukturáltak, ami nehezíti a gépi feldolgozást. Ezért további adatbetöltés-előkészítési és adattisztítási

feladatokat végeztünk, egyes esetekben kézi adatbetöltéssel. Így olyan egyedülálló adatbázisok jöttek létre, amelyek önmagukban is fontos kutatási és repozíciós infrastruktúrát jelentenek.

ADATMODELL KIALAKÍTÁSA

Az adatmodell logikai kialakítása kulcsfontosságú lépés

Az adatmodell logikai kialakítása kulcsfontosságú lépés az állatorvosi közegészségügyi adattó létrehozásában. A logikai tervezés során az adatmodellezők elméleti szinten határozzák meg az adatmodell struktúráját, ami magában foglalja az összes érintett entitás és azok közötti kapcsolatok azonosítását.

- *Entitások azonosítása*: Az entitások olyan objektumok vagy fogalmak, amelyekről adatokat szeretnénk gyűjteni és tárolni. Az állatorvosi közegészségügyi adattó esetében entitás pl. egy termék vagy vegyület.
- *Attribútumok meghatározása*: Minden egyes entitáshoz meghatározásra kerülnek azok az attribútumok, amelyek leírják az entitás jellemzőit. Az állatorvosi közegészségügyi adattó esetén a termékhez tartozó attribútumok pl.: termékazonosító, márkanév, tudományos (latin) név, élelmiszer- vagy adalékanyag-kategória stb. Vegyület entitás esetén attribútumok pl.: azonosító, szerkezet, inchi key stb.

ADATELEMEK KÖZÖTTI KAPCSOLATOK FELTÉRKÉPEZÉSE

Az adatelemek közötti kapcsolatok feltérképezésekor a logikai adatmodell fizikai leképezése történik meg – kialakul egy adatbázis-struktúra, amelyet ténylegesen implementálni lehet a későbbiek során az adattóban. A logikai adatmodell entitásait táblákká alakítjuk, minden egyes entitáshoz egy tábla tartozik. Pl. a termék entitásból a „termék” tábla, a vegyület entitásból a „vegyület” tábla kerül kialakításra. Az egyes táblák mezőit pedig az attribútumok alapján határozzuk meg. Minden mező egy-egy adatdarabot (entitásjellemzőt azaz attribútumot) reprezentál. Pl. a „termék” tábla mezői a fenti példa alapján a termékazonosító, márkanév, tudományos (latin) név, élelmiszer-kategória, adalékanyag-kategória stb. lesznek.

A következő lépés a kulcsok meghatározása. Elsődleges kulcsok, amelyek egyedileg azonosítják a tábla rekordjait (pl. „termék” tábla esetén a termékazonosító). A kulcsok segítségével történik meg a táblák (entitások) közötti kapcsolatok meghatározása.

ADATTÓ KIALAKÍTÁSA

Az adattó kialakítása az előzetesen specifikált adatkapcsolatok fizikai leképezését jelenti

Az adattó kialakítása tulajdonképpen az előzetesen specifikált adatkapcsolatok fizikai leképezését jelenti adatbázis formájában, másképp fogalmazva a releváns adatforrások integrációját a közöttük levő kapcsolatok mentén. Az adatforrások integrációja az állategészségügy, az élelmiszerlánc és a hatóanyag-repozíció területeinek komplexitása miatt rendkívül bonyolult feladat nemcsak informatikai technológiai (különböző bemeneti adattípusok, fájlformátumok pl. PDF, Excel, HTML) hanem szakmai szempontból is (pl. takarmány-adalékanyagok, növényi hatóanyagok, élelmiszerek stb. adatbázisainak logikai összekapcsolása). Az adatforrások adatmodell alapján történő betöltése az Extract, Transform, Load (ETL) logika mentén történik, ennek lépései a következők (a programozástechnikai részleteket jelen tanulmány nem tartalmazza):

- adatok kinyerése
- staging (tárolás nyers adatzónában) és transzformálás
- adatok betöltése az adatmodell sémába

EREDMÉNYEK

RENDELKEZÉSRE ÁLLÓ ADATFORRÁSOK FELTÉRKÉPEZÉSE

A projekt során azonosított és feltérképezett adatforrásokból létrejött két belső felhasználású adatbázis: az egyik több mint 190 különböző adatforrást tartalmaz az állategészségügy, élelmiszer- és takarmánybiztonság területéről, ill. külön adatbázis tartalmazza a természetes hatóanyagok adatforrásait ezek számossága miatt.

ADATFORRÁSOK PRIORIZÁLÁSA, RELEVÁNS ADATFORRÁSOK KIVÁLASZTÁSA, ADATELŐKÉSZÍTÉS

Az adatforrások kiválasztása relevanciájuk alapján történt, priorizálásuk azonban több szempont mentén:

- adatbázis mérete
- megbízhatósága
- aktualitása (frissítése)
- elérhetősége (ingyenesen elérhető)
- adatok letölthetősége
- formátum

A priorizálás irányt adott a véglegesen betöltésre kerülő adatforrások kiválasztásához, ugyanakkor bizonyos esetekben az akadályoztatott adatlehívás ellenére is kiválasztásra kerültek bizonyos adatbázisok az extrém szakmai relevancia miatt. Ilyenek pl. az EMA és a European Feed Additives adatbázisok, amelyekben szereplő adatok nehezen hozzáférhetők (pl. PDF formátum), ugyanakkor állatorvosi gyógyszer-repozíció szempontjából betöltésük elengedhetetlen volt.

A priorizálást követően az alábbi adatforrások kerültek kiválasztásra és betöltésre a projekt végéig (részletes jellemzés az adatbázisokról a *Táblázatban* található):

• *PubChem [8]:* A PubChem egy ingyenes, nyilvánosan elérhető adatbázis, amelyet az Amerikai Egyesült Államok Nemzeti Egészségügyi Intézeteinek (NIH, National Institute of Health) Kémiai Biológiai Központja tart fenn. Az adatbázis kémiai anyagok, biológiai aktivitások és egyéb kapcsolódó információk széles skáláját tartalmazza. A kutatók és szakemberek számára rendkívül hasznos eszköz, mivel lehetővé teszi a vegyületek tulajdonságainak, szerkezeteinek és biológiai aktivitásának gyors és hatékony keresését és elemzését. A PubChem adatbázis fontos az állatorvosi gyógyszer-repozíció szempontjából, mert rengeteg információt tartalmaz különböző vegyületekről és azok biológiai aktivitásáról. Az adatbázis lehetővé teszi a kutatók számára, hogy gyorsan hozzáférjenek a kémiai és biológiai adatokhoz, így felgyorsítva a meglévő humán gyógyszerek állatok kezelésére történő újrahasznosításának folyamatát.

• *ChEMBL [9]*: A ChEMBL egy nyilvánosan elérhető adatbázis, amelyet az Európai Bioinformatikai Intézet (EMBL-EBI, European Molecular Biology Laboratory, European Bioinformatics Institute) tart fenn. Az adatbázis bioaktív vegyületek, azok biológiai aktivitásának és farmakológiai tulajdonságainak átfogó gyűjteménye. A ChEMBL különösen hasznos a gyógyszerkutatásban és -fejlesztésben, mivel lehetővé teszi a kutatók számára, hogy gyorsan és könnyen hozzáférjenek fontos kémiai és biológiai adatokhoz, amelyek segíthetnek új gyógyszerek azonosításában és optimalizálásában [10].

• *CMAUP [11]:* A CMAUP (Comprehensive Medicinal Chemistry and Activity-Unified Platform) egy nyilvánosan elérhető adatbázis, amely integrált információkat tartalmaz bioaktív vegyületekről és azok farmakológiai aktivitásáról. Az adatbázis célja, hogy segítse a kutatókat a gyógyszerkutatás és -fejlesztés területén azáltal, hogy könnyen hozzáférhetővé teszi a különböző forrásokból származó kémiai és biológiai adatokat. A CMAUP hasznos eszköz lehet új gyógyszerek azonosításában és a meglévő vegyületek hatékonyságának és biztonságosságának értékelésében.

• *DisGeNET [12]:* A DisGeNET egy integrált platform, amely emberi betegségekkel kapcsolatos gén- és génváltozat-asszociációkat tartalmaz. Az adatbázis összegyűjti és rendszerezi a különböző forrásokból származó genetikai és genomikai adatokat, hogy segítsen a kutatóknak megérteni a betegségek genetikai hátterét. A DisGeNET hasznos eszköz a betegségek molekuláris mechanizmusainak feltárásában, új diagnosztikai markerek és terápiás célpontok azonosításában [13].

• *Drug Central [14]:* A DrugCentral egy nyilvánosan elérhető adatbázis, amely átfogó információkat tartalmaz jóváhagyott gyógyszerekről és azok biológiai célpontjairól. Az adatbázis célja, hogy segítse a kutatókat és az egészségügyi

TÁBLÁZAT. *Állatorvosi közegészségügyi adattóba integrált adatbázisok*

TABLE. *Databases integrated to veterinarian public health datalake*

Adatbázis/ Forrás neve	Elérhető adatok	Adat tulajdonos	Adat-formátum	Termék/Mátrix	Vegyi szennyező	Földrajzi lefedettség	Frissítési gyakoriság	Adatelérés	Megjegyzések
PubChem	Vegyi vegyületek, bioaktivitás, toxicitás, gyógyszer célpontok	NIH (USA)	SDF, XML, CSV, JSON	Kis molekulák, gyógyszerek, biológiai szerepek	Különféle	Globális	Napi	Ingyenes, nyilvános	NIH által fenntartott vegyi adatbázis
ChEMBL	Bioaktív molekulák, gyógyszer célpontok, ADMET adatok	EMBL-EBI (EU)	SDF, CSV, JSON	Gyógyszerek, bioaktív vegyületek	Gyógyszerek, metabolitok	Globális	Negyed-éves	Ingyenes, nyilvános	EMBL-EBI kezelésében
CMAUP	Gyógynövények, természetes anyagok, etnofarmakológia	BIDD Group (Kína)	CSV, Excel	Gyógynövényi készítmények	Fito-vegyületek	Globális (főként Ázsia)	Éves	Ingyenes, nyilvános	Hagyományos gyógyászatra fókuszál
DisGeNET	Gén-betegség asszociációk, gyógyszer-betegség kapcsolatok	ICBI (Spanyolország)	CSV, TSV, JSON	Gének, betegségek, gyógyszerek	N/A	Globális	Féléves	Ingyenes (regisztrációval)	Több forrás integrálása
Drug Central	Gyógyszer célpontok, indikációk, farmakológiai adatok	University of New Mexico (USA)	CSV, JSON	Jóváhagyott gyógyszerek	Gyógyszerek	Globális (FDA/EMA jóváhagyott)	Havi	Ingyenes, nyilvános	Az University of New Mexico fenntartja
EMA (Növényi gyógyászati termékek)	EU által jóváhagyott növényi gyógyszerek, monográfiák	Európai Gyógyszerügyi Ügynökség (EU)	PDF, HTML	Növényi termékek	Növényi kivonatok	Európai Unió	Változó	Ingyenes, nyilvános	Szabályozó hatóság adatbázisa
Európai Takarmány-adalékanyagok	EU által jóváhagyott takarmányadalékok, biztonsági adatok	EFSA (EU)	PDF, Excel, HTML	Állati takarmány	Adalék-anyagok, szennyező-dések	Európai Unió	Negyed-éves	Ingyenes, nyilvános	EFSA kezelésében
Európai Élelmiszer-adalékanyagok	EU által jóváhagyott élelmiszer-adalékanyagok, biztonsági értékelések	EFSA (EU)	PDF, Excel, HTML	Élelmiszerek	Adalék-anyagok, tartósító-szerek	Európai Unió	Féléves	Ingyenes, nyilvános	EFSA szabályozta
FooDB	Élelmiszer összetétel, tápanyagok, bioaktív vegyületek	The Metabolomics Innovation Centre (Kanada)	CSV, JSON	Élelmiszer termékek	Természetes vegyületek	Globális	Éves	Ingyenes, nyilvános	Átfogó élelmiszerkémiai adatbázis
Európai Állatgyógyászati Termékek	EU által jóváhagyott állatgyógyászati készítmények, biztonsági adatok	Európai Gyógyszerügyi Ügynökség (EU)	PDF, HTML	Állat-gyógyászati készítmények	Gyógyszerek	Európai Unió	Havi	Ingyenes, nyilvános	EMA szabályozta
HERB	Hagyományos kínai gyógyszer (TCM) vegyületek, célpontok	Kínai Tudományos Akadémia (Kína)	CSV, Excel	Gyógynövényi készítmények	Fito-vegyületek	Kína	Éves	Ingyenes, nyilvános	TCM-re fókuszál
NÉBIH: Magyarországon engedélyezett állatgyógyászati készítmények	Magyarországon engedélyezett állatgyógyászati készítmények	Nébih (Magyarország)	PDF, HTML	Állat-gyógyászati készítmények	Gyógyszerek	Magyar-ország	Negyed-éves	Nem nyilvános (magyar)	Nemzeti Élelmiszerlánc-biztonsági Hivatal
NÉBIH: Állatgyógyászati készítmények, ápolószerek	Magyarországon nyilvántartott állatgyógyászati készítmények, ápolószerek	Nébih (Magyarország)	PDF, HTML	Állat-gyógyászati készítmények, ápolószerek	Gyógyszerek	Magyar-ország	Negyed-éves	Nem nyilvános (magyar)	Hivatalos magyar nyilvántartás
NCBI Taxonomy	Organizmusok osztályozása, tudományos nevek	NIH (USA)	TXT, XML, JSON	Biológiai fajok	N/A	Globális	Heti	Ingyenes, nyilvános	NIH által fenntartott taxonómiai adatbázis
NPASS	Természetes anyagok, fajforrások, bioaktivitás	BIDD Group (Kína)	CSV, Excel	Természetes anyagok	Fito-vegyületek	Globális	Éves	Ingyenes, nyilvános	Természetes anyagok aktivitására fókuszál
MicroBIGG-E	Mikrobiális genomika, antibiotikum-rezisztencia gének	NIH (USA)	FASTQ, GenBank, CSV	Mikrobák, plazmidok	Anti-biotikumok	Globális	Féléves	Ingyenes, nyilvános	NCBI BioProject mikrobális rezisztenciáról
CARD (Átfogó Antibiotikum-rezisztencia Adatbázis)	Antibiotikum-rezisztencia gének, mechanizmusok	McMaster Egyetem (Kanada)	CSV, FASTA, JSON	Kórokozók, plazmidok	Anti-biotikumok	Globális	Negyed-éves	Ingyenes, nyilvános	McMaster Egyetem adatbázisa
UMLS (Egységes Orvosi Nyelvi Rendszer)	Orvosi terminológiák, ontológiák, szemantikus leképezések	NIH (USA)	RRF, CSV, XML	Orvosi fogalmak	N/A	Globális	Éves	Ingyenes (licenc szükséges)	NIH által fenntartott orvosi terminológiai rendszer
UniChem	Kémiai azonosítók közötti keresztreferenciák	EMBL-EBI (EU)	CSV, TSV	Kémiai vegyületek	N/A	Globális	Havi	Ingyenes, nyilvános	EMBL-EBI kémiai azonosító leképezés

szakembereket a gyógyszer-információk gyors és könnyű elérésében, beleértve a kémiai szerkezeteket, terápiás felhasználásokat, mellékhatásokat és interakciókat. A DrugCentral hasznos eszköz a gyógyszerkutatásban és -fejlesztésben, valamint a klinikai gyakorlatban, mivel megbízható és naprakész információkat nyújt a gyógyszerekről [15].

• *EMA (herbal medicine) [16]:* Az EMA herbal medicine adatbázisa egy nyilvánosan elérhető forrás, amely információkat tartalmaz gyógynövényalapú készítményekről. Az adatbázis célja, hogy támogassa a gyógynövény alapú gyógyszerek biztonságos és hatékony használatát azáltal, hogy részletes adatokat nyújt a növényi összetevőkről, azok terápiás alkalmazásairól, mellékhatásairól és klinikai vizsgálati eredményeiről. Az EMA herbal medicine adatbázis hasznos eszköz a kutatók, egészségügyi szakemberek és a nyilvánosság számára, mivel átfogó és megbízható információkat biztosít a gyógynövényekkel kapcsolatosan.

• *European Feed Additives [17]:* Az European Feed Additives egy nyilvánosan elérhető adatbázis, amelyet az Európai Élelmiszer-biztonsági Hatóság (Europaean Food Safety Agency, EFSA) tart fenn. Az adatbázis részletes információkat tartalmaz az Európai Unióban engedélyezett takarmány-adalékanyagokról, beleértve azok összetételét, funkcióit, biztonságossági értékeléseit és engedélyezési státuszát. Az adatbázis célja, hogy támogassa az állati takarmányozással kapcsolatos biztonságos és hatékony gyakorlatokat, és segítse a szakembereket és hatóságokat a szabályozási döntéshozatalban. Az European Feed Additives adatbázis hasznos eszköz a mezőgazdasági termelők, állatorvosok és kutatók számára.

• *European Food Additives [18]:* Az European Food Additives egy nyilvánosan elérhető adatbázis, amelyet az Európai Élelmiszer-biztonsági Hatóság (EFSA) kezel. Az adatbázis átfogó információkat tartalmaz az Európai Unióban engedélyezett élelmiszer-adalékanyagokról, beleértve azok kémiai összetételét, funkcióit, biztonságossági értékeléseit és jogi státuszát. Az adatbázis célja, hogy támogassa az élelmiszer-biztonsági szabályozást, valamint segítse a gyártókat, kutatókat és egészségügyi szakembereket az élelmiszer-adalékanyagok biztonságos és hatékony használatában. Az European Food Additives adatbázis megbízható és naprakész információkat nyújt az élelmiszer-adalékanyagokkal kapcsolatban.

• *FooDB [19]:* A FooDB egy átfogó adatbázis, amely információkat tartalmaz az élelmiszerekben található kémiai vegyületekről. Az adatbázis részletes adatokat nyújt az élelmiszerek összetevőiről, beleértve a tápanyagokat, bioaktív vegyületeket, valamint azok kémiai és biológiai tulajdonságait. A FooDB célja, hogy támogassa a kutatókat és az egészségügyi szakembereket az élelmiszerek kémiai összetételének és egészségügyi hatásainak megértésében. Az adatbázis hasznos eszköz lehet az élelmiszer-tudomány, táplálkozástudomány és élelmiszer-biztonság területén dolgozók számára, mivel megbízható és részletes információkat biztosít az élelmiszerekben található vegyületekről.

• *European Veterinary Medicinal Products [20]:* Az European Veterinary Medicinal Products adatbázis az Európai Unióban engedélyezett állatgyógyászati készítmények átfogó gyűjteménye. Az adatbázist az EMA kezeli, és részletes információkat nyújt az állatgyógyászati termékekről, beleértve azok összetételét, terápiás alkalmazásait, adagolását, mellékhatásait és biztonságossági értékeléseit. Az adatbázis célja, hogy támogassa az állatorvosokat, kutatókat és szabályozó hatóságokat a biztonságos és hatékony állatgyógyászati termékek használatában és engedélyezésében. Az European Veterinary Medicinal Products adatbázis megbízható és naprakész információkat biztosít az állatgyógyászat területén.

• *HERB [21]:* A HERB (Herbal Extracts and Phytochemicals Resource Base) egy nyilvánosan elérhető adatbázis, amely kifejezetten gyógynövényekkel és azok kivonataival kapcsolatos információkat tartalmaz. Az adatbázis részletes adatokat nyújt a különböző gyógynövények kémiai összetevőiről, biológiai aktivitásáról, és potenciális terápiás alkalmazásairól. A HERB célja, hogy támogassa a kutatókat,

egészségügyi szakembereket és a nyilvánosságot a gyógynövényekkel kapcsolatos tudományos ismeretek elérésében és alkalmazásában. Az adatbázis hasznos eszköz lehet a gyógynövénykutatásban, gyógyszerfejlesztésben és az alternatív gyógyászatban érdekeltek számára, mivel megbízható és részletes információkat biztosít a gyógynövényekről és azok hatóanyagiról.

• *NÉBIH: Magyarországon engedélyezett állatgyógyászati készítmények:* A NÉBIH (Nemzeti Élelmiszerlánc-biztonsági Hivatal) által üzemeltetett adatbázis Magyarországon engedélyezett állatgyógyászati készítményekkel foglalkozik. Az adatbázisban megtalálhatók az állatgyógyászati készítmények engedélyezésével, forgalomba hozatalával és felhasználásával kapcsolatos információk. Ezek az adatok lehetnek pl. az engedélyes neve, az engedély száma, a készítmény neve és típusa, valamint az alkalmazási területek és előírások. A NÉBIH által működtetett adatbázis fontos információforrás lehet az állatgyógyászati termékekkel kapcsolatos szabályozások, engedélyeztetési folyamatok és jogszabályi követelmények terén Magyarországon. Az adatbázis nem elérhető nyilvánosan, elemzési és kutatási céllal került beépítésre az adattóba.

• *NÉBIH: Magyarországon nyilvántartásba vett, állatgyógyászatban használatos gyógyhatású készítmények, állatgyógyászati ápolószerek:* A NÉBIH által nyilvántartott adatbázisban szerepelnek Magyarországon nyilvántartásba vett állatgyógyászatban használt gyógyhatású készítmények és állatgyógyászati ápolószerek. Ezek a termékek különböző formában és összetevőkkel rendelkeznek, és az állatok egészségügyi problémáinak kezelésére vagy megelőzésére szolgálnak. A nyilvántartott termékekkel kapcsolatos adatok között szerepelhetnek az engedélyes neve, a készítmény típusa, összetétele, alkalmazási területek, előírások és egyéb fontos információk, amelyek segítségével biztonságosan és hatékonyan alkalmazhatók az állatokon. Az adatbázis nem elérhető nyilvánosan, elemzési és kutatási céllal került beépítésre az adattóba.

• *NCBI taxonomy [22]:* Az NCBI Taxonomy adatbázis a National Center for Biotechnology Information (NCBI) által fenntartott rendszer, amely a biológiai szervezetek taxonómiai osztályozását tartalmazza. Ez az adatbázis átfogó információt nyújt a különböző fajokról, beleértve azok tudományos nevét, hierarchiáját és kapcsolatait más fajokkal. Az adatbázis célja, hogy támogassa a kutatást és az adatcserét, valamint egységes keretet biztosítson a biológiai adatok rendszerezéséhez. Az NCBI Taxonomy adatbázis széleskörűen használt a genomikai és bioinformatikai kutatások során.

• *NPASS [23]:* Az NPASS az NCBI PubChem adatbázis egyik forrása, amely természetes anyagok aktivitását és azok forrásait gyűjti össze. Az adatbázisban található információk segítségével lehetőség van természetes eredetű anyagok biológiai aktivitásának vizsgálatára és az ezekkel kapcsolatos kutatásokra. A NPASS adatbázis hasznos forrás lehet a gyógyszerfejlesztés, a kémiai biológia és a gyógynövényekkel kapcsolatos kutatások terén.

• *MicroBIGG-e [24]:* A MicroBIGG-E adatbázis a mikrobiális antimikrobiális rezisztencia (AMR) gének és egyéb stressz- és rezisztenciaelemek böngészésére szolgál. Célja, hogy segítse a kutatókat a rezisztencia gének azonosításában és elemzésében, amelyek fontos szerepet játszanak a rezisztens kórokozók vizsgálatában. Ez az adatbázis hasznos eszköz az AMR mechanizmusok és a mikrobiális adaptációk mélyebb megértéséhez.

• *CARD (Comprehensive Antibiotic Resistance Database) [25]:* egy átfogó adatbázis, amely az antimikrobiális rezisztencia géneket és azok molekuláris mechanizmusait tartalmazza. A CARD célja, hogy pontos és naprakész információkat nyújtson az antibiotikumrezisztencia génekről, azok funkcióiról, valamint a rezisztencia terjedésének módjairól. Az adatbázis bioinformatikai eszközöket is biztosít a kutatók számára, lehetővé téve a rezisztencia gének azonosítását és a genomikus adatok elemzését. A CARD fontos erőforrás az AMR kutatásban és a közegészségügyi stratégiák fejlesztésében.

• *UMLS (Unified Medical Language System) [26]:* egy átfogó orvosi nyelvi rendszer, amelyet az Egyesült Államok Nemzeti Orvosi Könyvtára (NLM) fejlesztett ki. Az UMLS célja, hogy összekapcsolja és integrálja a különböző orvosi és egészségügyi terminológiákat, megkönnyítve ezzel az információk megosztását és az interoperabilitást az egészségügyi rendszerek között. Az UMLS három fő komponenst tartalmaz: a Metathesaurust, amely különböző fogalmakat és azok kapcsolatait foglalja magában; a Semantic Networket, amely a fogalmak közötti kapcsolatok szemantikai kontextusát adja meg; és a SPECIALIST Lexicont, amely az orvosi nyelv feldolgozását segíti.

• *UniChem [27]:* egy kémiai források közötti adatkapcsolati rendszer, amelyet a European Bioinformatics Institute (EBI) fejlesztett ki. Az UniChem célja, hogy összekapcsolja a különböző kémiai adatbázisok azonosítóit, ezáltal megkönnyítve a vegyületek keresését és az adatok integrálását. Az UniChem nem tárol kémiai adatokat vagy struktúrákat, hanem az azonosítók közötti kapcsolatokat biztosítja, lehetővé téve a kutatók számára, hogy gyorsan hozzáférjenek a különböző forrásokban található információkhoz. Ez az eszköz különösen hasznos a gyógyszerkutatásban és a kémiai biológia területén.

ADATMODELL KIALAKÍTÁSA

Jelen publikációban az adatmodell egységes adatstruktúrát támogató jellege kerül bemutatásra, további részletes leírására az elméleti és technikai jellemzőkkel külön tanulmány készül. A kialakított modell képes tárolni az összes olyan biológiai és kémiai kapcsolatot, amelyre szükség van egy általános gyógyszerkutatási pipeline, ezen belül természetesen az állatorvosi gyógyszer-repozíció támogatásához.

A kialakított adatmodell képes tárolni az összes olyan biológiai és kémiai kapcsolatot, amelyre szükség lehet az állatorvosi gyógyszer-repozíció támogatásához

Az adatmodell meghatározásában központi cél volt a domén szakértők támogatása, a tudományos evidenciák hálózatos reprezentációja, a természetes nyelvfeldolgozási módszerek szövegtest és szótár képzésének biztosítása és az adatforrások eredetének és így a felhasználási jogi feltételek visszakövethetősége. Az elkészített modellből gráf adatbázis is generálható. Ez a fajta kutatási környezet felhasználóbarát és könnyen kezelhető, nem adattudománnyal foglalkozó szakértők is könnyen elsajátíthatják a működését és saját kutatási kérdéseik megválaszolására használhatják. A kitűzött célok közül tehát kiváltképp jellegzetes eredmény az adatmodell hálózatos jellege és az ebből adódó gráfgenerálási potenciál. A hálózatos jelleg továbbá a természetes nyelvfeldolgozási lehetőségeket magasabb szinten is támogatja, hiszen az evidenciák hálózatán keresztül lehetséges definiálni és lekérni egy dokumentum hálózatot. Nyelvfeldolgozási megfontolás volt továbbá egy szinonimatábla létrehozása, ahol az összes elsődleges entitás szinonimái és a szinonimaütközések is (azonos alakú szavak) tárolhatóak, ezzel segítve a névelem felismerési hibák könnyebb azonosítását.

Adatelemek közötti kapcsolatok feltérképezése

Az adatfeldolgozási folyamat során kiemelt fontosságú az adatforrások precíz és átgondolt felhasználása, amely a részletes adatmodell leképezésén alapul. Az adatmodell segít abban, hogy az entitásokat és az evidenciákat pontosan meghatározzuk, és világosan definiáljuk az ezek közötti kapcsolatokat és kulcsokat. Ez a strukturált megközelítés biztosítja, hogy az adatforrások zökkenőmentesen integrálódjanak a betöltési folyamatokba, ami elengedhetetlen a hatékony adattárolás és -elemzés szempontjából.

Az entitások és az ezekhez kapcsolódó adatok közötti kapcsolatok feltérképezése lehetővé teszi az adatok konzisztens és következetes kezelését. Az integráció során alkalmazott kulcsok és asszociációk meghatározása biztosítja, hogy az adatok összekapcsolása pontosan és megbízhatóan történjen. A kulcsok maguk

az azonosítók, amelyek egyedi rekordazonosítást vagy táblák közötti hivatkozást biztosítanak. Az asszociációk ezeken a kulcsokon keresztül valósulnak meg, tehát míg a kulcsok technikai eszközök, az asszociációk a kapcsolatok logikai jelentését írják le. A kapcsolatok helyességét nem kézi ellenőrzéssel, hanem az adatmodell alapos tervezésével és az automatizált betöltési folyamattal biztosítottuk. Ez a megközelítés elősegíti a különböző forrásokból származó adatok harmonikus egyesítését, minimalizálva az adatvesztést és az inkonzisztenciát.

Az így kialakított rendszer nemcsak az adatok integritását védi, hanem hozzájárul a vállalati döntéshozatal támogatásához is azáltal, hogy pontos és naprakész információkat biztosít a felhasználók számára. A gondosan megtervezett adatfeldolgozási folyamatok tehát alapvető szerepet játszanak a szervezet sikeres adatmenedzsmentjében és üzleti intelligencia stratégiájában.

Adattó kialakítása

Adatok kinyerése

Több esetben technikai kihívást okozott, hogy weboldalak tartalmát kellett letölteni, vagy PDF fájlokban található szöveget kellett felismerni és beolvasni. Ezen források tartalmukat tekintve rendkívül fontosak, pl. az Európai Unió által publikált takarmányadalékanyag-regiszter (European Feed Additives). Az állatorvosi gyógyszer-repozíció szempontjából releváns integrált adattó kialakításán túl ezen adatforrások elemezhető formába történő betöltése önmagában is értéket képvisel az élelmiszerlánc érintett szereplői, ill. kutatók számára, ugyanakkor megjegyzendő, hogy ezen teljes egészében történő kinyerése túlmutat a jelen tanulmány céljain, ahol a gyógyszerrepozíciós célú adatokra fókuszáltunk. Az adatbázisok jelentős részénél eredendően kiváló adatminőség és könnyű betölthetőség volt jellemző.

Staging és transzformálás

Staging másnéven nyers adatzónában történt az adatok első relációs formátumban történő tárolása, ekkorra már nagyvonalakban megtörtént az adattisztítás, de a betöltött adatok még nem kerültek módosításra, legfeljebb kiegészítésre a további folyamatok segítése érdekében. Az adattisztítás jelen esetben főként azon adatbázisrészek elhagyását, ill. ún. „child-adatbázisban" való eltárolását jelentette, amelyek nem szolgálták a specifikus, állategészségügyhöz és élelmiszerlánchoz köthető adattó konkrét célját, az antimikrobiális gyógyszerrepozíciót. Az ide betöltött adatokon már végrehajthatók azok a transzformálások összekapcsolások és egyéb műveletek, amelyek a betöltéshez kellenek.

Biztonsági mentés céljából ebben a fázisban eltárolásra kerültek bizonyos adatforrások arra az esetre, ha a forrásban valamilyen változás történne.

Adatbetöltés

Az adatbetöltéshez olyan függvények kerültek fejlesztésre, amelyek lehetővé teszik a mérnöki területen kevésbé jártas szakembereknek is az adatbázis frissítését. Ennek lényege, hogy minden entitás feltöltéséhez készült egy procedúra, ami feltölti a céltáblát, egyéb szükséges táblákat és automatikusan végrehajt számos átalakítási műveletet, a felhasználónak a bemeneti nyers adattáblát kell előkészítenie, valamint néhány paramétert megadnia.

A betöltés részben automatizált a frissítések tekintetében is, amelyeket manuálisan kell indítani, de automatizáltan futnak le az adatforrások eltérő frissítési gyakorisága miatt.

Dokumentáció

A dokumentációt tevékenységnapló és egy testreszabott adatbázis-dokumentáló külső eszköz (https://schemaspy.org/) végzi. Az adattó létrehozására szolgáló létrehozott dokumentáció tartalmazza az adattáblák, mezők, valamint, függvények leírását is, ami ezáltal adatkatalógusként is funkcionál.

Felhasználási jogi szabályozás visszakövethetőség
A betöltött adatforrásaink erősen különböző a felhasználásukat szabályzó jogi háttérrel bírnak. Akadémiai kutatási célra mind felhasználhatók, viszont a termékbe építés egyes esetekben engedékenyek más esetekben tárgyaláshoz kötöttek. Ezért az adatmodell minden entitás estén követi, hogy melyik forrásból származik és annak a jogi forrásnak mik a jogi felhasználási feltételei. Ezzel automatizálható az adatok jogszerű felhasználása.

MEGVITATÁS

Az állatorvosi közegészségügy és élelmiszerlánc-biztonság eddig nem használta ki az adatalapú döntéshozatal lehetőségeit sem hazánkban, sem az EU-ban. Az Európai Élelmiszer-biztonsági Hatóság (EFSA) ajánlásokat tett egy közös „European food safety data model" bevezetésére [28]. A jelenlegi relációs adatbázisok mellett szükséges a közép/hosszútávú megoldások feltérképezése, és kidolgozni a „future food safety data ecosystem of Europe"-ot. Az új repozíciós adattóplatform nemzetközileg is versenyképes és támogatja az állatgyógyászati, antibiotikum-rezisztencia elleni innovációt. A platform jövőbeli szerepe túlmutat ezen, lehetőséget biztosítva az élelmiszerlánc integrált elemzésére és biztonsági kockázatainak feltárására. A projekt megvalósulása után az adattó további adatforrások és elemzési-kutatási területek bevonására szolgálhat. Hosszú távú cél egy platform létrehozása az állategészségügyi, élelmiszer-biztonsági és humánegészségügyi adatok elemzésére, támogatva az agrár- és élelmiszer-gazdaság szereplőit. Jelenleg a bioinformatikai repozíciót nehézkes adatintegráció és adatelemzés jellemzi. Az új platform könnyíti a repozíciós stratégiák megvalósítását, csökkentve az időt és növelve a sikerességet. Az integratív repozíciós adatelemzést támogató adatbázis és az AI alapú módszerek széles körben hasznosítható terméket eredményeznek. Az új platform komponensei unikális termékként jelennek meg a piacon, licenszelhető formában gyógyszergyártó és takarmány-előállító cégek számára, valamint saját adatplatformként is értékesíthető, míg az adattó módszertani kialakítása publikus, főként a tudományos közönséget szolgáló ismeretanyagnak tekinthető.

Az új repozíciós adattó platform nemzetközileg is versenyképes és támogatja az állatgyógyászati, antibiotikum-rezisztencia elleni innovációt

HOSSZÚTÁVÚ STRATÉGIA ADAT- ÉS INFORMÁCIÓS RENDSZEREK EGYSÉGES HASZNÁLATÁHOZ

Az állatorvosi közegészségügyi adattér stratégiai fejlesztése és az Állatorvostudományi Egyetemen fejlesztett adat- és információs rendszerek egységes használata kapcsán az alábbi szempontok figyelembevétele szükséges.

Harmonizáció

Az egyik legfontosabb kihívás, amellyel az élelmiszerlánc-biztonság területén szembe kell nézni, a különböző nemzeti és nemzetközi szabályozások és a különböző rendszerek, adatkutatási infrastruktúrák közötti harmonizáció hiánya. Bár teljes harmonizáció nem érhető el, számos területen lehetőség van a meglévő szabványosított modellek alkalmazására. Fontos, hogy az állatorvosi közegészségügyi adattónak illeszkednie kell egy közös európai/nemzetközi terminológiához. Az európai állatorvosi közegészségügyi (élelmiszerlánc-biztonsági) adatoknak jelenleg még nincs egy egységes modellje, bár erre irányuló törekvések léteznek (a jelenlegi EFSA SSD2 rendszerén alapulva). Egy ilyen egységes adatmodell jelentős előnyökkel járna, különösen a jogszabályok összehangolása és az adatgyűjtés terén. A szabványosítás fontos szerepet játszhat az adatelemzés hatékonyságának javításában, valamint az innováció elősegítésében. Az interoperábilis terminológia használata, pl. az EFSA által használt ellenőrzött kifejezéskatalógusok, lehetővé teszik az adatok összekapcsolását különböző területek között.

Az állatorvosi közegészségügyi adattónak illeszkednie kell egy közös európai/nemzetközi terminológiához

A különböző adatmodellek közötti kapcsolatokat szorosabbá kell tenni, hogy az adatelemzést és az élelmiszerlánc-biztonsági ellenőrzést hatékonyabbá lehessen tenni, ideiglenes megoldásként az egységes adatmodell kialakításáig. A harmonizált adatkezelés, beleértve a közös adatmodellek és terminológiai rendszerek fejlesztését, kulcsszerepet játszik az interoperabilitás előmozdításában az egységes adatmodell kialakításának szolgálatában.

Szabványosítás

Az adatok és adatmodellek szabványosítása nem öncélú, hanem hozzáadott értéket kell teremtenie az elemzések révén. Az adatok új felhasználási lehetőségeinek feltérképezése, különösen az AI és az élelmiszerlánc-nyomonkövetésben is jelentős szerepet játszó blokklánc-technológia alkalmazásával, szintén előnyös lehet [29]. A szabványosítás nemcsak hazai vagy EU-s szinten fontos, hanem nemzetközi szinten is. A nemzetközi intézményekkel, pl. a WHO-val és a FAO-val való együttműködés lehetőséget ad az adatok szélesebb körű felhasználására és nemzetközi partnerségek építésére is. Ez különösen fontos az élelmiszerfogyasztási és élelmiszer-összetételi adatok terén, amelyek kulcsfontosságúak a kockázatértékelésben.

Interoperabilitás

Az interoperabilitás a különböző rendszerek közötti adatcsere és a csere során kapott adatok felhasználhatóságának képességét jelenti. Jelenleg a különböző élelmiszerlánc-biztonsági adatbázisok, rendszerek nem (mindig) kommunikálnak hatékonyan egymással, ezért már középtávon is az interoperabilitás fejlesztésére van szükség. Az interoperábilis rendszerek nem csak azt jelentik, hogy minden egyes megoldás képes kommunikálni más rendszerekkel, hanem azt is, hogy minden rendszer rendelkezik egy kommunikációs modullal, amely képes információkat cserélni.

Az EFSA ajánlásai [30] szerint is a jövő élelmiszerlánc-biztonsági adatökoszisztémájának decentralizált architektúrára kell épülnie, azaz több, egymástól független, de egymással interoperábilis adatbázis/adatmodell/terminológia/ontológia létezhet. Ez csökkentené az adatfeldolgozási szűk keresztmetszetek kockázatát, és általában elérhetőbbé teszi az adatokat. A biztonsági szempontok is kiemelt figyelmet kapnak, különösen az érzékeny adatok védelme és az átláthatóság biztosítása terén.

Ontológiák

Az állategészségügyi és élelmiszer-biztonsági ontológiák fejlesztése kulcsfontosságú a hatékony adatelemzés és adatkezelés területén

Az állategészségügyi és élelmiszer-biztonsági ontológiák fejlesztése kulcsfontosságú szerepet játszik a hatékony adatelemzés és adatkezelés területén. Az ontológiák olyan szabványosított fogalmi kereteket biztosítanak, amelyek segítenek rendszerezni az információkat, és összekapcsolni a különböző adatforrásokat [31]. Az élelmiszerlánc területén a megfelelő ontológiák nélkülözhetetlenek a különböző adatbázisok közötti interoperabilitás biztosításához, valamint az adatok helyes értelmezéséhez és felhasználásához. Az ilyen ontológiák lehetővé teszik a természetes nyelvi leírások automatikus feldolgozását és gépi olvasását, ami különösen fontos az egyre növekvő adatmennyiségek kezelésében. A jövőbeni adatelemzési folyamatok támogatása érdekében kiemelten fontos, hogy az élelmiszerlánc-biztonság területén naprakész, jól definiált ontológiákat hozzunk létre és alkalmazzunk, amelyek képesek összekapcsolni az adatok széles spektrumát, beleértve a tudományos, ipari és szabályozási területeket. Az ontológiák fejlesztése és karbantartása tehát elengedhetetlen ahhoz, hogy az adatok felhasználhatók legyenek a kockázatértékelésben, a döntéshozatalban, valamint az élelmiszerlánc biztonságának biztosításában.

A projekt során az alábbi konkrét ontológiafejlesztési igényeket (és egyben stratégiai célokat) definiáltunk:

- *Takarmány-adalékanyagok.* A jelenleg elérhető, az EU-ban engedélyezett takarmány-adalékanyagok adatbázisa bár rendelkezésre áll, ennek adattartalma nem teszi lehetővé a részletes (kémiai vegyület szintű) adatok közvetlen gépi olvasását vagy adattóba integrálását. Ezért ennek rendszerének (terminológia definiálás, standard ontológiához illesztés) kialakítása lenne szükséges, ami egyben egy hatékony kutatási infrastruktúra alapját is képezné.
- *Takarmány-alapanyagok.* A takarmány-alapanyagok az előzőnél egy jóval nagyobb, diverzebb, komplexebb és egyúttal kevésbé szabályozott területet jelentenek. Itt már maga a klasszifikáció/hierarchia kialakítása és más, hasonló osztályozási rendszerekhez való illesztése is feladat, ami meg kell, hogy előzze a terminológia szabványosítását és standard ontológiához illesztését.
- *Állatbetegségek nemzetközi osztályozása.* A humán betegségek nemzetközi osztályozási rendszeréhez (BNO) [32] hasonló klasszifikáció nem létezik az állatbetegségek területén. Ennek nemzetközi kialakításának létjogosultsága erős, és az Állatorvostudományi Egyetem ennek a folyamatnak egyik facilitátora is lehetne.

Összefoglalva, az állategészségügyi és élelmiszerlánc-biztonsági adatkezelés hatékonyságának javítása érdekében javasoljuk a harmonizált adatmodellek alkalmazását, a szabványosított adatgyűjtési és -jelentési formátumok használatát, az interoperabilitás biztosítását a különböző rendszerek között, valamint a szabványos ontológiák alkalmazását és fejlesztését. Az adatelemzés hatékonyságának növelése és az új technológiai megoldások, pl. az MI és a blokklánc-technológia alkalmazása szintén javasolt a jövő állategészségügyi és élelmiszerlánc-biztonsági rendszerének kialakításában.

KÖSZÖNETNYILVÁNÍTÁS

Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal által támogatott (2020-1.1.2-PIA-CI-KFI-2021-00246) kutatás.

IRODALOM

1. European Medicines Agency, EMA (2022) European Veterinary Big Data strategy 2022- 2027 A strategic vision towards implementation of new digital solutions in the Veterinary regulatory domain. https://www.ema.europa.eu/en/documents/other/european-veterinary-big-data-strategy-2022-2027_en.pdf

2. Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, Wool EE, Robles Aguilar G, Mestrovic T, Smith G, Han C, Hsu RL, Chalek J, Araki DT, Chung E, Raggi C, Gershberg Hayoon A, Davis Weaver N, Lindstedt PA, Smith AE et al (2024) Global burden of bacterial antimicrobial resistance 1990–2021: A systematic analysis with forecasts to 2050. Lancet 404:1199–1226. https://doi.org/10.1016/S0140-6736(24)01867-1

3. Pribis JP, García-Villada L, Zhai Y, Lewin-Epstein O, Wang AZ, Liu J, Xia J, Mei Q, Fitzgerald DM, Bos J, Austin RH, Herman C, Bates D, Hadany L, Hastings PJ, Rosenberg SM (2019) Gamblers: An Antibiotic-Induced Evolvable Cell Subpopulation Differentiated by Reactive-Oxygen-Induced General Stress Response. Mol Cell 74:785–800. https://doi.org/10.1016/j.molcel.2019.02.037

4. European Centre for Disease Prevention and Control (ECDC) | European Food Safety Authority (EFSA) | European Medicines Agency (EMA) (2024) Antimicrobial consumption and resistance in bacteria from humans and food-producing animals. EFSA J 22:e8589. https://doi.org/10.2903/j.efsa.2024.8589

5. Nzietchueng S, Kitua A, Nyatanyi T, Rwego IB (2023) Facilitating implementation of the one health approach: A definition of a one health intervention. One Health 16:100491. https://doi.org/10.1016/j.onehlt.2023.100491

6. Yang HT, Ju JH, Wong YT, Shmulevich I, Chiang JH (2017) Literature-based discovery of new candidates for drug repurposing. Brief Bioinform 18:488–497. https://doi.org/10.1093/bib/bbw030

7. Nowotka MM, Gaulton A, Mendez D, Bento AP, Hersey A, Leach A (2017) Using ChEMBL web services for building applications and data processing workflows relevant to drug discovery. Expert Opin Drug Discov 8:757–767. https://doi.org/10.1080/17460441.2017.1339032

8. https://pubchem.ncbi.nlm.nih.gov/

9. https://www.ebi.ac.uk/chembl/

10. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2011) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40(Database issue):D1100–1107. https://doi.org/10.1093/nar/gkr777

11. https://bidd.group/CMAUP/

12. https://disgenet.com/

13. Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, Garcia-Garcia J, Sanz F, Furlong LI (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 45(D1):D833–D839. https://doi.org/10.1093/nar/gkw943

14. https://drugcentral.org/

15. Ursu O, Holmes J, Knockel J, Bologa CG, Yang JJ, Mathias SL, Nelson SJ, Oprea TI (2017) DrugCentral: online drug compendium. Nucleic Acids Res 45(D1):D932–D939. https://doi.org/10.1093/nar/gkw993

16. https://www.ema.europa.eu/en/human-regulatory-overview/herbal-medicinal-products

17. https://ec.europa.eu/food/food-feed-portal/screen/feed-additives/search

18. https://ec.europa.eu/food/food-feed-portal/screen/food-additives/search

19. https://foodb.ca/

20. https://www.ema.europa.eu/en/veterinary-regulatory-overview/veterinary-medicinal-products-regulation

21. http://herb.ac.cn/

22. https://www.ncbi.nlm.nih.gov/taxonomy

23. https://bidd.group/NPASS/

24.https://ncbiinsights.ncbi.nlm.nih.gov/2021/01/25/microbigge/

25.https://card.mcmaster.ca/

26. https://www.nlm.nih.gov/research/umls/index.html

27. https://www.ebi.ac.uk/unichem/

28. EFSA (European Food Safety Authority), Abbinante F, Basic S, Foster D, Fuchs K, Georgieva N, Jozwiak Á, Nellemann Ch, Maldonado A, De Martino M, O'Dea E, Scharfenberg E, Stack M, Tuominnen M, Volatier J-L, Wienk K (2024) 2023 Annual Report of the Advisory Group on Data. EFSA supporting publication 2024 21:EN-8760. pp 13. https://doi.org/10.2903/sp.efsa.2024.EN-8760

29. Akinsulie OC, Idris I, Aliyu VA, Shahzad S, Banwo OG, Ogunleye SC, Olorunshola M, Okedoyin DO, Ugwu C, Oladapo IP, Gbadegoye JO, Akande QA, Babawale P, Rostami S, Soetan KO (2024) The potential application of artificial intelligence in veterinary clinical practice and biomedical research. Front Vet Sci 11:1347550. https://doi.org/10.3389/fvets.2024.1347550

30. Alvarez-Pinera J, Bager F, Bystrický M, Rasmussen SD, Foster D, Fuchs K, Gilsenan M, Grahek-Ogden D, Jozwiak Á, Moez S, Neagu M, O'Dea E, Perrella A, Richardson J, Scharfenberg E, Sokolic D, Stack M, Vermeersch K, Wienk K (2020) Report of the Advisory Forum Task Force on Data Collection and Data Modelling. EFSA Supporting Publications 2020:EN-1901 pp 63. https://doi.org/10.2903/sp.efsa.2020.EN-1901

31. Oberle D, Guarino N, Staab S (2009) What is an Ontology? In: Staab S, Studer R: Handbook on Ontologies. Springer-Verlag, Berlin Heidelberg. pp 1–17. https://doi.org/10.1007/978-3-540-92673-3_0

32. https://icd.who.int/en

Közlésre érk.: 2025. jan. 17.