

# A genetikai adatok szerepe a járványok elleni küzdelemben

Becsei Ágnes<sup>1\*</sup>, Pipek Orsolya Anna<sup>1</sup>, Pollner Péter<sup>2,3</sup>, Csabai István<sup>1</sup>

<sup>1</sup>ELTE, Komplex Rendszerek Fizikája Tanszék, Budapest, Magyarország

<sup>2</sup>MTA-ELTE, Statisztikus és Biológiai Fizika Kutatócsoport, Budapest, Magyarország

<sup>3</sup>Semmelweis Egyetem, Egészségügyi Menedzserképző Központ, Budapest, Magyarország

Beérkezett: 2021. 05. 11.; Elfogadva: 2021. 05. 25.

## Összefoglalás

A COVID-19-járvány alatt bizonyossá vált, hogy az adattudományok, az adatok gyors megosztása és a nemzetközi összefogás a hatékony járványkezelés kulcsfontosságú eszközei. A járvány előtt létrejött Újjonnan Felbukkanó fertőző betegségek Observatóriuma (Versatile Emerging infectious disease Observatory, VEO) nevű nemzetközi konzorcium célja egy olyan monitorozó rendszer kiépítése, amely a potenciálisan veszélyes kórokozókat még az előtt azonosítja, mielőtt azok tömeges megbetegedéseket okoznának, lehetőséget adva ezzel a gyors reagálásra. A járványok megelőzésére és kezelésére létrejött nemzetközi együttműködésekben, így a VEO-ban is a kórokozók, vagy a fertőzésnek kitett személyek genetikai szekvencia adatai kiemelkedő fontosságúak. Az ilyen típusú adatok kezelésével kapcsolatban az Európai Unióban többek között a Nagojai Jegyzőkönyv és a GDPR fogalmaz meg elveket, szabályokat.

**Kulcsszavak:** VEO, genetikai szekvencia adatok, Nagojai Jegyzőkönyv, GDPR

## Genetic data in the fight against pandemics

Ágnes Becsei<sup>1</sup>, Orsolya Anna Pipek<sup>1</sup>, Péter Pollner<sup>2,3</sup>, István Csabai<sup>1</sup>

<sup>1</sup>ELTE, Department of Physics of Complex Systems, Budapest, Hungary

<sup>2</sup>MTA-ELTE, Statistical and Biological Physics Research Group, Budapest, Hungary

<sup>3</sup>Semmelweis University, Digital Health and Data Utilisation Team, Health Services Management Training Centre, Budapest, Hungary

## Summary

Data science is proved to be a key tool in the fight against the ongoing COVID-19 pandemic, but it requires a huge amount of data shared between international research groups. The Versatile Emerging infectious disease Observatory (VEO) EU collaboration was established to generate and distribute high quality data for an evidence-based early warning system for emerging infectious diseases. Through an iterative process between data scientists, disease experts, social scientists and citizen scientists, a collaborative platform will be created for storing, secure sharing and analyses of traditional and new data sources. Next generation sequencing (NGS) has revolutionized genomic research. This versatile technology is broadly applicable to pathogens and human hosts. Rapid sharing of pathogen genetic resources, including physical samples of cultured pathogens and additionally genetic sequencing data of pathogens, is crucial in support of research and outbreak response. Access to genetic resources is regulated by the Nagoya protocol which is an internationally binding treaty to ensure equal sharing of benefits arising from the use of genetic resources. So far the Nagoya protocol has been applied only to biological samples, but digital data from genetic sequencing doesn't necessarily fall under the treaty. Effects of diseases can differ based on genetic backgrounds, as certain gene variants may provide protection against or susceptibility to viral diseases. Human genomic data is an important resource for medical research. The General Data Protection Regulation (GDPR) lists identifiable human genetic data as sensitive, which is a subset of personal data. Sharing and analysis of this kind of data are strictly regulated and they are also subject to ethical challenges. These concerns become less pronounced when analyzing environmental samples like sewage. Samples collected from wastewater treatment plants can be used as pooled samples, containing naturally anonymized genetic information of the human population, near the wastewater treatment plant.

**Keywords:** VEO, genetic sequencing data, Nagoya protocol, GDPR

## Bevezetés

Napjaink egyik legnagyobb kihívása a COVID-19-járvány, melyhez hasonlót az emberiség a spanyol nátha 1918-as kitörése óta nem tapasztalt (Ashton 2020). Az azóta eltelt több mint száz év alatt a tudomány jelentős fejlődésen ment keresztül. Az utóbbi évtizedekben a világról gyűjtött adatok robbanásszerű növekedése egy új terület, az adattudomány születését idézte elő. Az adattudomány egy olyan interdiszciplináris tudományág, mely tudományos módszerek, algoritmusok és rendszerek segítségével nyer ki információt strukturált és strukturálatlan adathalmazokból (Navarro et al. 2019). Az adatforradalom nem kerülte el a biológiát sem, így körülbelül egy évtized alatt az adattudományok gyökeresen átforgalmazták a genetikát is. A korábban egy-két gén hatásainak tanulmányozása helyett az ún. új generációs szekvenálási módszereknek köszönhetően a kutatók nyakába hatalmas mennyiségű genetikai információt tartalmazó adat szakadt, mellyel az egyes gének vizsgálata helyett már az élőlények vagy vírusok teljes „tervjaja” napok alatt vizsgálhatóvá vált (Aly-Bolbol 2015). Az adattudományok elvitathatatlan előnyei mellett azonban külön figyelmet kell szentelni a biztonságra. Adatok kapcsán számos kutató felhívja a figyelmet a személyes adatokkal a való visszaélésre vagy az adatokból származó hasznok egyenlőtlen elosztására.

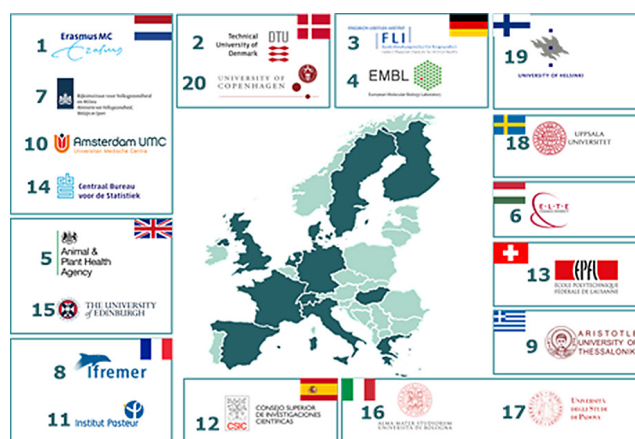
## COVID-19 járványadatok

A klasszikus járványügyi statisztikai adatok, mint az igazolt fertőzöttek száma, kórházban kezelt vagy intenzív kezelésre szoruló betegek száma, igazoltan COVID-19 megbetegedésekhez köthető halálozások száma, pozitív PCR-tesztek aránya stb. (Ritchie et al.) mellett az emberek közösségi médiában történő járvánnyal kapcsolatos megnyilvánulásai (Panuganti et al. 2020; Abd-Alrazaq 2020), mobilitási adatai vagy a szennyvízben található SARS-CoV-2-vírus koncentrációja (Kitajima 2020) is értékes adatnak számítanak a járvány kapcsán. Utóbbiak különösen hasznosak lehetnek a tünetmentes, vagy csak enyhe tünetekkel küzdő fertőzöttek nagy számával együtt járó ún. csendes járványterjedés követésében. Egyes kutatók a közösségi médiában megjelenő COVID-19 tünetekkel kapcsolatos megnyilvánulások segítségével próbálják megbecsülni a betegséget enyhe tünetekkel átvészelő emberek számát (Guo 2020; Durazzi 2020). Mások a szennyvízben található SARS-CoV-2 örökítőanyagának koncentrációja alapján igyekeznek a fertőzöttek számának növekedését előre jelezni (Róka et al. 2020). A SARS-CoV-2 elsősorban egy cseppfertőzéssel terjedő, légúti megbetegedést okozó vírus, de az emésztőrendszer sejtjeiben is jelen lehet. Ennek eredményeként a fertőzött, akár tünetmentes személyek jelentős része széklettel is üríti a vírusrészecskéket, amelyek így a szennyvízbe kerülhetnek (Izquierdo-Lara 2021). A SARS-CoV-2 örökítőanyagának növekedése 4–10

nappal megelőzheti a tömeges megbetegedéseket. Bár egyelőre a legtöbb ilyen vizsgálat kutatási célú, Európában már több ország, ideértve Magyarországot is, heti rendszerességgel méri a szennyvízminták koronavírus-koncentrációját (Róka et al. 2020).

## Újonnan Felbukkanó fertőző betegségek Obszervatóriuma

Talán minden EU-s tagállam egyetért abban, hogy a COVID-19 elleni küzdelem fontos közérdek, amely során a nemzetközi kutatószervezetekhez történő adat-továbbítás elengedhetetlen (03/2020. iránymutatások). Ilyen nemzetközi együttműködés a VEO (Versatile Emerging infectious disease Observatory), mely húsz kutatócsoportot, köztük az ELTE Természettudományi Karán működő kutatócsoportunkat is magába foglaló nemzetközi konzorcium.



1. ábra

Az Újonnan Felbukkanó fertőző betegségek Obszervatóriuma (VEO) tagországai és intézetei. A konzorcium 12 európai ország 20 intézetét foglalja magába. Nyugat-, észak-, dél- és kelet-európai partnereivel a VEO átfogó képet kaphat a különböző európai ökoszisztémákról, egészségügyi rendszerekről és nemzeti szabályozási környezetekről

(Forrás: <https://www.veo-europe.eu/about-veo/partners>)

A VEO célja egy olyan virtuális megfigyelőrendszer létrehozása, amely képes előre jelezni és nyomon követni az újonnan felbukkanó fertőző betegségeket. A projektben adattudósok, virológusok, járványügyi és más szakértők dolgoznak közösen a szükséges adatok összegyűjtésén, az eszköztár kidolgozásán, kollaboratív adatelemző platformok felállításán és az új betegségek kialakulásának és fő útvonalainak elemzésén, Európában és világszerte (VEO). A járványok kialakulása rendkívül összetett, soktényezős folyamat. Például, a klímaváltozással a betegségeket terjesztő vektorok (szúnyogok, kullancsok) elterjedési területe megváltozik (Semenza–Suk 2017), a természetes élőhelyek csökkenésével a vadállatok (denevérek, madarak) (Reed 2003) könnyebben érintkeznek az emberekkel vagy haszonállatokkal. Valamint a sarki régiókban az esetlegesen veszélyes kórokozókat tartal-

mazó permafroszt gyors kiolvadása (Sajjad 2020) újabb, talán a mostaninál is jelentősebb járványok kialakulásához vezethet. A megfelelő védekezéshez az újonnan felbukkanó betegségeket (beleértve a gyógyszerekre rezisztens kórokozókat) a jelenleginél sokkal korábbi szakaszban fel kell fedeznünk, így időt adva a gyors reakcióra, hogy a járványok és betegségek hatását csökkentsük. A VEO célkitűzésében szereplő előrejelző és monitorozó rendszer működésének alapja a nagy mennyiségű releváns és sokrétű adat komplex elemzése. Ezek lehetnek klasszikus járványügyi vagy újszerű információk, amiket a szennyvízből és más környezeti mintákból vagy páciensekből nyert genetikai információ szolgáltat. De a zoonózisok (olyan fertőző betegségek, amelyek állatról emberre képesek terjedni) kapcsán fontosak lehetnek más biológiai adatok is, mint a vándormadarak vonulási útvonalaira, trópusi szúnyogfajok vagy bizonyos vadállatok (pl. denevérek) elterjedési területeire vonatkozó adatok. További értékes információkkal szolgálhatnak egy járványról a meteorológiai mérések is (VEO). A VEO együttműködésben feldolgozott adatok döntő többsége azonban az új generációs szekvenálási (NGS) módszerek valamelyikével kinyerhető genetikai adat. A DNS szekvenálási technológiák fejlesztésében elért sikereknek és a költségek drámai csökkenésének köszönhetően ezen technológiák ugrásszerű elterjedése figyelhető meg az állategészségügyi és élelmiszer-biztonsági intézetekben, a klinikai mikrobiológiai laboratóriumokban és a fertőző betegségek kiterjedésének megelőzésére létrejött konzorciumokban. A szekvenálás során keletkező hatalmas mennyiségű adat fejlett infrastruktúrát igényel mind az adatok tárolása, megosztása és elemzése tekintetében. A technikai feltételek biztosítása mellett azonban külön akadályt jelenthetnek a különböző politikai, etikai, adminisztratív és jogi tényezők, így a platformok kialakítása során ezekre is külön figyelmet kell fordítani. A VEO-t megelőző COMPARE együttműködés keretein belül elindított KOOPLEX (Visontai et al. 2019) informatikai adatelemzést segítő platform az újgenerációs szekvenálási adatok feltöltésére, elemzésére és megosztására egy integrált, felhasználóbarát lehetőséget biztosít. A „COMPARE data hub” egy zárt platform, ahol az adatok megosztása csak kollaborációs partnerek között lehetséges, biztosítva ezzel például a publikálás előtt álló adatok biztonságos megosztását (Amid et al. 2019).

## Patogén szekvenciák megosztása

Egy vírusfertőzés során a gazdaszervezetbe jutó vírus arra kényszeríti a gazdaszervezet sejtjeit, hogy róla másolatokat készítsenek, kezdve a vírus genomjával. A másolás nem mindig tökéletes, ezért a vírus genomjának másolata már nem mindig lesz teljesen azonos az eredeti genommal, kisebb hibák, mutációk lehetnek benne, amik tovább öröklődnek a másolatok másolataiba, és potenciálisan újabbakkal egészülnek ki. Hogy a vírus genomban hol és milyen gyakorisággal jelennek meg

ezek a mutációk, az számos tényezőtől függ. Az RNS-vírusok, mint például a SARS-CoV-2 genomja gyorsabban mutálódik, mint a DNS-vírusoké (Gough, 2020). A mutáció eredményeként újabb és újabb SARS-CoV-2 változatok jönnek létre, melyeket genomszekvenciájuk ismeretében lehet hatékonyan azonosítani. Egyes típusok, attól függően, hogy az ún. „signature” mutációikban mennyire különböznek az eredeti változattól, külön elnevezést kaphatnak, mint például B.1.1.7 és 501.V2, vagy ismertebb nevükön brit és dél-afrikai variáns (Jahn et al. 2021). A variánsok hatékony követéséhez a SARS-CoV-2 szekvencia adatok nemzetközi megosztása elengedhetetlen. Ezt a célt szolgálja az egyik legnagyobb SARS-CoV-2- és influenzavírus szekvenciákat tartalmazó adatbázis, a GISAID, valamint a nyers szekvenálási adatokat is tartalmazó európai COVID-19 adatplatform (<https://www.covid19dataportal.org/>). A teljes genom szekvenálási módszerek fejlődése új lehetőségeket kínál a kórokozók monitorozására. Egyre több és több patogén szekvencia érhető el világszerte, így a patogén genom-szekvencia adatbázisok részben átveszik a mikrobiális és víruskultúrák tárolására létrehozott biobankok szerepét (S. Ribeiro 2018). A Biológiai Sokféleség Egyezmény nyilatkozata alapján mind a biobanknak, mind az adatbázisoknak tiszteletben kell tartaniuk a minta, illetve az adatok birtoklásával kapcsolatos jogokat, ideértve az államok a saját területükön gyűjtött minták és adatok felhasználásáról való rendelkezési jogát (<https://www.cbd.int/convention/text/>). 2010-ben ezen egyezmény keretein belül jött létre, majd lépett hatályba 2014-ben a genetikai erőforrásokról szóló Nagojai Jegyzőkönyv. Egyelőre azonban vitatott, hogy a genetikai szekvencia adatok mennyire tartoznak bele az elsősorban fizikai mintákra vonatkozó jegyzőkönyv hatáskörébe (S. Ribeiro 2018). A mostaninhoz hasonló népegészségügyi vészhelyzetekben az adatmegosztás, köztük a genetikai szekvencia adatok megosztása elősegíti a vészhelyzet okának mélyebb megértését és a bizonyítékalapú döntéshozást, ezért a releváns adatok szabad és gyors megosztása kulcsfontosságú. Amennyiben azonban a genetikai szekvencia adatok a Nagojai Jegyzőkönyv hatálya alá esnek, akkor az ilyen adatok megosztása illegálisnak minősül, és a biobankok és az adatbázisok esetében ismert komplex adminisztratív teherrel kell megküzdenie mind a fogadó, mind pedig a megosztó félnek.

## Humán szekvenciák megosztása

Adatvédelmi szempontból kiemelt jelentőséggel bírnak a személyes adatok és az ezen belül különleges adatkategóriába tartozó genetikai, biometrikus vagy egészségügyi adatok. Ez utóbbiak nem kezelhetők, kivéve például, ha az érintett személy kimondottan hozzájárul, ha az adatkezelést alapvető közérdek indokolja, illetve ha uniós vagy tagállami jog szükségessé teszi. A személyes adatok védelméről a részletes követelményeket a 2016-ban elfogadott, majd 2018-ban hatályba lépett GDPR

(General Data Protection Regulation), más néven általános adatvédelmi rendelet rögzíti (03/2020. iránymutatások). Eszerint személyes adatnak minősül minden olyan adat, amely azonosított vagy azonosítható élő személlyel kapcsolatos. Amennyiben a személyes adatok úgy anonimizáltak, hogy az érintett személy semmilyen módon nem azonosítható, akkor nem tekinthetők többé személyes adatnak.

A humán genomi adatokból nyert értékes információk egyre inkább mindennapos eszközévé válnak a betegségek diagnosztikájának és a kutatásnak. Az ilyen adatok szigorúan védett különleges adatoknak minősülnek. A GDPR szerint genetikai az „egy természetes személy örökölt vagy szerzett genetikai jellemzőire vonatkozó minden olyan személyes adat, amely az adott személy fiziológiájára vagy egészségi állapotára vonatkozó egyedi információt hordoz, és amely elsősorban az említett természetes személyből vett biológiai minta elemzéséből ered” (Európai Bizottság). Az ilyen jellegű adatok kezelése számos etikai kérdést is felvet. Ilyen például az ún. terápiás szakadék helyzete (Ábránd–Sándor–Brand 2012). Ez azt jelenti, hogy a genom ismeretében egyre nagyobb biztonsággal lesznek jósolhatók a jövőbeli betegségek, úgy, hogy sokszor nincs még ismert terápia az adott betegségre, mint például a súlyos szellemi és fizikai leépüléssel járó genetikailag örökölt betegség, a Huntington-kór (McColgan–Tabrizi 2018). A Huntington-kórt egy dominánsan öröklődő gén hibája okozza, mely akár már születéskor kimutatható, viszont hatása csak jóval későbbi életkorban jelentkezik (McColgan–Tabrizi 2018). Egyes betegségek esetében azonban nem ilyen egyértelmű a genetikai kapcsolat, vagy egyszerre több, részben ismeretlen gén mutációja felelős kialakulásukért (pl. tumoros megbetegedések, Lafin et al. 2019). Ebben az esetben az eredmények bizonytalansága nehezíti a megfelelő terápia kiválasztását.

Kockázatot jelenthet a humán genomi adatok egészségügyön kívüli potenciális felhasználása, melyek így visszaélések eszközeivé válhatnak. Ezen adatok birtokában például biztosítási társaságok sorolhatnak egyeseket magasabb kockázati csoportba, vagy leendő munkáltatók diszkriminálhatnak a jelöltek genomi adatai alapján feltételezett képességei szerint (Ábránd–Sándor–Brand 2012). Talán ezek okán a humán genetikai adatokkal való visszaéléstől való félelem a genomi adatokkal kapcsolatos kutatások egyik legnagyobb akadálya (Ábránd–Sándor–Brand 2012). Személyes adat (így a különleges adat) definíciója szerint az egyik kulcskérdés az azonosíthatóság. A szennyvízben található humán eredetű DNS mondhatni természetes módon anonimizált, továbbá sokrétű, hiszen akár egy nagyobb emberi populáció vizsgálatát is lehetővé teszi. Egy ilyen minta sok esetben akár több millió ember DNS-ét ömlesztve tartalmazza, általában jóval kisebb mennyiségben, mint más élőlények DNS-ét. Konkrét személyek nem azonosíthatók vissza.

Populációsintű elemzések, statisztikai összegző elemzések és döntéstámogató kimutatások azonban kinyer-

hetők az anonim szennyvízadatokból. Kutatócsoportunk felvetette, hogy a szennyvízmintákban található kis mennyiségű, de minden emberi sejt mitokondriumában megtalálható DNS-szekvenciák elemzésével értékes információk nyerhetők a szennyvíztelep környékén élő népességről. A mitokondriumok a sejtek energiaellátásáért felelős sejtszervecskék, amelyek különlegessége, hogy saját genetikai állománnyal rendelkeznek és ún. anyai öröklődésűek, azaz kizárólag a petesejteken keresztül öröklődnek. A mitokondriális DNS így csak a nagy ritkán bekövetkező mutációk révén változik. A mutációval keletkező mitokondriális DNS-változatok, a haplotípusok, a mitokondriumokkal öröklődnek anyáról a gyermekekre. A haplotípusokat hasonlóságuk alapján haplocsoportokba sorolhatjuk (Pipek 2019). Egy-egy haplocsoport adott földrajzi területen gyakoribb lehet, így megkülönböztethetünk például nyugat-ázsiai vagy európai haplocsoportokat. Ezzel a módszerrel a szennyvízgyűjtő területen élő populáció etnikai és genetikai összetételét, illetve egyes betegségekkel összefüggésbe hozott haplotípusok jelenlétét, azaz a potenciálisan veszélyeztetett személyek jelenlétét anonim módon vizsgálhatjuk.

## Összefoglalás

A 21. század talán legnagyobb egészségügyi kihívását az újonnan megjelenő fertőző betegségek jelentik. A COVID-19 világszintű elterjedésével sajnos a kérdés igen aktuális lett, és a jövőben is számíthatunk hasonló vagy ennél súlyosabb járványokra. A járványok hatékony megelőzése vagy kezelése nemzetközi összefogást igényel, amely együtt jár az adatok országok közötti megosztásával. A rohamosan fejlődő új generációs szekvenálási módszerek kiváló új lehetőségeket rejtenek az élővilág megismerésére. A genetikai anyag, mint egy molekuláris „tervrajz”, minden élő sejtben és vírusban ott van. Megismerésével a humán betegségekről és kórokozókról mindaddig példátlan mennyiségű információ nyerhető ki. Az így nyert információk társadalmi hasznossága elvitathatatlan, de az adatokkal való visszaélés elkerülése érdekében nagy hangsúlyt kell fektetni a személyes és különleges adatok védelmére, valamint az adatokból származó előnyök méltányos elosztására. A patogénnel kapcsolatos adatok megosztása részben a Nagojai Jegyzőkönyvben foglaltak szerint szabályozott, de a genetikai szekvencia adatok esetében nem egyértelmű, hogy jegyzőkönyv hatálya alá tartoznak, vagy sem. A személyes adatok és az ezen belül egy szigorúbb alkategóriát képviselő különleges adatok védelmét az Európai Unió belül a GDPR szabályozza. Eszerint a személyes adatok egyik legfontosabb tulajdonsága az azonosíthatóság, vagyis a lehetőség arra, hogy egy adatot egy konkrét személyhez kössünk. Egy konkrét személyből származó humán genom szekvenálása esetén a kapcsolat egyértelmű. A kérdés azonban nem mindig ennyire egyszerű. Egyes esetekben a minta jellege miatt a humán szekvenciák természetes módon anonimizáltak a

mintában, mint például a szennyvízminták esetében. Bár ezek a lehetőségek rövid távon megkönnyítik az adatok biztonságos kezelését, hosszú távon a kutatók és adatvédelmi szakértők közös munkája szükséges, hogy az értékes kutatási eredmények mellett az adatok védelme is biztosított legyen.

## Köszönetnyilvánítás

Ez a cikk az EU Horizon 2020 program „VEO” No.874735 támogatási megállapodás és a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal No. FIEK\_16-1-2016-0005 támogatásával jött létre.

A kutatást részben az Innovációs és Technológiai Minisztérium Tématerületi Kiválósági Programja (2020-4.1.1.-TKP2020) finanszírozta, a Semmelweis Egyetem Digitális Biomarker tématerületi programja keretében.

## Irodalomjegyzék

- 03/2020. iránymutatások az egészségügyi adatoknak a Covid-19 járvánnyal összefüggésben végzett tudományos kutatás céljából történő kezeléséhez. [https://edpb.europa.eu/sites/edpb/files/files/file1/edpb\\_guidelines\\_202003\\_healthdatascientificresearch-covid19\\_hu\\_0.pdf](https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202003_healthdatascientificresearch-covid19_hu_0.pdf)
- Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020) Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. *J Med Internet Res.*, Vol. 22. No. 4. pp. 491–197. DOI: 10.2196/19016
- Ábránd R., Sándor J., & Brand A. (2012) Népegészségügyi genomika. Budapest, Medicina
- Aly, S. & Bolbol, D. (2015) Next generation sequencing (NGS): A golden tool in forensic toolkit. *Archives of Forensic Medicine and Criminology*, Vol. 65. No. 4. pp. 260–271. DOI: 10.5114/amsik.2015.61029
- Amid, C., Pakseresht, N., Silvester, N., Jayathilaka, S., Lund, O., Dynowski, L. D., et al. (2019) The COMPARE Data Hubs. *Database*, Vol. 2019. DOI: 10.1093/database/baz136
- Ashton J. (2020) COVID-19 and the ‘spanish’ flu. *Journal of the Royal Society of Medicine*, Vol. 113. No. 5. pp. 197–198. DOI: 10.1177/0141076820924241
- COMPARE: Collaborative management platform for detection and analyses of (re-)emerging and foodborne outbreaks in europe. <https://www.compare-europe.eu/about>
- Durazzi, F., Müller, M., Salathé, M., & Remondini, D. (2020) International expert communities on twitter become more isolated during the COVID-19 pandemic. *arXiv preprint arXiv:201106845*
- Európai Bizottság (é. n.) Mit nevezünk személyes adatnak? [https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data\\_hu#pldk-nem-szemlyes-adatnak-minsl-adatra](https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_hu#pldk-nem-szemlyes-adatnak-minsl-adatra)
- GISAID: The GISAID initiative for rapid sharing of data from all influenza viruses and the coronavirus causing COVID-19. <https://www.gisaid.org/>
- Gough, N. R. (2020) Defining a new strain of a virus. *BioSerendipity*, <https://www.bioserendipity.com/defining-a-new-strain-of-a-virus/>
- Guo, J., Radloff, C. L., Wawrzynski, S. E., & Cloyes, K. G. (2020) Mining twitter to explore the emergence of COVID-19 symptoms. *Public Health Nursing*, Vol. 37. No. 6. pp. 934–940.

- Izquierdo-Lara, R. W., Elsinga, G., Heijnen, L., Munmink, B. B. O., Schapendonk, C. M. E., Nieuwenhuijse, D., et al. (2021) Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing. *Emerging Infectious Diseases*, Vol. 27. [https://wwwnc.cdc.gov/eid/article/27/5/20-4410\\_article](https://wwwnc.cdc.gov/eid/article/27/5/20-4410_article)
- Jahn, K., Dreifuss, D., Topolsky, I., Kull, A., Ganesanandamoorthy, P., Fernandez-Cassi, X., et al. (2021) Detection of SARS-CoV-2 variants in switzerland by genomic analysis of wastewater samples. *medRxiv*. DOI: 10.1101/2021.01.08.21249379
- Kitajima, M., Ahmed, W., Bibby, K., Carducci, A., Gerba, C. P., Hamilton, K. A., Haramoto, E., & Rose, J. B. (2020) SARS-CoV-2 in wastewater: State of the knowledge and research needs. *Science of The Total Environment*, Vol. 739. DOI: 10.1016/j.scitotenv.2020.139076
- Lafin, J. T., Bagrodia, A., Woldu, S., Amatruda, J. (2019) New insights into germ cell tumor genomics. *Andrology*, Vol. 7. No. 4. pp. 507–515. DOI: 10.1111/andr.12616
- McColgan, P. & Tabrizi, S. J. (2018) Huntington’s disease: A clinical review. *European Journal of Neurology*, Vol. 25. No. 1. pp. 24–34. DOI: 10.1111/ene.13413
- Nagójai Jegyzőkönyv. A genetikai erőforrásokhoz való hozzáférés és a hasznosításukból származó előnyök megosztása. [https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=LEGISSUM:20110303\\_1](https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=LEGISSUM:20110303_1)
- Nations U. UN Convention on Biological Diversity. [www.cbd.int/convention/text/default.shtml](http://www.cbd.int/convention/text/default.shtml)
- Navarro, F., Mohsen, H., Yan, C., Li, S., Gu, M., Meyerson, W. et al. (2019) Genomics and data science: An application within an umbrella. *Genome Biology*, 20. DOI: 10.1186/s13059-019-1724-1
- Panuganti, B. A., Jafari, A., MacDonald, B. & DeConde, A. S. (2020) Predicting COVID-19 incidence using anosmia and other COVID-19 symptomatology: Preliminary analysis using google and twitter. *Otolaryngology–Head and Neck Surgery*, Vol. 163. No. 3. pp. 491–197. DOI: 10.1177/0194599820932128
- Pipek O. A., Medgyes-Horváth A., Dobos L., Stéger J., Szalai-Gindl J., Kaas D. V. R. S., et al. (2019) Worldwide human mitochondrial haplogroup distribution from urban sewage. *Scientific Reports*, Vol. 9. DOI: 10.1038/s41598-019-48093-5
- Reed, K. D., Meece, J. K., Henkel, J. S., Shukla, S. K. (2003) Birds, migration and emerging zoonoses: West Nile virus, Lyme disease, influenza A and enteropathogens. *Clinical Medicine & Research*, Vol. 1. No. 1. pp. 5–12. DOI: 10.3121/cmr.1.1.5
- Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald B. et al. Coronavirus pandemic (COVID-19). <https://ourworldindata.org/coronavirus>
- Róka E., Khayer B., Schuler E., Kovács B. L., Magyar N., Kiss Z., et al. (2020) Koronavírus a szennyvízben: Országos korai előrejelző rendszer kialakítása. *Hírcsatorna*, 2020: 17–25.
- S. Ribeiro, C. dos, Koopmans, M. P., & Haringhuizen, G. B. (2018) Threats to timely sharing of pathogen sequence data. *Science*, Vol. 362. No. 6413. pp. 404–406. DOI: 10.1126/science.aau5229
- Sajjad, W., Rafiq, M., Din, G., Hasan, F., Iqbal, A., Zada, S., et al. (2020) Resurrection of inactive microbes and resistome present in the natural frozen world: Reality or myth? *Science of The Total Environment*, Vol. 735. DOI: 10.1016/j.scitotenv.2020.139275
- Semenza, J. C. & Suk, J. E. (2017) Vector-borne diseases and climate change: a European perspective. *FEMS Microbiology Letters*, Vol. 365. No. 2. DOI: 10.1093/femsle/fnx244
- The european COVID-19 data platform. <https://www.covid19data-portal.org/>
- Versatile emerging infectious disease observatory (VEO). <https://www.veo-europe.eu/>
- Visontai D., Stéger J., Szalai-Gindl J. M., Dobos L., Oroszlány L., & Csabai I. E. (2019) Kooplex: Collaborative data analytics portal for advancing sciences. *arXiv preprint arXiv:191109553*.

A cikk a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>) feltételei szerint publikált Open Access közlemény, melynek szellemében a cikk bármilyen médiumban szabadon felhasználható, megosztható és újraközölhető, feltéve, hogy az eredeti szerző és a közlés helye, illetve a CC License linkje és az esetlegesen végrehajtott módosítások feltüntetésre kerülnek. (SID\_1)