

Kutatási adatkezelést támogató szolgáltatások előkészítése a Szegedi Tudományegyetemen

Zeller Rozália – Hoczopán Szabolcs – Nagy Gyula

A Szegedi Tudományegyetem számára biztosított kutatási adatkezelési (RDM) szolgáltatásainkat megvalósítandó, 2020 őszén a mindennapi adatkezelési praxist vizsgáló kérdőívet készítettünk, hogy az egyetem munkatársai körében felmérjük a kutatási adatkezeléssel kapcsolatos általános hozzáállást, továbbá valós képet kapjunk a felmerülő elvárásokról, igényekről. Különösen kíváncsiak voltunk véleményükre a vitatottabb RDM-témákról. A kérdőív összeállításához több hasonló, korábban zajlott felmérés kérdéseit, illetve eredményeit vizsgáltuk meg, illetve igazítottuk a Szegedi Tudományegyetem speciális adottságaihoz, gyakorlatához. Jelen tanulmányban a kérdőív kiértékelése kapcsán levont tanulságokat, tapasztalatokat közöljük. Az elemzés alapján elmondható, hogy a kérdőív hasznos eszközül szolgált, hogy reális képet kapjunk az SZTE sokszínű kutatói állományának aktuális, RDM-hez való hozzáállásáról.

Tárgyszavak: nyitott tudomány kutatási módszer, adatkezelés, repositórium, egyetemi könyvtár, felmérés

1. Bevezető

Hazánkban is egyre több kutatóhely és kutatástámogató szervezet ismeri fel a kutatási adatok kezelésének (Research Data Management = RDM) fontosságát; az elmúlt években igazán forró témává vált ez a részterület az egyébként is igen gyorsan változó Open Science mozgalom kurrens témái között. Az Open Science alapelvek térhódítása mellett legalább ennyire fontos szerep jut az akadémiai szféra ráeszmélésének, miszerint nemcsak egy tudományos folyóiratban megjelenő tanulmány jelentheti az elszámolható tudományos produktumot, hanem az odavezető út során keletkező adatok szisztematikus dokumentálása, tárolása, megosztása, esetleges újrafelhasználása legalább annyira fontos az Open Science megközelítés szerint. A szereplők – a fentieket felismerve és tudomásul véve – azt is rövid úton belátják, hogy mindez tudatosan végiggondolt kutatási adatmenedzsment-stratégia és megfelelő támogató környezet nélkül nagyon nehézkes. Ezt a támogatást a világban tapasztalható trendeknek megfelelően

jellemzően az akadémiai környezetben egyébként is jelenlévő kutatástámogató szervezet, a könyvtár tudja leginkább biztosítani a kutatók számára.

Az elmúlt években a területen itthon is elinduló történések szintén ebbe az irányba mutatnak, több magyarországi, erős könyvtári kötődésű RDM-műhely is kezd kibontakozni, főként a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár¹ és az MTA Könyvtár és Információs Központ² jóvoltából. Az ilyen módon kialakuló könyvtári „élmezőnyhöz” az ország egyik vezető tudományegyetemének könyvtáraként az SZTE Klebelsberg Könyvtár is szeretne felzárkózni. Mindez a törekvés tudatos,

1 Görögh Edit, Száldobágyi Ádám. (2020). Kutatási adatkezelés egyetemi környezetben. Workshop 2020, Debrecen.

<https://kifu.videotarium.hu/hu/recordings/43365/kutatasi-adatkezeles-egyetemi-kornyezetben>

2 Holl András. (2018). A nyílt kutatási adatok kezelésének három oldala In: Tick, József; Kokas, Károly; Holl, András (szerk.) NET-WORKSHOP 2018. Budapest, Magyarország : Hungarnet (2018) pp. 59-63.

<https://doi.org/10.31915/nws.2018.8>

ezért egy nemrégiben zárult EFOP 3.6.1. projektünk keretében új kutatási adatmenedzsment-szolgáltatások elindítását és az RDM alapelveinek térnyerését mindinkább támogató környezet megteremtését tűztük ki célul a Szegedi Tudományegyetem polgárai számára. Mindezen törekvéseinkről beszámoltunk a 2021-es Networkshop előadásunkban³, amely a NWS konferenciakötetben tanulmányként is megjelenik. Jelen írás ezt az előadást hivatott kiegészíteni, a kutatási adatmenedzsment támogató szolgáltatások elindítását megalapozó előkészítő lépések egyik legfőbb momentumának tekinthető kérdőíves felmérésünk eredményeinek részletes bemutatásával. Fontosnak gondoltuk, hogy jó példával járjunk elöl, ezzel is növelve mondanivalónk hitelességét, ezért a tanulmány alapjául szolgáló felmérés kutatási adatait szabadon hozzáférhető módon magunk is elhelyeztük egy adatrepozitóriumban⁴.

2. A kérdőív megtervezése

2020 őszén egy, a mindennapi adatkezelési praxist vizsgáló kérdőívet készítettünk annak érdekében, hogy az SZTE-kutatók között felmérjük a kutatási adatkezeléssel kapcsolatos elvárásokat, igényeket. Szintén kíváncsiak voltunk véleményükre néhány különösen vitatott RDM-témáról. A kérdőív összeállításához több hasonló hazai^{5,6} és nemzetközi⁷ felmérés kérdéseit, illetve eredményeit vizsgáltuk meg, illetve igazítottuk a Szegedi Tudományegyetem speciális adottságaihoz, gyakorlatához.

A kutatási adatkezeléssel kapcsolatos kérdőív kitöltésére 2020. november 18. és 2021. január 7. között volt lehetőség, ez idő alatt többször is próbáltuk népszerűsíteni a felmérést különböző kommunikációs csatornáinkon (a könyvtár Facebook oldala, Szerzői Eszköztár, szakreferensi hálózat, egyetemi honlap, hírlevelek). Végül összesen 116 kutató töltötte ki a kérdőívet – ez a részvételi arány megfelelt előzetes várakozásainknak, sőt valamelyest meg is haladta azt.

A kérdőív öt fő kérdéscsoportból állt, ezek: 1) Személyes adatok; 2) Formátumok, adattípusok, etikai és jogi kérdések; 3) Adattárolás, archiválás és adatvesztés; 4) Az adatok hozzáférhetősége, megosztása és újrhasználtsága; 5) Adatkezeléssel kapcsolatos igények. A kérdések összeállítása során elméleti és praktikus szempontokat egyaránt igyekeztünk figyelembe venni, ezért a válaszadók kutatási adatkezeléssel és -megosztással kapcsolatos motivációi, attitűdjei mellett az ezekkel kapcsolatos napi gyakorlatokra is rákérdeztünk.

Célzott kérdéseket tettünk fel a kitöltők esetleges, az RDM egyes területein jelentkező hiányosságaira; illetve ezzel összefüggésben arra vonatkozóan, hogy a könyvtár a kutatási adatkezelés mely területein tudná leginkább támogatni a kutatóközösséget. Összességében elmondható, hogy a kérdőív elérte célját, és segítségével hatékonyabban tudtuk megtervezni a könyvtár kutatási adatkezeléssel kapcsolatos szolgáltatásait.

3. A kérdőív elemzése

3.1 Személyes adatok

Felmérésünk első részében a kitöltők tudományterületére, publikációs szokásaira voltunk kíváncsiak. A kérdőívet kitöltő szerzők szakterületi megoszlása nem tükrözi teljesen a Szegedi Tudományegyetem publikációs arányait. A válaszadók között felülreprezentáltak voltak a bölcsészettudományok és a társadalomtudományok képviselői; ez okozta az első meglepetést, mert sztereotip módon közelítve nem ezek a legtipikusabb területek, melyek kapcsán a kutatási adatkezelési kérdések felmerülését várhatná a laikus szemlélő.

A természettudományok területén működő kutatók hozzávetőlegesen az egyetemi arányoknak megfelelően, az élettudományok képviselői viszont a vártnál jóval kisebb arányban szerepeltek a kitöltő

3 Zeller Rozália, Hoczopán Szabolcs, Nagy Gyula. (2021). Kutatási adatkezelést támogató szolgáltatás a Szegedi Tudományegyetemen. Networkshop 2021.
<https://kifu.videotorium.hu/hu/recordings/42348>

4 Zeller Rozália, Hoczopán Szabolcs, Nagy Gyula. (2021). Kutatási adatkezelést támogató szolgáltatások előkészítése a Szegedi Tudományegyetemen kérdőív és válaszok [Data set]. Zenodo.
<https://doi.org/10.5281/zenodo.5166625>

5 Discovering Research Data Practices at University of Debrecen (Hungary)
<https://www.openaire.eu/blogs/discovering-research-data-practices-at-university-of-debrecen-hungary-1>

6 Szabo Zsuzsa. (2019). Discovering Research Data Practices at University of Debrecen. COAR 2017 Annual Meeting in Venice (Italy) 8-10 April 2017.
<https://doi.org/10.5281/zenodo.800523>

7 Springer Nature State of Open Data 2019
https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_Report_2019/9980783

tők között. Mindezeket figyelembe véve értékes információkkal szolgált a kérdőív, mivel olyan kutatói, szerzői kutatási adatkezelési igényekről szerezhattünk információt, melyek egyelőre kevésbé szerepelnek a köztudatban.

A kitöltők nagy része tudományos fokozattal rendelkező, aktívan publikáló szerző volt, státuszuk alapján legnagyobb részük egyetemi docens és adjunktus. Némi meglepetést okozott, hogy a PhD hallgatók kisebb arányban vettek részt a felmérésben. Különösen amiatt volt ez váratlan, mert egyébként nagyon aktívan veszik igénybe az SZTE Klebelsberg Könyvtár különböző kutatástámogatási szolgáltatásait.

Megnyugtató, egyúttal figyelemre méltó, hogy a kérdőívet kitöltők szinte egészének – tudományterülettől függetlenül – kevesebb mint egy éve jelent meg az utolsó publikációja. Azaz valóban aktív, gyakorló szerzők válaszai alapján vonhattunk le következtetéseket.

Kiemelendő motívum – annak ellenére, hogy a válaszadók nagyobbik része a bölcsészettudományok és a társadalomtudományok képviselői voltak – hogy több mint háromnegyedük felé már van tudományterületi elvárás a kutatási adatok publikálására és ennek a szerzők többsége tudatában is van. A kutatók további közel tíz százaléka elképzelhetőnek tartja, hogy számára is elvárás a kutatási adatok publikálása, de ebben nem volt teljesen biztos.

3.2 Formátumok, adattípusok, etikai és jogi kérdések

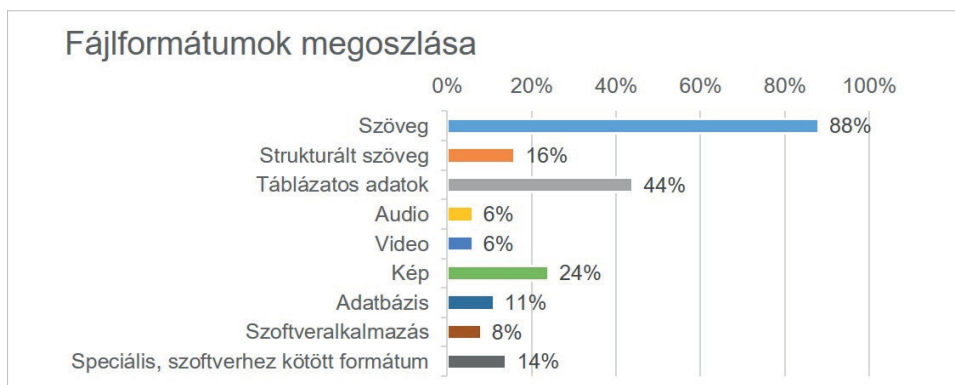
A következő kérdésblokkban a megszülető adatok típusára, formátumára és ezek kezelésére voltunk kíváncsiak. A kitöltők tudományterületi eloszlásából következően magától értetődően legtöbbször szöveges, táblázatos és képi formátumú adatokat állítanak elő. A kérdőívre adott válaszok alapján ezek a formátumok dominánsnak tekinthetők a bölcsész- és társadalomtudományokon túl minden más tudományterületen. Szoftveralkalmazások és speciális szoftverhez kötött formátumú adatok elsősorban a természettudományi területen születnek, de meglepetésünkre számos bölcsész-kutató is jelezte, hogy dolgozik ilyen jellegű adatokkal. Nagyon hasonló képet mutatnak az adatbázisokban (MS Access, MySql stb.) gyűjtött adatok, azzal

a különbséggel, hogy egyes társadalomtudományi területeken még hangsúlyosabbak ezek a megoldások. Audio és video formátumú adatfájlokat kevesen állítanak elő, elsősorban a bölcsészettudomány kutatóira jellemző ez a tevékenység (1. ábra).

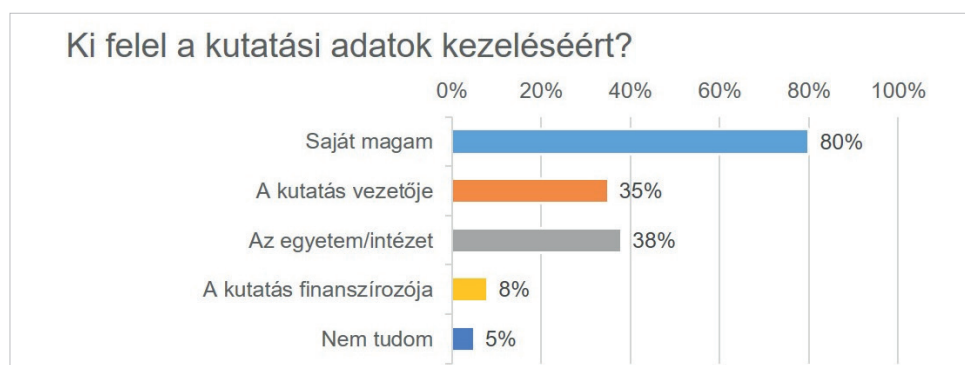
A következő kérdés az adatok szenzitivitásával foglalkozott. Nagyon kevés szegedi kötődésű adatsomag tartalmaz szabadalmakat, vagy üzleti szempontból kényes adatokat. Ez a néhány eset is leginkább a természettudomány és a mezőgazdaság területéről került ki. Személyes adatokkal viszont a kutatók nagyobb része, több mint harmaduk dolgozik, gyakorlatilag minden tudományterületen. A kutatók meggyőző aránya komoly gondot fordít a személyes adatok megfelelő anonimizálására.

A könyvtári szolgáltatások további fejlesztése miatt különösen fontosnak tekinthető a következő kérdés, miszerint a kutatók kitartanak felelősnek kutatási adatok kezeléséért. Több opció is választható volt, hiszen a kutatás különböző fázisaiban könnyen elképzelhető, hogy más és más lesz az adatokért felelős entitás. Szerzőink többsége magától értetődőnek találta, hogy az előállított adatokért elsősorban saját maguk felelősek – legalábbis addig mindenképpen, amíg a kutatás tart – bár ez is összefüggést mutathat azzal, hogy a bölcsészettudományok területén leggyakrabban egyedül, vagy kevés társszerzővel publikálnak a kutatók. A kitöltők harmada segítséget várna intézményétől a kutatási adatok kezelésében, szintén harmaduk pedig úgy véli, hogy ennek megszervezése főként a kutatás vezetőjének lenne a feladata (2. ábra).

A következő kérdésben arra voltunk kíváncsiak, hogy a kitöltők szerint kinek vagy kiknek a feladata biztosítani a kutatási adatok nyílt hozzáféréseinek költségeit. A válaszok a várt mintát mutatták, a nagy többség úgy vélte, hogy elsősorban az intézmény és a kutatásfinanszírozó feladata az adatkezelési költségek biztosítása. Saját véleményünk alapján keveselltük az olyan irányú választ, miszerint ezeket a költségeket, a megfelelő adatkezelési tervek alapján, már a pályázati költségvetésbe be kellene tervezni, majd a megszerzett forrásból dedikáltan finanszírozni. A kutatók mindössze negyede vélte úgy, hogy ez saját, kutatói feladat lenne. Ezt figyelembe véve a jövőbeli, az SZTE Klebelsberg Könyvtár által szervezett RDM-képzéseken erre nagyobb hangsúlyt érdemes fektetni.



1. ábra A kérdőív 2.1-es kérdésére adott válaszok százalékos megoszlása



2. ábra A kérdőív 2.3-as kérdésére adott válaszok százalékos megoszlása

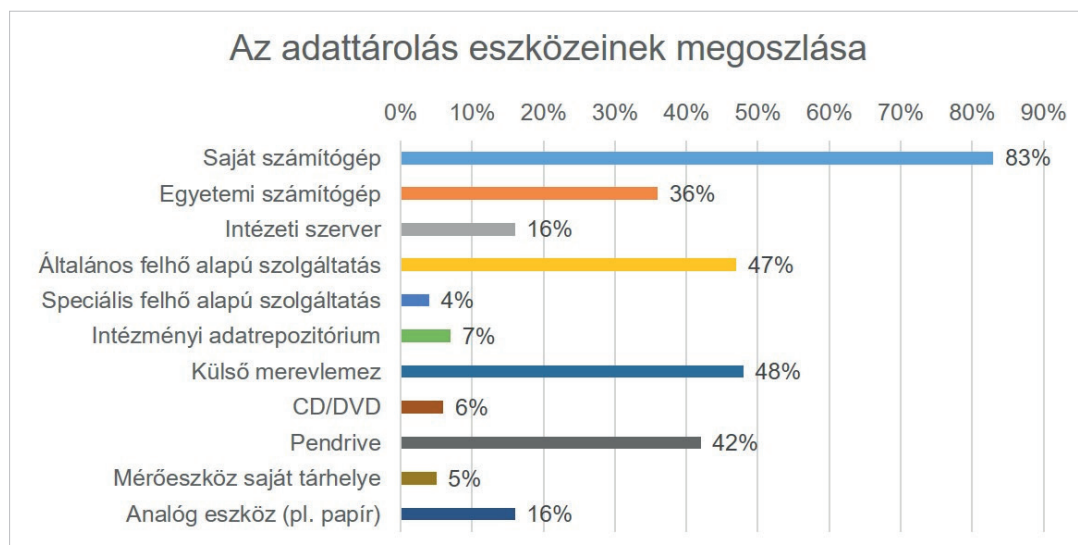
A kutatók nagy többsége evidensnek tartja, hogy az előállított adatok az intézmény birtokában maradnak, abban az esetben is, ha az egyes személyek elhagyják az egyetemet, negyedük viszont úgy vélte, hogy az adatokat alapértelmezetten magukkal vihetik következő munkahelyükre. Szerencsére azt csak egy kisebbség gondolta, hogy egy személy kilépése után az általa előállított kutatási adatokat megsemmisíti az egyetem a kilépést követően.

3.3 Adattárolás, archiválás és adatvesztés

A következő kérdéscsoportban a tárolandó adatok mennyiségét, méretét, illetve az egyetemen bevett gyakorlatnak számító adattárolási, archiválási szokásokat igyekeztünk megvizsgálni. A válaszadók fele éves szinten nem állít elő 10 GB-nál több kutatási adatot, így kutatóink többségének nem okoz különösebb problémát az adattárolás hardveres kapacitásának kérdésköre. A válaszadók alig tíz százaléka viszont 100 GB fölötti, egy százalékuk pedig a terabájtnyi mennyiséget is meghaladó adatmennyiséget állít elő évente. Ők már szinte egyöntetűen igényelnének valamilyen formájú intézményi segítséget az adattároláshoz és adatrepozitáláshoz.

A kérdőív alapján a leggyakrabban használt adattárolási eszközök a várható módon oszlottak meg (3. ábra). A legtöbben saját számítógépükön is tárolják az adataikat, de a kitöltők csaknem fele jelölte meg a külső merevlemezt, pendrájvot, DVD-t, valamint az általános felhőalapú adattároló szolgáltatásokat (pl. Google Drive, OneDrive stb.). A speciálisan kutatási adatok tárolására szolgáló felhőalapú szolgáltatásokat, adatrepozitáriumokat (pl. FigShare, Zenodo) csupán a kitöltők néhány százaléka használja, amely arány reményeink szerint a közeljövőben emelkedni fog.

A kitöltők több mint fele rendszeresen, egyharmada csak néha készít biztonsági mentést a kutatása során. Azoknak az aránya, akik sohasem készítenek biztonsági mentést, szerencsére alacsonynak mondható. Ennek ellenére a válaszadók több mint egyharmadánál fordult már elő adatvesztés a kutatás során, legtöbbször figyelmetlenség vagy hardverhiba miatt. Az ebből adódó kellemetlenségeket, ha többségében nem is élték meg tragédiának, azért mindenképpen a kutatást hátráltató tényezőnek tartották a kutatók.



3. ábra A kérdőív 3.3-as kérdésére adott válaszok százalékos megoszlása

A kérdőív válaszai alapján a kutatók túlnyomó többsége vezet dokumentációt az előállított kutatási adatairól, bár csak alig a felük teszi ezt rendszeresen. Metaadatokkal sajnos csak elenyésző részük látja el rendszeresen az eredményeket. Problémásnak ítéltük, hogy akik egyáltalán használnak metaadatokkal, azok sem egységesen teszik ezt. Sokszor ugyanazon a projekten, kutatáson dolgozók is eltérő módszereket alkalmaznak.

3.4 Az adatok hozzáférhetősége, megosztása és újrahaznosítása

A következő, negyedik blokkban a kutatási adatok hozzáférhetősége, megosztása és újrahaznosítása volt fókuszban. Sajnos a kutatási adatok kezelésének FAIR (Findable, Accessible, Interoperable, Reusable) alapelveiről⁸ nagyon kevesen nyilatkoztak úgy, hogy száz százalékgig tisztában lennének azok mibenlétével. Tudományterülettől függetlenül a válaszadók több mint fele nem is hallott a FAIR alapelvekről és a fennmaradók további, több mint 40%-a is bizonytalanak ítélik saját tudását a témában. Ennek megfelelően a következő kérdésre adott válaszok, miszerint „Kutatási adatai mennyire felelnek meg a FAIR alapelveknek?“, a bizonytalanságot és az ismeretek hiányát tükrözi.

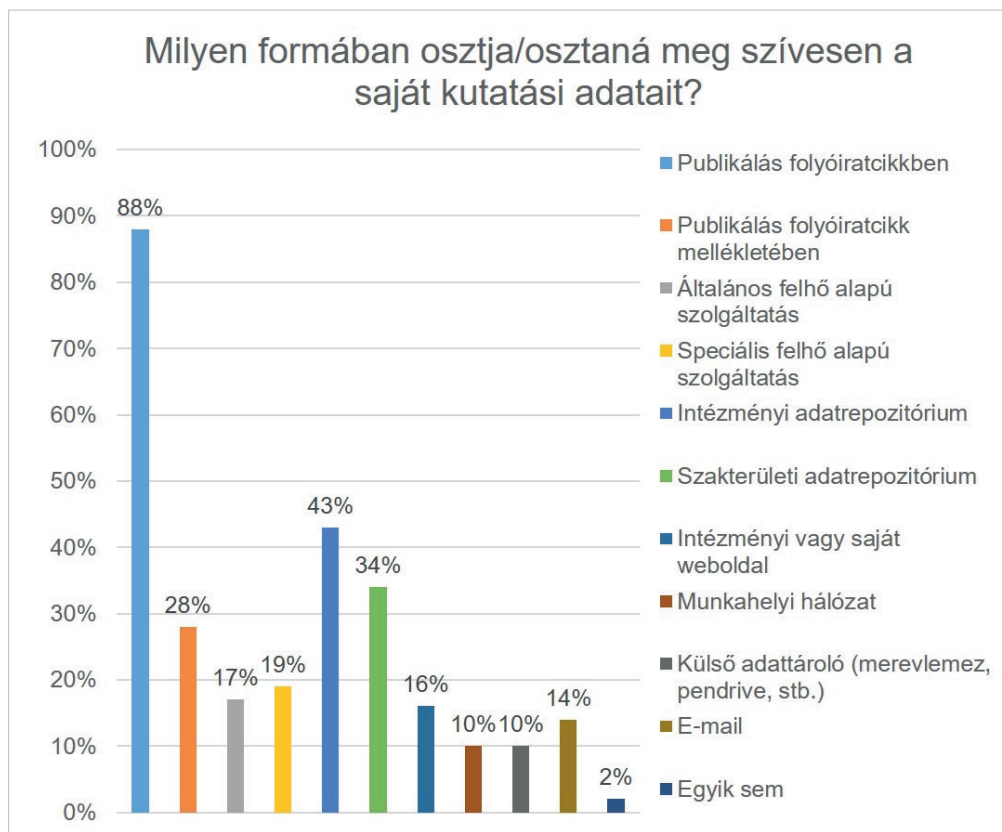
A kutatási adatok újrafelhasználhatóságával a kitöltők többsége nagy mértékben egyetért. Két-

harmaduk – bár változó gyakorisággal – már használt fel munkája során mások által előállított adatokat és szívesen használna ilyeneket a jövőben is. A kutatók negyede rendszeresen, fele pedig alkalmanként teszi nyíltan elérhetővé saját kutatási adatait szélesebb körben, nem csak a közvetlen kollégák számára elérhető módon.

A kutatási adatok megosztását válaszadóink elsősorban a hagyományos formában folyóiratcikkek, vagy folyóiratcikk-melléklet formájában tervezik. Bár működő példát nem nagyon találunk rá Magyarországon, a kitöltők több mint 40 százaléka szívesen használna intézményi adatrepozitóriumot az adatok megosztására, nyilvánvalóan az intézményi publikációs repozitóriumok ismertsége és sikere miatt. A szakterületi adatrepozitóriumokat, valamint multidiszciplináris repozitóriumokat a kutatóink több mint fele szívesen használna, vagy már használja is. Megjegyzendő, hogy újfent tudományterülettől függetlenül kaptunk ilyen válaszokat, amely kellemes meglepetést jelentett, mert a bölcsészek és a társadalomtudósok is szép számmal választották ezt az opciót.

Az adatok megosztásának több, hagyományos, illetve bizonyos esetekben kevésbé professzionális eszköze is előkerült a válaszok között. A kitöltők egészen nagy arányban jelezték, hogy pendrájvon, mobil winchesteren, Google Driveon keresztül szívesen osztanak meg kutatási adatokat (valószínűleg csak közeli kollégákkal) (4. ábra).

8 FAIR kutatási adatkezelés - Open Science. Hozzáférés: 2021. 07. 30 <https://openscience.hu/f-a-i-r-kutatasi-adatkezeles>



4. ábra A kérdőív 4.9-es kérdésére adott válaszok százalékos megoszlása

A válaszadók fő motivációja az adatok megosztására természetesen kutatásaik láthatóságának és hatásának elősegítése, kutatásaik igazolása, az adatok jobb átláthatósága és újrafelhasználhatósága, valamint a szakmai elismertség növelése volt.

Harmaduk a szokásos kiadói, pályázati, tudományterületi és intézményi elvárásokat is megjelölte az adatok megosztásának indokaként. A válaszadók kevesebb, mint két százaléka vallotta azt, hogy soha, semmilyen körülmények között sem osztaná meg adatait.

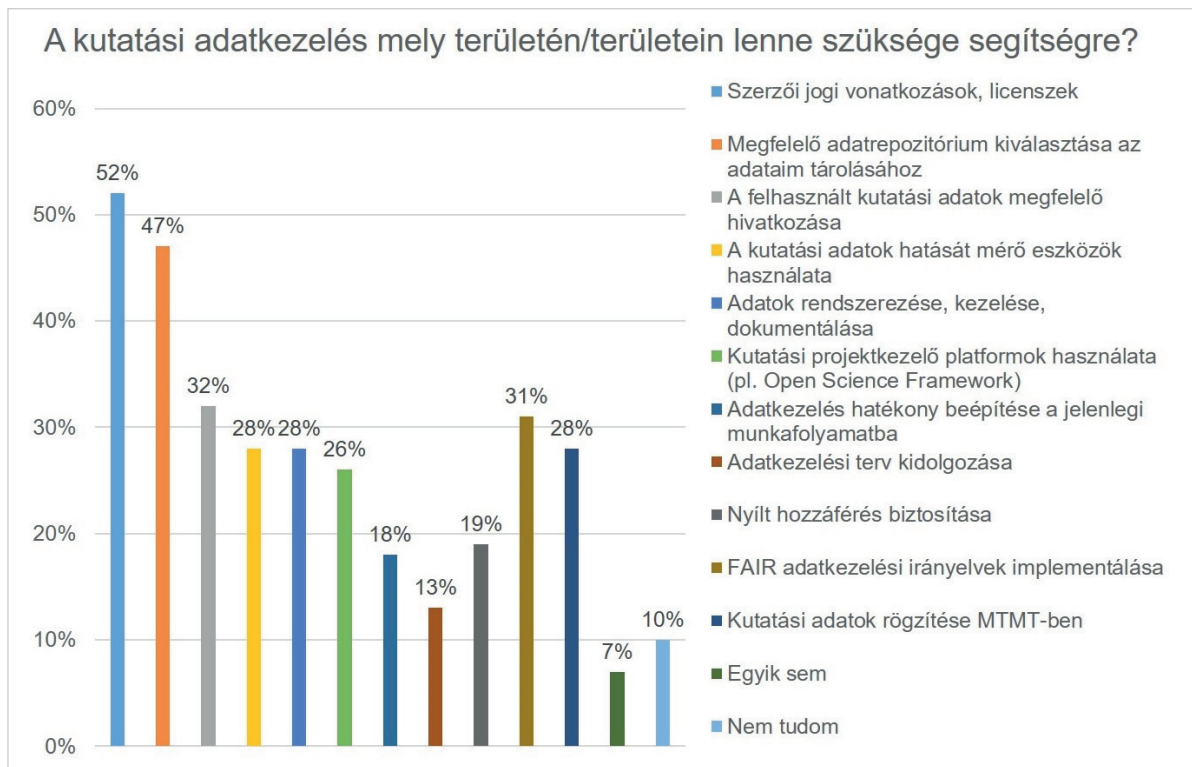
Mindezek ellenére majdnem minden kutatónak van valamilyen fenntartása vagy félelme az adatok megosztásával kapcsolatban. Legtöbben nem jártasak eléggé a szerzői, adatmegosztási jogok és a licencek használatában, illetve bizonytalanok a szenzitív adatok megosztásának legális módjában. A válaszadók harmada fél attól, hogy megosztott adatait illetéktelenül használnák fel mások, megfelelő idézés nélkül. A válaszadók ötöde jelölte meg, hogy az adatok megosztása előtt szükséges rendszerező munkára és a kapcsolódó adminisztratív tevékenységekre nincsen elég ideje.

3.5 Adatkezeléssel kapcsolatos igények

Saját, könyvtárosi nézőpontunkból talán az utolsó, 5. kérdéscsoport volt a legérdekesebb, mert ebben mértük fel a kutatók adatkezeléssel kapcsolatban megfogalmazott konkrét igényeit, és a könyvtár irányában támasztott elvárásaikat (5. ábra).

A grafikonon jól látszik, hogy a válaszadók jelentős része szinte minden felkínált lehetőséget megjelölt. Nem meglepő, hogy a kitöltők több mint fele igényelne szerzői jogi kérdésekben és a licencváltáshoz segítséget, hiszen ezek a hagyományos publikálásnak is kényes területei. Az, hogy a kitöltők közel fele érzi úgy, hogy segítségre szorul a megfelelő adatrepozitórium kiválasztásában, egyúttal azt is jelentheti, hogy a kutatási adatok hozzáférhetővé tételére irányuló külső nyomás és/vagy saját belső motiváció jelentősen növekedett az utóbbi időben.

Az 'Egyik sem' és 'Nem tudom' válaszlehetőségeket leszámítva a válaszokból az derül ki, hogy az adatkezelési terv kidolgozásához van legkevésbé szüksége segítségre a kitöltőknek. Ez azért volt meglepő számunkra, mert az ebben a témában tartott útmutató webinarjaink rendkívüli népszerűség-



5. ábra A kérdőív 5.1-es kérdésére adott válaszok százalékos megoszlása

nek örvendtek a kutatóink körében, és az ezidáig elindított szolgáltatások közül is az adatkezelési terv konzultáció a legkihasználtabb.

Rákérdeztünk, hogy a kitöltők kihez fordulnának először kutatási adatkezeléssel kapcsolatos tanácsért, és örömmel láttuk, hogy 30–30 százalékuk fordulna közvetlen kollégái mellett a könyvtárhoz. Ezzel, illetve az előző kérdésre adott válaszokkal szemben némileg ellentmondásos, hogy a válaszadók 15,5%-a egyáltalán nem venne részt kutatási adatkezeléssel kapcsolatos tréningen, amely ellentmondás áthidalása érdekes kihívást jelenthet a könyvtárosok számára. A kitöltők fele azonban szerencsére hajlandóságot mutat a részvételre RDM témájú online workshopon.

A könyvtár adatkezelési szolgáltatásaival kapcsolatos elvárások tekintetében pozitívként értékeljük, hogy a kitöltők több mint fele szinte az összes felkínált válaszlehetőséget megjelölte. Leginkább egy saját, SZTE-kutatók számára fenntartott adatrepozitórium létrehozására, valamint egy kutatási adatkezelésre vonatkozó, elméleti és módszertani segédlet összeállítására lenne igény. Hasznosnak tartanák továbbá, ha a PhD képzés tanrendjében megjelenne egy adatkezelési tematikájú kur-

zus, illetve ha a könyvtár kidolgozna egy intézményi szinten egységes FAIR adatkezelési irányelvet.

A 116 válaszadó közül mindössze kettő nyilatkozott úgy, hogy nincsenek elvárásai RDM-témában a könyvtárral szemben, és nincs szüksége a könyvtár által nyújtott kutatási adatkezeléssel kapcsolatos szolgáltatásokra. Ők ketten RDM-témájú tréningen sem vennének részt, és nem tartanak igényt segítségre az RDM egyik területén sem.

4. Tipikus kutatói hozzáállások

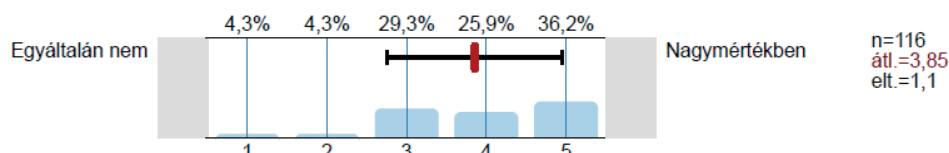
A kérdőív egyes, főleg az adatok hozzáférhetőségére, megosztására és újrahaznosítására vonatkozó részében feltett kérdéseire adott válaszok alaposabb elemzésével érdekes kutatói attitűdökre derült fény. Elmondható, hogy a továbbiakban vizsgált kérdések tekintetében az egyes tudományterületek között nincsen számottevő különbség.

Habár a kitöltők kétharmada (90 fő) nyilatkozott úgy, hogy létezik vele szemben tudományterületi elvárás a kutatási adatai publikálására; mégis, ahogyan az a 6. ábrán is látszik, az egynegyedük soha, mintegy felük pedig csak néha teszi nyíltan elérhetővé a saját kutatási adatait. Ez vagy arra utal, hogy az elvárás még csak informális, és sok esetben nin-

Van tudományterületi elvárás Önnel szemben a kutatási adatai publikálására?		Milyen gyakran teszi nyíltan elérhetővé a saját kutatási adatait?		
		Mindig	Néha	Soha
	Igen	27,8%	45,6%	26,7%
	Nem	12,5%	68,8%	18,8%
	Nem tudom	10,0%	60,0%	30,0%

6. ábra A kérdőív 1.4 és 4.6 kérdéseire adott válaszok közötti összefüggés

4.3) Egyetért-e azzal, hogy a kutatási adatok újrahaznosíthatóvá váljanak?



7. ábra A kérdőív 4.3-as kérdésére adott válaszok megoszlása

Használt-e már mástól (pl. adatrepozitóriumból) származó kutatási adatokat?		Milyen gyakran teszi nyíltan elérhetővé a saját kutatási adatait?		
		Mindig	Néha	Soha
	Igen, többször	38,9%	47,2%	13,9%
	Ritkán	23,5%	70,6%	5,9%
	Nem	13,0%	37,0%	50,0%

8. ábra A kérdőív 4.4 és 4.6 kérdéseire adott válaszok közötti összefüggés

csen következménye annak, ha nem teljesül; vagy arra, hogy ha az adatokat meg is osztják, azt nem nyíltan teszik, tehát az Open Science mozgalomnak van még feladata ezen a téren.

Pozitív irányba mutató tendencia, hogy akikkel szemben nincsen tudományterületi elvárás a kutatási adatok publikálására, vagy nem tudnak róla, azoknak is jóval több, mint fele teszi legalább esetenként nyíltan elérhetővé a kutatási adatait.

Ehhez társul még a 7. ábráról leolvasható egyértelmű tény, hogy a válaszadók túlnyomó része valamilyen mértékben egyetért azzal, hogy a kutatási adatoknak újrahaznosíthatóvá kell válniuk. Ugyanez az arány figyelhető meg egyébként azon kutatók körében is, akikkel szemben van tudományterületi elvárás a kutatási adatok publikálására: kisebb-nagyobb mértékben, de ők is egyetértenek a kutatási adatok újrahaznosíthatóvá tételére irányuló törekvésekkel. Ezek alapján úgy tűnik, hogy a nemzetközi szinten egyre erősödő trendek^{9,10}, a különböző szak-

mai szervezetek¹¹ törekvései a tudomány nyíltabbá tételére, és az egyes kutatók saját motivációja nagymértékben megegyeznek.

Akad azonban némi ellentmondás a kitöltők kutatási adatok újrahaznosíthatóságához való hozzáállásában. Egyértelmű, hogy a kutatók szívesebben használják fel mások kutatási adatait, mint teszik nyíltan elérhetővé a sajátjaikat. Annak a 30 kitöltőnek, aki soha nem teszi nyíltan elérhetővé az adatait (akár van elvárás vele szemben, akár nincs), több mint fele, 18-an értenek egyet nagymértékben az adatok újrahaznosíthatóvá válásával. A más kutatók adatai iránti érdeklődést jelzi az is, hogy bár a kitöltők majdnem fele még sosem használta fel mások adatait saját kutatásához, egyharmaduk nyilatkozott úgy, hogy szívesen használna fel ilyen adatokat a jövőben.

A 4.4–4.6 kérdésekre adott válaszok közötti összefüggéseket vizsgálva kiderül, hogy tulajdonképpen a nyíltan elérhető adatok felhasználásának gyakorlata jelentős korrelációt mutat a saját kutatási adatok nyíltan elérhetővé tételével. A 8. ábrán látható százalékos arányokból arra tudunk követ-

9 The EU's open science policy. Hozzáférés: 2021. 08. 02
https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en

10 UNESCO Recommendation on Open Science. Hozzáférés: 2021. 08. 02
<https://en.unesco.org/science-sustainable-future/open-science/recommendation>

11 Többek között: OpenAIRE (<https://www.openaire.eu/>), SPARC (<https://sparcopen.org/>), COAR (<https://www.coar-repositories.org/>). Hozzáférés: 2021. 08. 02

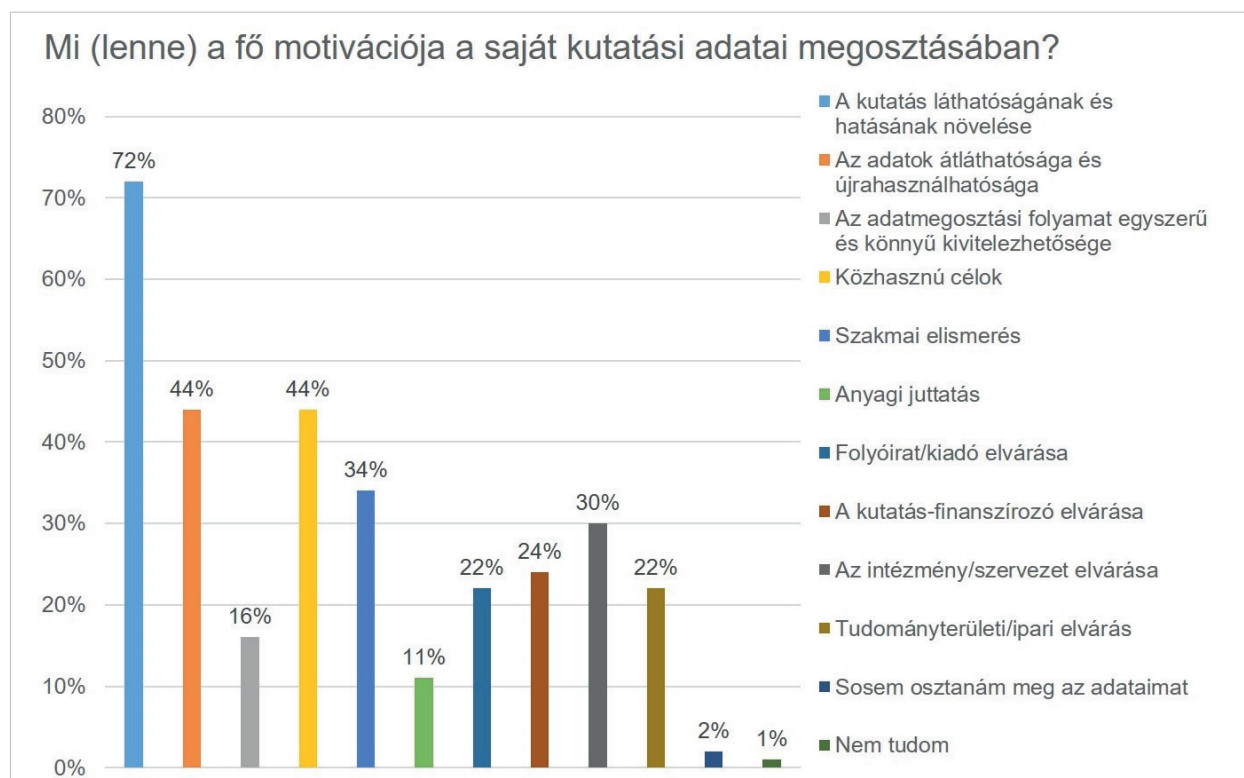
keztetni, hogy azok a kutatók, akik felhasználták már mások kutatási adatait, hajlamosabbak megosztani a sajátjukat. Ez a megfigyelés azon egyszerű tényből is következhet, hogy az aktív használók jobban megismerik azokat a felületeket, ahol a kutatási adatok megosztása történik, így oda-vissza alapon kedvet kapnak a nagyobb mértékű bevonódásra.

Azon kitöltők legnagyobb része (majdnem fele), akik már többször is felhasználták mások kutatási adatait, néha (tehát legalább egyszer) teszik nyíltan elérhetővé saját adataikat. Akik ritkán használják fel mások adatait, azoknak már csaknem a háromnegyede teszi nyíltan elérhetővé saját kutatási adataikat – amint látható az ábrán, ez a korreláció a legerősebb. Azon válaszadók fele, akik még sosem használták fel mások adatait, soha nem is teszi nyíltan elérhetővé a sajátját. Felmerül a kérdés, hogy azokat a kutatókat, akik eddig nem, vagy csak kismértékben járultak hozzá a kutatási adatok újrafelhasználhatóvá tételéhez (azaz soha nem tették még nyíltan hozzáférhetővé a saját adataikat), mivel lehetne erre motiválni.

A válaszok alapján elsősorban a kutatás láthatóságának és hatásának növelése jelent motivációt a kutatóknak, második helyen pedig a közhasznú

célok mellett az adatok átláthatósága és újrahaználhatósága áll. Ez az eredmény egybevág az Open Science mozgalom filozófiájával, ám azt is láttuk, hogy még mindig nagyon magas azoknak az aránya, akik ennek ellenére sem osztják meg kutatási adataikat. Pozitívum azonban, hogy a „Sosem osztanám meg az adataimat” válaszlehetőséget csak nagyon kevesen jelölték meg; ebből az feltételezhető, hogy akik eddig ugyan még sosem osztották meg az adataikat, azok nagy része nem zárkózik el ettől a jövőben. (9. ábra) Szükségesnek tűnik tehát, hogy a könyvtárak is aktívabb szerepet vállaljanak az adatok átláthatóságának és újrahaználhatóságának népszerűsítésében és elsajátításában.

Az adatok megosztása mögött rejlő motiváció mellett érdekes lehet azt is megvizsgálni, hogy milyen visszatartó tényezők miatt dönthet valaki úgy, hogy nem osztja meg az adatait. A 4.6 („Milyen gyakran teszi nyíltan elérhetővé a saját kutatási adatait?”) kérdésre „Soha” választ adók kétharmada a szerzői, adatmegosztási jogokkal és a licencekkel kapcsolatos bizonytalanságot jelölte meg aggodalomként. Ez szintén egy olyan terület, amelyben a könyvtár hasznos segítséget tud nyújtani, és amint az erre vonatkozó kérdésekből láttuk, erre igény is



9. ábra A kérdőív 4.10-es kérdésére adott válaszok százalékos megoszlása

mutatkozna. A kitöltőknek további komoly fenntartásaik vannak – teljesen érthető módon – a megosztott adatokkal való visszaélések és a szennyezett információk megosztása miatt is.

Sokan kételkednek továbbá a mások által megosztott adatok megbízhatóságában – ez szintén jogos félelem, és ez az a terület, ahol a kutatási adatok megosztására szolgáló platformoknak, illetve még inkább azok kiválasztásának nagy szerep jut. Ebben a kérdéskörben a könyvtárnak mindenképpen szakszerű és releváns segítséget kell tudni nyújtania a megfelelő adatrepozitóriumok kiválasztásában. A probléma egy másik, párhuzamos megközelítésű megválaszolását jelenthetné egy saját, az SZTE kutatói számára létrehozott és fenntartott, az SZTE Egyetemi Könyvtár által üzemeltetett adatrepozitórium.

Általánosságban elmondható, hogy az adatrepozitóriumoknak valamilyen módon megoldást kell találniuk a bennük tárolt adatok minőségbiztosítására¹², ahogyan legalább ennyire fontos volna a tudományos anyagok valóban hosszú távú, biztonságos, stabil megőrzési és elérési feltételrendszerének kialakítása. Ez utóbbi különösen olyan közelmúltbeli történések fényében érdekes, mint bizonyos COVID-19 génszekvenciák National Institutes of Health (NIH) adatbázisából való törlése¹³, vagy a Vuhani Virologiai Intézet teljes vírusadatbázisának internetről való eltűnése 2019 szeptemberének közepén¹⁴.

5. Összefoglalás

A bemutatott eredmények alapján elmondható, hogy érdemes volt összeállítani a kérdőívet, mert láthatóan van létjogosultsága az ilyenfajta felméréseknek. Hasznos eszközként szolgált ahhoz, hogy reális képet kapjunk az SZTE sokszínű kutatói álló-

mányának aktuális, RDM-hez való hozzáállásáról. Bár a minta kétségkívül lehetett volna nagyobb és heterogénebb, a kitöltés idején fennálló pandémiás helyzetet és a hasonló jellegű önkéntes egyetemi kérdőívek kitöltési arányát figyelembe véve elégedettek lehetünk. Külön öröme adott okot, hogy egy kivétellel minden, az SZTE-n művelt tudományterület képviseltette magát legalább egy kitöltővel.

A kutatási adatkezelés számos területén szerepet vállalni kívánó könyvtárosként mindenképpen fontos és érdekes tapasztalatokat szereztünk a kérdőív kiértékelésével. Jól látszik, hogy a kutatási adatok kezelésének problematikája egyre nagyobb szerepet kap a tudományos kommunikációban, ilyenformán minden kutatót érint, a kezdő PhD hallgatótól a pályája vége felé járó professzorig. A felmérés segítségével első kézből tapasztalhattuk, hogy nincs ez másként a Szegedi Tudományegyetemen sem.

A kutatók egy – vélhetően egyre kisebb – része érezhető ellenállást mutat a kutatási adatok szabványok szerinti, „felülről szabályozott” kezelésével, illetve azok nyíltan hozzáférhetővé tételével szemben. Felesleges tehernek, sőt, akár a munkájukra nézve komoly veszélyforrásnak tekintik ezeket a törekvéseket. Emellett azonban megfigyelhető egy sokkal pozitívabb, és főleg proaktívabb szemléletmód is a kutatók körében, amely már egyértelműen az RDM- és az Open Science mozgalom sikerét jelzi előre ezen a részterületen is.

Az írásunkban részletesen bemutatott és elemzett válaszokon túl legalább ennyire fontos kiemelnünk a háttérben zajló folyamatot, amely a könyvtáros kollégáktól nagyfokú nyitottságot és elköteleződést igényel egy új típusú szemlélet iránt. A kérdőív összeállítása, elemzése, illetve az ennek nyomán elinduló RDM-szolgáltatások¹⁵ megtervezése, majd működtetése komoly tanulási folyamatot igényelt a résztvevő kollégáktól. Mivel a kutatási adatkezelés esetében egyértelműen egy most formálódó területről beszélünk, melynek jó gyakorlatai, sőt az alapinfrastruktúrája csak most van kialakulóban, ezért különösen fontosnak gondoljuk, hogy bekapcsolódjunk ebbe a szakmai diskurzusba.

12 Lsd. például: CoreTrustSeal Trustworthy Data Repositories Requirements. Hozzáférés: 2021. 08. 03

<https://www.coretrustseal.org/why-certification/requirements/>

13 Bloom J. D. (2021). Recovery of deleted deep sequencing data sheds more light on the early Wuhan SARS-CoV-2 epidemic. bioRxiv.

<https://doi.org/10.1101/2021.06.18.449051>

14 Anon, A, Bostickson, B, & Demaneuf, G. (2021). Investigation into Wuhan Institute of Virology Pathogen Databases (Version 2). Zenodo.

<https://doi.org/10.5281/zenodo.4512260>

15 Kutatási adatkezelési szolgáltatások – SZTE Klebelsberg Könyvtár

<http://szerzoknek.ek.szte.hu/szolgáltatásaink>

Beérkezett: 2021. augusztus 6.



Zeller Rozália

SZTE Klebelsberg Könyvtár
osztályvezető

E-mail: rozalia.zeller@ek.szte.hu

ORCID: 0000-0003-2501-8760



Hoczopán Szabolcs

SZTE Klebelsberg Könyvtár
osztályvezető

E-mail: szabolcs.hoczopan@ek.szte.hu

ORCID: 0000-0002-7892-9974



Nagy Gyula

SZTE Klebelsberg Könyvtár
főigazgató-helyettes

E-mail: gyula.nagy@ek.szte.hu

ORCID: 0000-0002-8391-2851