

PERSISTENT IDENTIFIERS (PIDs - ÁLLANDÓ, EGYEDI AZONOSÍTÓK) KUTATÁSI EREDMÉNYEK AZONOSÍTÁSÁRA

Holl András; Bilicsi Erika
MTA Könyvtár és Információs Központ

<https://doi.org/10.14755/mtakik.arp.2023>

TARTALOMJEGYZÉK

ÖSSZEFOGLALÓ	2
Bevezető	2
1.) Elterjedt egyedi azonosító megoldások	3
Handle.....	3
Digital Object Identifier (DOI).....	4
Archival Resource Key (ARK).....	6
Persistent Uniform Resource Locator (PURL).....	7
2.) Alkalmazás körülményei, granularitás, verziók	8
3.) A DOI azonosító alkalmazásának előnyei	8
4.) A DataCite DOI ügynökség szolgáltatása	10
5.) MTA KIK DOI Iroda.....	19
6.) DataCite DOI regisztráció tapasztalatok	22
7.) ARK és DOI	23
8.) DOI és ARK használat integrálása, hierarchikus azonosítás, verziók azonosítása	24
9.) Adatok hivatkozása	25
10.) További egyedi azonosítók (kutatók, szervezetek azonosítása).....	25
ORCID.....	26
ROR.....	27
Irodalom.....	28
URL-ek:	30

A kézirat lezárva: 2023. november 7.

Tanulmányunk az egyedi digitális objektumazonosítók használatát tárgyalja. A különböző dokumentumazonosítók használatának előnyeit és hátrányait összefoglalva megállapítjuk, hogy míg az ARK azonosító alkalmas és ajánlható digitális kutatási adatelemek (alacsony szintű, és ezért nagyszámú, elemi adatok) adatbázison belüli vagy technikai azonosítására, a publikálandó, várhatóan hivatkozott, magas szintű adat-aggregátumok azonosításának nemzetközileg elterjedt és általunk ajánlott módja a DataCite ügynökség DOI azonosítóinak használata. A DOI-k alkalmazása képes támogatni a hosszú távú megőrzést, a globális láthatóságot, de legfőképpen a hivatkozások nemzetközi keringésbe való bekerülését, tudományometriai felhasználhatóságát (amely hozzájárul az egyéni és intézményi globális rangsorokban való értékeléshez). Mindezen túl a DOI is biztosítja az elérési állandóságot, amit a többi azonosító használata is megold.

Az egyedi azonosítók alkalmazása – legfőképpen a nemzetközi gyakorlatban bevett DOI-ké – hatalmas potenciállal bír. A remélt előnyök eléréséhez azonban szükséges a szabványoknak megfelelő precíz használat, és az alkalmazott infrastruktúrák és szervezetek állandósága és működési biztonsága.

A digitális objektumok azonosítására alkalmazható egyedi azonosítók

	Hosszútávú elérés biztosítása	Széles körben elterjedt	Szoftveresen létrehozható	Metaadatokat tárol	Támogatja a tudományometriai felhasználást	Támogatja a hierarchia jelölését	Támogatja a verziókezelést	Használata díjfizetéshez kötött
DOI	x	x	x	x	x	x	x	x
Handle	x		x			x		x
ARK	x		x	x		x	x	
PURL	x		x					

Az egyedi, állandó azonosítók lehetővé teszik a kutatásokban résztvevők és a kutatási eredmények – azaz a kutatók, intézmények, támogatók, támogatások, publikációk, kutatási adatok, szoftverek stb. – azonosítását és összekapcsolását, ezáltal pedig az automatizált adatgyűjtést és lekérdezések, elemzések, riportok készítését. Ennek az ökoszisztémának az egyik eleme a kutatási adatok azonosításának a kérdése.

A kutatási adatok idézhetőségének biztosításával, egyedi, állandó azonosítókkal (persistent identifiers, PIDs) való ellátásával – ezen belül a DOI (digital object identifier) használatával erre a célra – már több, mint egy évtizede foglalkozik a szakirodalom (Simons, 2012).

Egy, az (digitális) objektum életét végigkísérő egyedi azonosító megléte önmagában fontos. Azonban a modern, egyedi, állandó azonosítók további, lényeges követelményeknek kell megfeleljenek:

a.) Lehetővé kell tenniük a digitális objektum elérését hosszú időtávon – akkor is, ha a tárolás/szolgáltatás helye időközben megváltozik. Ezt egy közbülső adatbázis közbeiktatásával érik el: a hivatkozásban használandó cím állandó, amíg az objektum elérési címét a közbülső adatbázisban karbantartják, a HTTP kérést oda átirányítják.

b.) Gépileg kezelhetők kell legyenek (machine actionable): nem csupán a dokumentumhoz kell elvezessenek, de a metaadataihoz is; gépileg létrehozhatónak, bejegyezhetőnek kell lenniük.

c.) Metaadatokat kell tároljanak egy (központi) API-n keresztül elérhető adatbázisban az azonosítóval rendelkező objektumokról, melyek segítenek a digitális objektum felderítésében, láthatóságának növelésében.

d.) Be kell tagozódjanak a tudományos kommunikáció alapvető technológiai közé, lehetővé téve az azonosított objektumok láthatóságának növelését, azaz használatának elősegítését, megteremtve a tudományometriai felhasználás lehetőségét, módot adva az adatbázisok közötti (automatikus) adatáramlás technikáinak kialakítására.

Az elterjedten használt egyedi, állandó azonosítók a fenti kritériumoknak különbözőképpen felelnek (vagy nem felelnek) meg. A technológiai különbségeken túl különbözhetnek az egyes azonosító-félék alkalmazásának költségei és adatbiztonsága is, amint ezzel összefüggésben az egyes azonosítófajták globális vagy lokális jellege is különbözik.

1.) ELTERJEDT EGYEDI AZONOSÍTÓ MEGOLDÁSOK

Az egyedi azonosítókat ismerteti egy évtizeddel ezelőtti, az MTMT számára készült tanulmány (HU-DOI: Karácsony és Holl, 2011), a THOR Persistent Identifier Project weblapja (Getting Started -> What are persistent identifiers? Overview / Features / Selection), valamint az NRL LibGuides tájékoztató oldala. Az alábbi leírásokat a HU-DOI tanulmányból idézett leírásokkal gazdagítjuk.

HANDLE

Nem kereskedelmi alapon fejlesztett decentralizált azonosító rendszer, az Egyesült Államokban működő Corporation for National Research Initiatives kezelésében. A DOI mögött álló állandó elérési technológia.

A b., c. és d. kritériumnak nem felel meg, mivel globális (vagy lokális) bővebb metaadattára nincsen (de az URL a DNS-hez hasonló, elosztott adatbázisban tárolt).

Az azonosító felépítése egyezik a DOI-ével (ld. lentebb). A „20.” előtag a HANDLE.NET rendszert jelöli, a „10.” előtag a DOI rendszert.

A gyakorlatban a 20-as előtagot nem használják, egy Handle azonosító így néz ki:

<http://hdl.handle.net/2437/154879>

A Handle System egy általános célú elosztott információs rendszer, melyet eredetileg a DARPA támogatásával fejlesztettek. A DOI a Handle System egyik megvalósítása és alkalmazása.

A HDL© Azonosító és Névfeloldó Szolgáltatás némileg a Domain Name System-hez hasonlít.

A HANDLE.NET szabad szoftveren alapul. Használatához azonban regisztrációs díj ellenében prefixet kell bejegyeztetni a Global Handle Registry-ben. Egy globálisan bejegyzett „handle” egyedi azonosító, ilyen azonosítókat használnak a DSpace szoftveren alapuló repozitóriumok. A handle azonosítók – ha globálisan regisztrálva vannak - egyediek, a probléma csupán az velük, hogy általában nem a dokumentum létrehozói, kiadói rendelik hozzá ezeket a dokumentumokhoz. Hiába van egy Dspace alapú repozitóriumban elhelyezett dokumentumnak handle-ja, ha a dokumentum szerzője a dokumentumot egy másik repozitóriumban is elhelyezte, ott ez az azonosító valószínűleg nem szerepel.

Az RFC 3650, 3651, 3652 szabványok írják le. Általános formája:

<Handle Naming Authority> "/" <Handle Local Name>

azaz egy, az egyes dokumentumok „számozásáért” felelős szervezetet jelölő előtagból és törtvonallal elválasztva az általa kiosztott utótagból áll.

Pl.: hdl:2437/2787

Ez URL-ként így érhető el: <http://hdl.handle.net/2437/2787>

(Forrás: HU-DOI: Karácsony és Holl, 2011)

ELŐNYÖK	HÁTRÁNYOK
<ul style="list-style-type: none">- világszerte használt- szabványos	<ul style="list-style-type: none">- a szerzők által nem ismert- nem tárol metaadatokat- tudományometriai célú felhasználásra nem alkalmas- használata regisztrációs díjhoz kötött

DIGITAL OBJECT IDENTIFIER (DOI)

A DOI a legelterjedtebben alkalmazott azonosító – havonta kb. 1 milliárd kattintást old fel a DOI rendszer – mind tudományos publikációk, mind kutatási adatok esetében. 2012 óta ISO szabvány ([ISO 26324](#)). A rendszerrel kapcsolatos információk fő forrása a [DOI Handbook](#). Az [International DOI Foundation](#) irányítása alatt áll, ami az ún. DOI regisztrációs ügynökségek irányító testülete. Jelenleg [12 DOI ügynökség](#) együttműködésével biztosítja a DOI rendszer szolgáltatásait. Az egyes ügynökségek más és más ügyfélkör számára megfelelő szolgáltatási portfóliót biztosítanak: pl. a CrossRef ügynökség a tudományos publikációk azonosítására, a DataCite ügynökség a kutatási adatok és szürkeirodalom számára, a Chinese DOI pedig az ázsiai ügyfelek kiszolgálására specializálódott. Előnye az elterjedtségen, szabványosságon túl

- a metaadattár: minden DOI ügynökség saját metaadattárat működtet, azaz csak ügynökségenként létezik központi adatbázis, de ez már megteremti a metaadatok cseréjének lehetőségét;
- az adatáramlást elősegítő megoldások léte és terjedése (CrossRef - ORCID kapcsolat);
- a scientometriai beágyazódás (DOI aratás a cikkekből).

Hátránya, hogy pénzbe kerül: a rendszer önfinanszírozó, azaz a költségeket a tagok befizetései fedezik.

Az azonosító felépítése:

<rendszer prefix (directory indicator, a DOI esetében 10)>.<szolgáltató prefix>/<szuffix>

A szuffix (utótag) e szolgáltatón belül egyedi, de a DOI rendszer szintjén nem hordoz információt. Az egyes szolgáltatók belekódolhatnak információt a felhasználható karakterek alkalmazásával, azonban DOI rendszer szinten ennek az információnak nincs funkciója.

A Digital Object Identifier (DOI©) egy, a HDL rendszeren alapuló dokumentumazonosító. A rendszert az International DOI Foundation tartja fenn, melynek főbb tagjai nagy tudományos kiadványok. A jelentősebb tudományos szakfolyóiratok cikkei DOI-val rendelkeznek, így az MTMT számára legjelentősebb dokumentumkör nagy részében létezik egyedi azonosító.

A DOI jelenleg ISO szabványosítás alatt áll (ISO/DIS 26324).

A DOI azonosítók általános formája a következő:

<DIR>". "<REG>"/"<DSS>

A DIR kód értéke 10, a DOI rendszert jelöli az előtagban. A REG az előtag második része a dokumentumok azonosításáért felelős szervezet számára kiosztott szám. A ponttal elválasztott DIR és REG együtt képezik a handle előtagját. A DSS a handle utótagja, mely lehet egy beszélő karakterlánc is. (Pl.: 10.1088/0004-637X/703/1/L72 ami a következő cikk DOI-ja: The Astrophysical Journal Letters, Volume 703, Issue 1, pp. L72-L75 (2009) Ebből a DOI-ból a következő módon képezhető URL, ami a cikkhez vezet: <http://dx.doi.org/10.1088/0004-637X/703/1/L72>).

A doi.org DOI feloldója (rezolvere) a <http://dx.doi.org/>. Ez használható manuálisan (egy mezőbe kell az azonosítót bemásolni a weblapon) vagy mint a fenti példa mutatja, az URL-ben mögé kell toldani.

A handle és a DOI a böngészők szintjén is támogatott - pl. Firefox kiegészítő létezik a kezelésükhöz. DOI azonosítókat csak regisztrációs ügynökségeken keresztül lehet szerezni. Ilyenek pl. a CrossRef és a DataCite.

(Forrás: HU-DOI: Karácsony és Holl, 2011)

(Megjegyzés: Ma már a <https://doi.org/10.1088/0004-637X/703/1/L72> alkalmazzák.)

ELŐNYÖK	HÁTRÁNYOK
<ul style="list-style-type: none"> - világszerte használt - a szerzők által is ismert - szabványos - metaadatokat is tárol - alkalmas tudományometriai célokra is - a közösségi szabályok elősegítik a megőrzést 	<ul style="list-style-type: none"> - használata díjfizetéshez kötött

ARCHIVAL RESOURCE KEY (ARK)

1995 óta létező rendszer, nincs mögötte szabvány. A California Digital Library szervezi, egy EZID (easy-eye-dee) nevű, kombinált DOI/ARK infrastruktúrára alapozva.

Lokális adatbázisokkal kiszolgálható - ez előnye is (költségek), de egyúttal hátránya is.

[https://NMA/ark:/NAAN/Name\[Qualifier\]](https://NMA/ark:/NAAN/Name[Qualifier])

Pl.: <http://example.org/ark:/12025/654xz321/s3/f8.05v.tiff>

A "Qualifier" komponens alkalmas a hierarchikus azonosító séma és a verziókezelés megvalósítására.

ARK, Archival Resource Key, California Digital Library.

Nem csupán állandó azonosítási koncepció, hanem teljes protokoll és szoftverkészlet, mely az állandó azonosítás teljes keretrendszerét és az azonosítók különböző információforrások irányába történő feloldását alapozza meg.

Az alapötlet abban rejlik, miszerint az állandóságot nem az elnevezési szintaxis, hanem a szolgáltatás adja. Az azonosító az azonosított objektum gondozottságát ígéri; az azonosító elvezet az azonosított objektum leírásához (metaadatok); az azonosító lehetőség szerint elvezet magához az objektumhoz.

AZ ARK koncepció egy állandó azonosításra szolgáló nyílt keretrendszer. Adminisztratív modellből, a hierarchiát és verziókat kifejező elnevezési szabályokból, valamint hálózati protokollokból áll.

ARK-k adminisztrálása

ARK kiosztást Name Assigning Authority (NAA) vagy egy NAA által meghatalmazott al-hatóság végezhet. Mindegyik NAA kap egy számot (Name Assigning Authority Number, NAAN). A NAA felelős a nevek objektumokhoz rendeléséért. A feloldási oldalon: mindegyik NAAN egy vagy több NMAH-hoz (Name Mapping Authority Hostport) kapcsolódik, melyek levágják az azonosító NMAH részét, s ebből döntenek el, hogy két ARK ugyanazt az objektumot azonosítja-e. A NMAH a szolgáltatás technikai helye, s egy vagy több NAA-t szolgál ki. A NAAN-ek aktuális listája publikus.

Az ARK névtér

[“http://” <NMAH> “/”] “ark:” <NAAN> “/” <Name> [<Qualifier>]

Kötelező része: „ark:” <NAAN> “/” <Name>

Az ARK olyan koncepció, mely az állandó azonosítást a szükséges technikai és adminisztratív kerettel együtt dolgozta ki.

Az ARK kétszintű hierarchikus névtér. Az első a gyökér, a második a névtérrel rendelkező NAA. Az ARK-ba integrálható más azonosító rendszer, s az ARK is integrálható valamely más rendszerbe.

Teljesen nonprofit, nincs kereskedelmi háttere.

(Forrás: HU-DOI: Karácsony és Holl, 2011)

Alacsony technikai feltételekre dolgozták ki (DNS, webservert, és webböngészőt a kliensoldalon), ezért a fenntartása könnyebb, mint a kifinomult, komplex szoftverek esetében.

ELŐNYÖK	HÁTRÁNYOK
<ul style="list-style-type: none">- nincs regisztrációs díj- alacsony technikai feltételekre épül- lokális adatbázissal igénybe vehető- támogatja a hierarchia jelölését és a verziókezelést	<ul style="list-style-type: none">- nem szabványos- lokális adatbázissal igénybe vehető- a szerzők által nem ismert- nem tárol metaadatokat- tudománymetriai célú felhasználásra nem alkalmas

PERSISTENT UNIFORM RESOURCE LOCATOR (PURL)

Általános azonosító és névfeloldó koncepció, alapvető internetes szabványokra építve. Olyan URL, ami (a fentebb említett azonosítókhoz hasonlóan) egy névfeloldó rendszeren keresztül éri el az azonosított objektumot. 1995-ben dolgozták ki az OCLC-ben. Az alap névfeloldó a purl.org (Weibel és Jul, 1995).

ELŐNYÖK	HÁTRÁNYOK
<ul style="list-style-type: none">- nincs regisztrációs díj- alacsony technikai feltételekre épül	<ul style="list-style-type: none">- nem szabványos- a szerzők által nem ismert- nem tárol metaadatokat- tudománymetriai célú felhasználásra nem alkalmas

2.) ALKALMAZÁS KÖRÜLMÉNYEI, GRANULARITÁS, VERZIÓK

Amíg egy kutatási adat repozitóriumon belül való azonosítására van szükség, bármilyen belső azonosító használható. A felhasználhatóságot növeli, ha valamilyen szabványos azonosító technológia kerül alkalmazásra, külső feloldhatóságot is biztosítva, metaadattárat is alkalmazva (mely elvileg adatbányászatra használható).

További előnye egy szabványos, de helyi azonosító alkalmazásának az, ha ennek nincsenek anyagi terhei (nincs külső félnek fizetendő regisztrációs díj). Ez különösen hasznos, ha az azonosítók viszonylag alacsony szinten (pl. adatrekordonként, mérési pontonként, rajzonként) alkalmazásra kerülnek. Ezt mutatja a használatban lévő azonosítók száma is: amíg 8.2 milliárd ARK azonosító van bejegyezve, addig 240 millió DOI.

Abban az esetben, ha nagyobb adatsomagokat kell azonosítani (projektenként, kísérletenként, lekérdezésenként, esetleg mérésenként – de nem mérési pontonként), olyan egységként, amelyek egyben való újrafelhasználása – így hivatkozása – elképzelhető, már érdemes globális (globális metaadattárral rendelkező) azonosítókat alkalmazni. Lehetséges megközelítés az is, hogy a globális azonosító használatát a külső felhasználáshoz, hivatkozáshoz kell kötni. A rendszer abban az esetben ad külső azonosítót, ha egy SQL lekérdezés eredményét, egy adatsomagot külső felhasználó igényel, vagy az adatgazda külső felhasználásra átad. Ilyen esetben lehet javallat, hogy a felhasználó a globális azonosítót használja az adatokra való hivatkozásnál.

Ezt a technológiát alkalmazzák dinamikus adatbázisok SQL lekérdezése útján keletkező, későbbi kutatás számára bemenetet képző adatok esetében is (pl. OpenBioMaps). A forrás adatbázis folyamatosan, dinamikusan változik, bővül – az adatbázis műveletek naplózása, verziókezelés, a mentések belső célokat szolgálnak. Viszont amikor kutatási célból egy SQL lekérdezés történik, ennek eredményét külön elmentik, és egyedi, globális azonosítóval látják el az ellenőrizhetőség és reprodukálhatóság érdekében. (Ugyanaz az SQL lekérdezés egy más időpontban más eredményt adna.)

A DOI szabvány nem nyújt lehetőséget egy dokumentum vagy adatsomag verzióinak kezelésére. A preprintek esetében a szakterületi repozitóriumok különbözőképpen kezelik a verziókat – egy lehetőség a „concept DOI” amit pl. a Zenodo is alkalmaz (COAR-ASAPbio, 2022).

3.) A DOI AZONOSÍTÓ ALKALMAZÁSÁNAK ELŐNYEI

Amikor az adatok tényleges (publikációban megtestesülő) felhasználásáról vagy újrafelhasználásáról van szó, kulcsfontosságú az adatok hivatkozhatóságának biztosítása és a hivatkozások összegyűjtésének lehetősége. A publikációk esetében egy azonosító van, amelyik kiemelkedően jó lehetőségeket biztosít ebben a tekintetben a többi közül: a DOI.

A tudományos kommunikáció és a tudománymetria terén egy kutatás hatását a hivatkozások számával mérik. Mind a hivatkozások számának növelésében, mind a kapott hivatkozások összegyűjtésének lehetőségében fontos lehetőségeket nyújt a DOI alkalmazása. A CrossRef megköveteli az irodalomjegyzékekben a DOI-k feltüntetését és linkként való alkalmazását (ez növeli a hivatkozott információ láthatóságát, használatát), illetve a DOI-k gyűjtése lehetőség arra, hogy hivatkozáskapcsolatokat kis befektetéssel gyűjtsenek (pl. a Digital Science Dimensions adatbázisa).

A kutatók közzétett adatainak hivatkozásai sokkal hatékonyabban gyűjthetőek össze a publikációk világában alkalmazott technológiák használatával, mint bármilyen más módon. Ésszerű – és esetenként már alkalmazott is – az adatállományokra vonatkozó hivatkozások a publikációs hivatkozásokkal együtt való kezelése. Korábban is működött már ez, egyedi azonosítók nélkül is: amikor egy kutatásban más kutatások által előállított adatokat használtak, arra a cikkekre hivatkoztak, amelyek ezeket az adatokat leírták, közzétették. Így az adatok használata esetén is azok szerzőinek publikációs hivatkozása keletkezett. Ugyancsak van arra példa, hogy amikor közvetlenül kutatási adatokat hivatkoztak, ahhoz publikációs egyedi azonosítókat használtak.

[Példa: DENIS adatok - az adatállomány:

<https://ui.adsabs.harvard.edu/abs/1999yCat.2240....0E/abstract>

BIBCODE: 1999yCat.2240....0E

MTMT azonosító: 2885732

Az adatállományra hivatkozó egyik cikkben (<https://doi.org/10.3847/1538-4357/ab881c>)

így hivatkoznak rá, az irodalmi hivatkozások között:

Epchtein N., Deul E., Derriere S. et al. 1999 yCat 2240 0]

A DOI alkalmazásának előnye, hogy a DataCite, a CrossRef és az ORCID együttműködése eredményeképpen a tudományos információáramlás fontos szolgáltatásai épülnek rá (ORCID Auto-update – mind a DataCite, mind a CrossRef adatbázisából át tudja venni az ORCID a szerzők profiljába az újonnan DOI-val azonosított közleményeket; CrossRef/DataCite Event Data – idézési visszacsatolás; FREYA projekt keretében fejlesztett PID Graph; COUNTER – megtekintések és letöltések jelentése). A DOI azonosítókhoz fejlett szolgáltatások állnak rendelkezésre, mint a DataCite Commons vagy a Google Dataset Search.

A CrossRef adatbázisának bányászata a Digital Science Dimensions adatbázisának egyik pillére. Ez, és az ehhez hasonló vállalkozások, illetve pl. a Clarivate Analytics Data Citation Index-ének bővítése arra ösztönzik a kutatói közösségeket, hogy az idézhető produktumaikhoz DOI azonosítókat használjanak.

Mindazonáltal figyelembe kell venni, hogy a DataCite külön metaadattárral rendelkezik, és a CrossRef-es technológiák az eltérő metaadatséma és metaadattár miatt nem alkalmazhatóak közvetlenül. Itt a különböző nemzetközi szervezetek közös fejlesztéseire, átjárhatósági kezdeményezéseire lehet támaszkodni (pl: DOI Citation Formatter, Freya, THOR).

A Research Data Alliance munkacsoportjai is foglalkoztak és foglalkoznak a publikációk és a publikált adatállományok összekötésével.¹

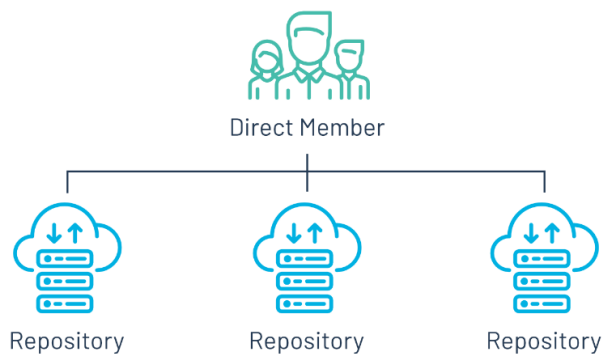
Azt is meg kell jegyezzük, hogy a DOI azonosító alkalmazása a kutatási adatok hosszútávú megőrzését is segíti, hiszen a DOI rendszerben előírás, hogy a csatlakozott intézményeknek gondoskodniuk kell az azonosított objektumok hosszútávú megőrzéséről és szükség szerint a DOI-khoz kapcsolt URL-ek módosításáról. Azaz ennek az azonosítónak a használata hozzájárul az adat-repozitóriumok használatának elterjedéséhez és a publikációkhoz kapcsolódó mellékletek megőrzéséhez is. Ugyanis megkönnyíti a kiadók munkáját is, hiszen ha a szerzők DOI-val hivatkoznak a tanulmányokhoz kapcsolódó adatokra, akkor a repositóriumok és nem a kiadók felelőssége a publikációkhoz kapcsolódó mellékletek megfelelő metaadatokkal való ellátása és hosszútávú elérhetőségének biztosítása – ami összetett feladat, hiszen ezek formátuma, mérete igen változatos. Végül, de nem utolsósorban pedig hozzájárul a Plan S-ben foglalt célok megvalósulásához is, hiszen a publikálási felületek számára határozottan ajánlott kritériumok között szerepel a közleményekben szereplő eredményeket alátámasztó adatokra történő hivatkozás alkalmazása.

4.) A DATACITE DOI ÜGYNÖKSÉG SZOLGÁLTATÁSA

A DataCite egy nonprofit szervezetként működő DOI regisztrációs ügynökség, ami elsősorban kutatási adatok és -gyűjtemények, szoftverek, modellek, ill. egyéb kutatási eredmények, ún. szürke irodalom (olyan tudományos kutatáshoz kapcsolódó dokumentumok, melyek repositóriumban jelennek meg – mint pl. kutatási jelentések, technikai dokumentációk, specifikációk, riportok, nem publikált konferenciaközlemények, disszertációk) számára biztosítja a DOI-val való azonosítást. A kutatásban résztvevő szervezeteknek tagként (Direct Member) kell csatlakozniuk² a DataCite-hoz vagy egy az egyesület tagjaként működő konzorciumhoz (Consortium Lead and Consortium Organisations), hogy DOI-val azonosíthassák kutatási produktumaikat. Létezik egy támogatói típusú tagság is, az ekként csatlakozó szervezetek mellett, hogy financiálisan segítik a DataCite működését, részt vesznek a DOI-t regisztráló közösség munkájában és lehetőségük van részt venni az egyesület irányításában is, de DOI-t nem regisztrálnak. Jellemzően kutatásfinanszírozó szervezetek döntenek emellett a csatlakozási forma mellett. A Direct Member-ként csatlakozó szervezetek a saját intézményük által fenntartott adattár(ak) számára regisztrálnak DOI azonosítókat. Konzorciumokat egy-egy szakterület vagy ország kutatóintézetei szervezhetnek, minimum 5 intézményből jöhet létre konzorcium. Ennek a csatlakozási modellnek előnye, hogy kisebb intézmények számára is lehetővé teszi a csatlakozást, egyrészt mert ebben a formában a szervezet számára kedvezőbb a díjszabás, másrészt mert a konzorciumvezetést felvállaló szervezet általában biztosít adattárat – de lehet sajátja is a konzorciumba csatlakozó intézménynek – és a regisztrációt támogató szolgáltatásokat is, így helyben elegendő alacsonyabb szintű szaktudás és kisebb létszámú személyzet fenntartása.

¹ [RDA/WDS Publishing Data Services WG](#) ; [RDA/WDS Scholarly Link Exchange \(Scholix\) WG](#)

² A csatlakozási információkat ld. részletesen: <https://datacite.org/become.html>



1. kép: A DataCite tagként csatlakozott szervezetek több repozitóriumi gyűjteményük számára is regisztrálhatnak DOI-kat



2. kép: A DataCite konzorciumban a konzorciumot vezető szervezet is DOI-t regisztráló tagként van jelen, akár több repozitóriumi gyűjteménnyel is

Az egyesület évente számol el tagjaival a tárgyévét megelőző évben regisztrált DOI-k száma alapján és konzorciumok esetén a december 31-én aktuális tagok száma alapján. A díjszabás³ eltérő a non-profit és a for-profit szervezetek számára. A fizetendő díj három összetevőből áll: az egyesületi tagdíjból, ill. a szolgáltatási díjből, aminek két eleme az intézményi díj és a DOI-k száma után fizetendő díj. Az egyesületi tagdíj jelenleg 2000 EUR, konzorciumok esetén ezt a konzorcium tagjai közösen fizetik meg (az MTA KIK ezt a díjat megfizeti, azaz nem szedi be a konzorciumi tagoktól). Ezen felül minden csatlakozott tag 500 EUR intézményi díjat kell fizessen, valamint a DOI

³ A részletes díjszabást a lehetséges kedvezményekkel ld. <https://datacite.org/feemodel.html>

regisztráció díját, ami 1999 DOI regisztrációjáig 0,8 EUR, 2000 DOI regisztrációja felett pedig ez az összeg csökken (pl. 2500 DOI esetén 0,64 EUR/DOI, 9000 DOI esetén 0,18 EUR/DOI adódik).

Annual DOI range	Tier	Organization Fee	DOI Fee	Annual Service Fee
0 - 1999	Tier 1	500€	0.80€ per DOI	500€ + 0.80€ per DOI
2.000 - 10.000	Tier 2	500€	1.600€	2.100€
10.001 - 100.000	Tier 3	500€	2.500€	3.000€
100.001 - 250.000	Tier 4	500€	3.500€	4.000€
250.001 - 1.000.000	Tier 5	500€	8.500€	9.000€
1.000.001 - 2.000.000	Tier 6	500€	13.000€	13.500€
2.000.001 - 5.000.000	Tier 7	500€	18.500€	19.000€
5.000.001 - 10.000.000	Tier 8	500€	25.500€	26.000€

3. kép: A DataCite aktuális díjszabása

Amennyiben egy konzorciumi tag 10.000 DOI-nál többet regisztrál egy évben, taggá kell válnia (Direct Member) vagy dönthet a konzorciumi tagság mellett, de meg kell fizetnie az évi 2000 EUR tagsági díjat is.

A DataCite saját metaadatsémát⁴ alakította ki, amit folyamatosan gondoz a Metadata Working Group koordinálásával. Az egyesület igen aktív kapcsolatban áll tagjaival, havi webinárokat szervez és fórumot tart fenn, ahol a résztvevők megoszthatják egymással tapasztalataikat és a jó gyakorlatokat. Az MTA KIK DOI Irodájának munkatársai is részt vesznek ezeken a konzultációkon, gyakorta tapasztaljuk, hogy az itt elhangzottakat figyelembe veszik a séma módosításakor. Az adatbeviteli űrlap kifejezetten a kutatási adatok és szürkeirodalom bibliográfiai leírására felel meg, számos adatbázis-szolgáltató használja világszerte.

A séma háromféle metaadatot tartalmaz: a kötelező, az ajánlott és az opcionálisan megadható adatok körét. Ennek alapján a metaadatséma dokumentációjában a hivatkozás formátumára is javaslatot tesz a DataCite: Creator (PublicationYear): Title. Version. Publisher. (resourceTypeGeneral). Identifier⁵

⁴ DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. <https://doi.org/10.14454/3w3z-sa82>
További információk, a WG bemutatása és módosítás-javaslati űrlap: <https://schema.datacite.org/>

⁵ DataCite Metadata Schema. Version 4.4. 10. old.

Kötelező adatok:

- DOI azonosító (Identifier),
- alkotó (Creator),
- kiadó (Publisher),
- megjelenés éve (PublicationYear),
- forrás típusa (ResourceType - az azonosított objektum típusa)

Azaz így néz ki egy a DOI regisztrációhoz összeállított XML fájl, ami csak a minimális metaadatokat tartalmazza:

```
<identifier identifierType="DOI">10.18426/OBM.kd8ouii28u00</identifier>
<creators>
  <creator>
    <creatorName>Bán, Miklós</creatorName>
  </creator>
</creators>
<titles>
  <title>Dead Animals occurrence database</title>
</titles>
<publisher>OpenBioMaps Consortium</publisher>
<publicationYear>2014</publicationYear>
<resourceType resourceTypeGeneral="Service"/>
</resource>
```

Igen hasznos, ha ennél azért több adatot veszünk fel a DOI regisztráció során. A javasolt (R), azaz az adatszerét, használatot jelentősen támogató adatokat és az opcionálisan (O) megadható adatokat az alábbi táblázat tartalmazza:

Table 2: DataCite Recommended and Optional Properties

ID	Property	Obligation
6	Subject (with scheme sub-property)	R
7	Contributor (with optional given name, family name, name identifier, and affiliation sub-properties)	R
8	Date (with type sub-property)	R
9	Language	O
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point, box, place, and polygon sub-properties)	R
19	FundingReference (with name, identifier, and award related sub-properties)	O
20	RelatedItem (with identifier, creator, title, publication year, volume, issue, number, page, publisher, edition, and contributor sub-properties)	O

4. kép: A DataCite metaadatséma ajánlott és opcionális adatelemei

Íme egy adatbázis-lekérdezést azonosító DOI regisztrációs XML-e,⁶ melyet az MTA KIK DOI Iroda munkatársai az ügyféllel szoros együttműködésben állítottak össze:

```
<identifier identifierType="DOI">10.18426/obm.36vn3g36r3m0</identifier>
<creators>
  <creator>
    <creatorName>OBMQuery dinpi.openbiomaps.org</creatorName>
  </creator>
</creators>
<titles>
  <title>Database query for Violet click beetle from Duna-Ipoly National Park
species occurrence database (doi: 10.18426/obm.23mj4em6bca0)</title>
</titles>
<publisher>Duna-Ipoly National Park</publisher>
<publicationYear>2017</publicationYear>
<subjects>
  <subject xml:lang="eng">OpenBioMaps</subject>
  <subject xml:lang="eng">species occurrences</subject>
  <subject xml:lang="eng">Limoniscus violaceus</subject>
  <subject xml:lang="eng">Violet click beetle</subject>
  <subject xml:lang="hun">kék pattanó</subject>
</subjects>
<contributors>
<contributor contributorType="DataCollector">
<contributorName>Bérces, Sándor</contributorName>
</contributor>
<contributor contributorType="DataCollector">
<contributorName>Grabant, Aranka</contributorName>
</contributor>
<contributor contributorType="DataCollector">
<contributorName>Kotán, Attila</contributorName>
</contributor>
<contributor contributorType="DataCollector">
<contributorName>Merkl, Ottó</contributorName>
</contributor>
<contributor contributorType="DataCollector">
<contributorName>Nádai, László</contributorName>
</contributor>
</contributors>
<dates>
  <date dateType="Created">2017-12-07</date>
  <date dateType="Collected">From 2008-01-29 to 2017-10-21</date>
</dates>
<resourceType resourceTypeGeneral="Dataset"/>
<alternateIdentifiers>
  <alternateIdentifier alternateIdentifierType="URL">
http://dinpi.openbiomaps.org/projects/dinpi/IQ/51@i79JYVfCjK8r6DZq/</alternateIdentifier>
  <alternateIdentifier alternateIdentifierType="URL">
Raw data: http://dinpi.openbiomaps.org/projects/dinpi/IQ/51@i79JYVfCjK8r6DZq/data/</alternateIdentifier>
  <alternateIdentifier alternateIdentifierType="URL">
Map: http://dinpi.openbiomaps.org/projects/dinpi/IQ/51@i79JYVfCjK8r6DZq/map/</alternateIdentifier>
</alternateIdentifiers>
<relatedIdentifiers>
  <relatedIdentifier relatedIdentifierType="DOI" relationType="IsCompiledBy">
10.18426/obm.23mj4em6bca0</relatedIdentifier>
</relatedIdentifiers>
<sizes>
  <size>212567 bytes</size>
</sizes>
<formats>
  <format>text/json</format>
</formats>
<rightsList>
  <rights rightsURI="http://creativecommons.org/licenses/by-nc/4.0">The data included in this query
are provided to the user under a Creative Commons BY-NC 4.0 license which means that you are free
to use, share, and adapt the data provided that you give reasonable and appropriate credit (attribution)
and that you do not use the material for commercial purposes (non-commercial)</rights>
</rightsList>
<descriptions>
  <description descriptionType="Abstract" xml:lang="eng">A dataset containing 1 species occurrences
available in dinpi.openbiomaps.org matching the query: obm_taxon_id: 4033.
The dataset includes 208 records.</description>
</descriptions>
</resource>
```

⁶ A teljes XML letölthető (Download Metadata -> DataCite XML, ill. számos további formátum):
<https://commons.datacite.org/doi.org/10.18426/obm.36vn3g36r3m0>

A regisztráció kis volumenben végezhető a DataCite Fabrica felületén keresztül az ott elérhető űrlap kitöltésével vagy előre megírt XML fájl feltöltésével (ez utóbbi esetben van lehetőség adatgazdagításra, a webes felületen elérhető űrlap nem tartalmazza a metaadatsémában szereplő minden elemet). A DOI-khoz kapcsolt webcím és metaadatok karbantartása is elvégezhető ezen a felületen.⁷ A regisztráció természetesen elvégezhető a DataCite rendszere és a tagok adatgyűjteményei között beállított API⁸ kapcsolattal is. Mindazonáltal javasoljuk, hogy a regisztrációt előzze meg a metaadatok validálása, mert a DOI rendszerére épített szolgáltatások akkor lesznek valóban hatékonyan használhatóak, ha a metaadatok az arra kijelölt mezőben és hibátlanul kerülnek az adatbázisba.

Ugyancsak a használatot segíti elő az egyesület tagjai által megfogalmazott alapvető követelmények betartása, hiszen a DataCite úgy fejleszti szolgáltatásait, hogy azok a legmegfelelőbbek legyenek az azonosított kutatási produktumok számára és az egyéb adatbázisok üzemeltetői is ennek alapján döntenek a DataCite adatbázisának felhasználásáról. Tehát érdemes inkább a CrossRef-hez csatlakozni, ha például saját kiadásában megjelenő folyóiratban vagy kötetben közreadott tudományos publikációkat kíván egy kutatóintézet DOI-val azonosítani, hiszen a DataCite adatbázisát kutatási adatokkal foglalkozó más adatbázisok használják – például a Magyar Tudományos Művek Tára csak a CrossRef adatbázisából veszi át publikációk adatait, ugyanakkor a kutatási adatok adatainak átvétele kifejezetten a Kutatási adat típusba tervezett (még csak fejlesztési terv). Az adatminőség romlásához és számos félreértéshez vezet, ha nem a meghatározott forrástípusokat azonosítjuk a DataCite-nál DOI-val vagy nem a megfelelő metaadat mezőkbe kerülnek a leíró adatok. Így az adatcserére épülő elemzések is megbízhatatlanná válnak. Természetesen figyelembe kell venni azt is, hogy a kutatók számára is jól látható helyen elérhető legyen minden a hivatkozáshoz szükséges információ, hiszen hiába készítjük fel a szoftvereket az automatizált feldolgozásra, ha a bemeneti oldalon a felhasználók nem használják a DOI-kat.

A hatékony működéshez a következőkre van szükség:⁹

- Válasszuk az azonosítani kívánt tartalomnak megfelelő szolgáltatást nyújtó DOI ügynökséget a regisztrációhoz.
- Csak a szervezet saját közreadásában megjelenő objektumoknak regisztrálható DOI. Egyrészt a szervezet csak ezek működésének fenntartásáért vállalhat felelősséget. A szükség szerinti frissítéseket a DataCite el is várja. Így az is elkerülhető, hogy ugyanaz a forrás több DOI-val

⁷ Bár a DOI használatának alapjai nem témája jelen tanulmánynak, hangsúlyozzuk, hogy a DOI-khoz kapcsolt URL-ek szükség szerinti cseréje a mindenkori online elérés biztosításának kulcsa. A megtalálhatóságot és a használatot alapvetően befolyásoló tényező a metaadatok helyessége, azaz esetlegesen felmerülő hiba esetén ezek javítása, ill. lehetőségek szerinti gazdagítása is elengedhetetlen. Tehát a DOI folyamatos, szükség szerinti karbantartást igényel, ennek megvalósításáról pedig a DataCite-hoz csatlakozott kutatóintézetnek gondoskodnia kell.

⁸ DataCite REST API Guide: <https://support.datacite.org/docs/apis> ; Knowledge base: <https://support.datacite.org/docs/api>

⁹ Id. DataCite DOI Basics: <https://support.datacite.org/docs/doi-basics>
DataCite DOI Registration Policy: <https://support.datacite.org/docs/doi-registration-policy>
DataCite DOI Display Guidelines: <https://support.datacite.org/docs/datacite-doi-display-guidelines>
Best Practices for DOI Landing Pages: <https://support.datacite.org/docs/landing-pages>
Best Practices for DOI Registration: <https://support.datacite.org/docs/best-practices-for-datacite-members>
Best Practices for Tombstone Pages: <https://support.datacite.org/docs/tombstone-pages>

rendelkezzen. Tehát egy repozitóriumban elhelyezett kézirat például kaphat DOI azonosítót, természetesen a hozzá kapcsolt bibliográfiai leírást a kéziratról kell elkészíteni, mint a repozitóriumban elhelyezett tanulmányról. Ha később ez a kézirat valahol megjelenik, a kiadó elláthatja az új verziót másik DOI-val (természetesen nem használhatja a repozitórium által regisztrált DOI-t, de a megjelentetett műben – és a DOI regisztráció során a metaadatok között is – hivatkozhat a DOI-val annak korábbi közreadására). Így kiválóan elkülöníthető a kéziratra érkezett idézetek és a megjelent publikációra érkezett idézetek köre. Tehát adott tartalom különböző verziói azonosíthatóak DOI-val, az azonos verziók azonban nem rendelkezhetnek több DOI azonosítóval akkor sem, ha több adatbázisban is szerepel ugyanaz a verzió.

- Javasolt, hogy a DOI utótag egy random karaktersorozat legyen – a DataCite Fabrica tud suffixet generálni –, ugyanis ha jelentést hordoz, az félreértésekre adhat okot, hiszen a jelentés idővel változhat (pl. egy intézmény nevének rövidítése annak átnevezése esetén) vagy a vélt séma alapján a humán felhasználók megpróbálhatnak további DOI-kat összeállítani és használni (pl. az utótagban a v2 átírása v1-re nem feltétlenül vezet el egy tartalom korábbi verziójához). Figyelemmel kell lenni arra is, hogy maga a DOI azonosító nem módosítható vagy törölhető, csak a hozzá kapcsolt adatok módosítása lehetséges.¹⁰ A suffix ideális hossza 6-10 karakter, a megengedett karakterek köre pedig: a-z, 0-9 és a - (kötőjel), mivel más karakterek egyéb jelentést hordoznak az azonosítóban. Az azonosító nem kis- és nagybetű érzékeny.
- A DOI-knak egy nyilvánosan elérhető céloldalra kell vezetniük, ahol elérhető az azonosított objektum bibliográfiai leírása és a letöltési lehetőség. A teljes tartalomhoz való nyílt hozzáférés nem alapkövetelmény, de javasolt a FAIR elveknek megfelelő közlés.¹¹ Ezen az oldalon a DOI-t linkként kell feltüntetni, hogy a szoftverek is könnyedén felismerjék és használják és ne csak szövegnek tekintsék (pl. nagyobb valószínűséggel kerüljenek be a Google Dataset Search-be¹²). Az idézés elősegítése érdekében célszerű a kutatóknak különböző formátumokban elérhetően kínálni ezen az oldalon az objektum bibliográfiai leírását a DOI-val együtt.
- A DOI-val azonosított források hosszútávú elérhetőségét biztosítani kell, azaz ha nem repozitóriumban elhelyezett objektumokhoz regisztrált a DOI, akkor azok archiválásáról tartós tárhelyen és hosszútávú megőrzésre és szolgáltatásra alkalmas formátumban gondoskodni kell.¹³ Továbbá érdemes arra is felkészülni, ha ez az archivált példány mégsem

¹⁰ Meg kell jegyezzük, hogy a DataCite szolgáltatása egyedülálló abban a tekintetben, hogy a Fabrica rendszerébe háromféle státuszban vihető fel adat: a Draft státuszú rekordok törölhetőek, mivel nem kerülnek beküldésre a központi DOI rendszerbe. Tulajdonképpen a kutatóintézet saját munkaterületén álló azonosítókról van szó, URL-t sem kell kapcsolni a felvitelük során, pusztán az azonosító lefoglalására alkalmas ez a státusz. A Registered státuszú azonosító már bekerül a központi rendszerbe, azaz működik, kell hozzá URL és nem törölhető, de a DataCite Commons adatbázis nem indexeli. A Findable státuszú azonosítók azok, amik az indexelőkben is megjelennek. A központi rendszerbe bekerült azonosítók Draft státuszba sem kerülhetnek vissza.

A részletes tájékoztatót ld. <https://support.datacite.org/docs/doi-states>

¹¹ Guides for Researchers by OpenAIRE. How to make your data FAIR : Basic information with links to resources: <https://www.openaire.eu/how-to-make-your-data-fair>

¹² <https://support.datacite.org/docs/how-do-i-expose-my-datasets-to-google-dataset-search>

¹³ A DataCite kutatásiadat repozitórium adatbázist is működtet, a re3data.org szolgáltatást: <https://support.datacite.org/docs/datacite-re3data-user-documentation>

fenntartható. Ilyen esetben a DOI-k ún. sárga weboldalra irányíthatóak át – ezeket persze szintén a regisztrációt végző szervezetnek kell létrehozni és fenntartani. Ezeket a sárga weboldalakon mindazoknak a bibliográfiai információknak szerepelnie kell, amik az eredeti céloldalon szerepeltek, valamint tájékoztatást kell adnia a tartalom elérhetetlenségéről.

- Javasolt a kutatási produktumok letisztított, végleges verziójának DOI-val való azonosítása, de egyes esetekben lehet létjogosultsága a verziókezelésnek.¹⁴ Amennyiben csak kisebb módosítások – például helyesbítések, kisebb kiegészítések a könnyebb használhatóság érdekében – történnek, nem szükséges új DOI-val ellátni az új verziót, de a DOI metaadatait frissíteni kell és megadni az új verziószámot. Nagyobb változtatások esetében, ill. például szoftverek új példányainak kiadásakor, olyan kutatási eredmények közreadása során, melyek periodikusan frissítve kerülnek publikálásra van lehetőség arra is, hogy az egyes verziók DOI azonosítóit összekapcsoljuk és a köztük lévő viszony is jelölhető (pl. `IsPreviousVersionOf`, `IsNewVersionOf`, `IsVersionOf`, `HasVersion`). Ha egy forrásnak sok verziója létezik, rendelhető hozzá egy ún. *canonical DOI*, ami az összes verziót képviseli. Ehhez kapcsolni kell az egyes verziók azonosítóit. A DataCite a verziókezelés terén a döntést a kutatóintézetekre bízta, de együttműködik az RDA Data Versioning munkacsoportjával. Ennek tájékoztató dokumentumait¹⁵ ajánljuk az adatkezelők figyelmébe.

A fenti technikák segítségével a kutatási eredmények a lehető leggyorsabban megtalálhatóvá, felhasználhatóvá és idézhetővé válnak, hiszen a DOI-hoz kapcsolt metaadatok a regisztrációt követően szinte azonnal elérhetőek a tudományos közösség számára a nyilvános kereső, a [DataCite Commons](#) segítségével. A metaadatok a központi adatbázisból bárki által lekérdezhetőek, azaz könnyedén és gyorsan bekerülnek közös keresőszolgáltatásokba. A DOI regisztrációról és az azonosítók használatáról pedig statisztika is elérhető.¹⁶

Az egyesület a DOI regisztrációs szolgáltatás mellett ismeretterjesztő tevékenységet folytat, ill. szolgáltatásokat fejleszt, hogy elősegítse a kutatási eredmények megosztását és idézését. A DataCite szolgáltatásai mellett, hogy támogatják az adatok idézését, a felhasználás elemzését is lehetővé teszik. Azaz a DOI-k segítenek a kutatási projektek eredményeinek követésében, az idézettség összegyűjtésében, már a publikációk megjelenését megelőzően.

Lehetőség van arra, hogy a DOI regisztráció során rögzítsük a DataCite adatbázisában egy adott forrás felhasznált irodalom listájában szereplő más források vagy egyéb kapcsolódó források DOI azonosítóit.¹⁷

A DataCite ingyenes, bárki számára használható konvertáló eszközt biztosít a különböző idézet stílusok előállításához (Citation Formatter).¹⁸ A Make Data Count projekt részeként az egyesület lehetővé tette, hogy a repozitóriumok betölthessék a DataCite Commons-ba az azonosított

¹⁴ A verziókezelésről részletesen ld. <https://support.datacite.org/docs/versioning>

¹⁵ <https://www.rd-alliance.org/groups/data-versioning-ig>

¹⁶ DataCite Statistics: <https://support.datacite.org/docs/datacite-statistics>

¹⁷ Contributing Citations and References: <https://support.datacite.org/docs/contributing-citations-and-references>

¹⁸ DataCite Citation Formatter: <https://support.datacite.org/docs/datacite-citation-formatter>

objektumok használati, azaz megtekintési és letöltési adatait.¹⁹ Arra is van lehetőség, hogy a DataCite rendszereiből összegyűjthető használati és idézettségi adatokat a Data Metrics Badge segítségével a repozitóriumok megjelenítsék a források weblapjain.²⁰

5.) MTA KIK DOI IRODA

A Magyar Tudományos Akadémia Könyvtár és Információs Központ (MTA KIK) 2014 óta segíti a hazai tudományos publikációk és kutatási adatok DOI-val való azonosítását.

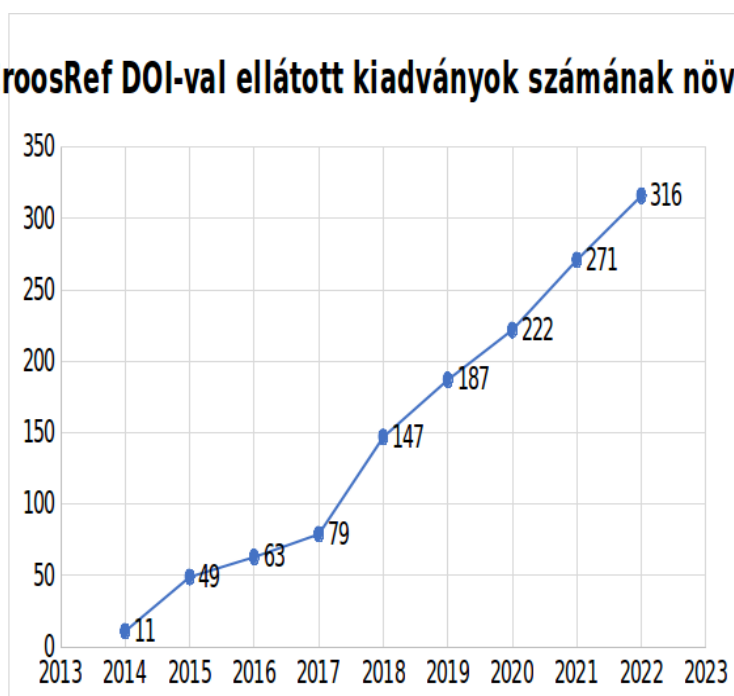
A DOI Iroda megszervezését a felsőoktatási törvény 2013. január 1-től hatályos módosítása hívta életre, mely szerint „*A doktori értekezést és téziseit elektronikus formában a Magyar Tudományos Művek Tárában, az általánosan elfogadott nemzetközi gyakorlatnak megfelelő (DOI) azonosítóval ellátva, mindenki számára hozzáférhetővé kell tenni.*”²¹ Ezért áttekintettük a regisztrációs ügynökségek szolgáltatásait – melyek más-más dokumentumtípusra, kiadói körre specializáltak – és a DataCite portfólióját ítéltük legmegfelelőbbnek a PhD dolgozatok számára (ekkor a CrossRef metaadatsémájában a tézis még nem szerepelt). Bár az ügynökség elsősorban a kutatási adatok azonosítására optimalizálja működését, de az ún. szürke irodalom hivatkozhatóságát is képes biztosítani, így ehhez az egyesülethez csatlakozott elsőként a könyvtár. Még ugyanebben az évben aláírtuk a tagsági szerződést a CrossRef egyesülettel is, hogy megfelelő szolgáltatást nyújthassunk a tudományos folyóiratok és könyvek publikálásához is. A DOI Iroda három munkatársa ma már több mint 300 partner munkáját segíti. A közös munka együttműködés keretében valósul meg, azaz kollégáink a regisztrációval kapcsolatos feladatokat látják el, partnereink pedig kiadványaik adataival gyarapítják a Magyar Tudományos Művek Tára (MTMT) adatbázis tartalmát, valamint a publikációk PDF példányával a REAL repozitórium gyűjteményét, ami garantálja számukra a közlemények elektronikus változatának hosszútávú online megőrzését.

¹⁹ Usage Stats in DataCite Commons: <https://support.datacite.org/docs/usage-stats>

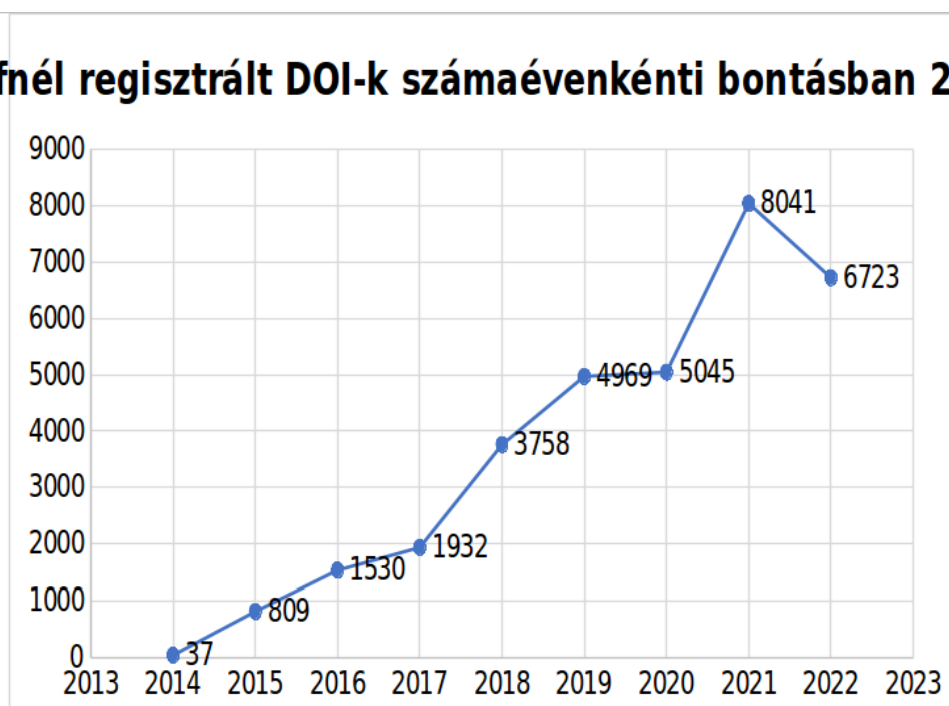
²⁰ Displaying Usage and Citations in your Repository: <https://support.datacite.org/docs/displaying-usage-and-citations-in-your-repository>

²¹ [2011. évi CCIV. törvény a nemzeti felsőoktatásról 53/A. § 2.](#)

Az MTA KIK által CrossRef DOI-val ellátott kiadványok számának növekedése 2014-2022



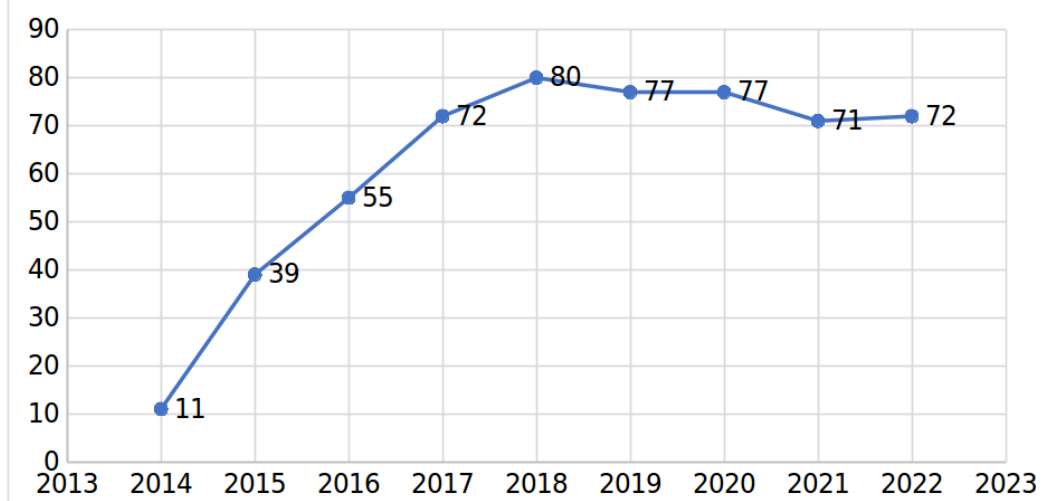
CrossRef-nél regisztrált DOI-k számaévenkénti bontásban 2014-2022



A CrossRef-nél elérhető regisztráció évről évre népszerűbb: jelenleg 162 kiadó összesen 316 kiadványának DOI-val való azonosítását segítjük.

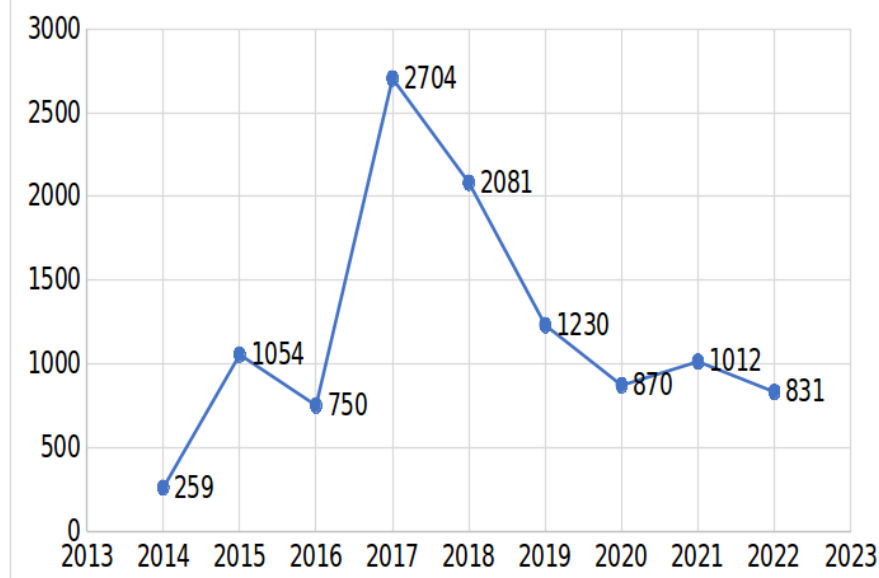
Ennek az ügynökségnek a szolgáltatása mellett, hogy a DOI alapvető funkcióit biztosítja – egyedi, permanens azonosító, ami garantálja a fulltext online elérését –, a köré szerveződő munkamenetnek köszönhetően támogatja a nemzetközi, automatizált adatcserét, ezáltal megkönnyíti a hivatkozások gyűjtését is. Ezt pedig az teszi lehetővé, hogy a csatlakozott kiadók a publikációk irodalomjegyzékében kölcsönösen feltüntetik a DOI-val ellátott források azonosítóit.

DataCite partnerek számának alakulása 2014-2022



A DataCite-nál elérhető szolgáltatást jelenleg 72 partnerünk számára biztosítjuk. Ügyfeink számának csökkenését több ok is indokolja: 2020-ig az MTA KIK önerőből finanszírozta az egyesületi tagság díját, így a DOI regisztráció partnereink számára ingyenes volt. 2020-ban azonban az egyesület módosította díjszabását, így az új csatlakozások száma megtorpant, mert a legalacsonyabb éves díj elérte az 500 EUR összeget. Emellett a tudományos kiadványokat megjelentető ügyfeleinknek a CrossRef szolgáltatását ajánljuk és örömmel tapasztaljuk, hogy egyre többen vállalják fel az ügynökség által előírt technikai feltételek teljesítését annak érdekében, hogy a publikációk adatai bekerülhessenek a nemzetközi, gépi adatcserébe.

DataCite-nál regisztrált DOI-k számaévenkénti bontásban 2014-2022



Tájékoztató weboldalt is üzemeltetünk 2015 óta, mely az openaccess.mtak.hu címen érhető el. A könyvtár ezen a honlapon nyújt információkat a fenti szolgáltatásokról, nyílt hozzáférést támogató projektjeiről és ilyen témájú rendezvényeiről is.

2016 óta a folyóirat-publikálást támogató Open Journal Systems szoftvert is kínálunk együttműködő partnereinknek, ami a DOI regisztráció automatizálásának és a hivatkozások lekérdezésének lehetőségével segíti munkánkat.

Terveink közt szerepel a REAL repozitóriumokat üzemeltető Eprints szoftver DataCite DOI regisztrációt támogató kiegészítőjének használatba vétele, valamint a DOI használatára és a REAL repozitórium gyűjteményére alapozva hivatkozáskereső szolgáltatás kialakítása, mely az MTMT adatbázis tartalmának gyarapítását segíti majd.

6.) DATACITE DOI REGISZTRÁCIÓ TAPASZTALATOK

Az MTA KIK mintegy 8 éves gyakorlatában mindössze egy ügyfél élt a kötelező metaadatoknál több adat megadásának lehetőségével a DOI regisztráció során: az OpenBioMaps konzorcium. Eddig 8 adatbázist és 4 adatbázis lekérdezést azonosítottak DOI-val (ld. <https://commons.datacite.org/repositories/4ygcw7>).

Az első adatbázis DOI-át 2016-ban regisztráltuk, ez volt a Duna halas adatbázis (<https://commons.datacite.org/doi.org/10.18426/obm.28kj80ftej0>). A regisztrált adatok XML vagy JSON formátumban letölthetők a DataCite nyilvános keresőjéből. A kötelező adatok mellett egy rövid leírást, angol és magyar nyelvű kulcsszavakat, felhasználási jogi információkat, nyelv, méret adatokat tartalmaz a regisztráció, ill. az alkotók szerepe is megadott, vezeték- és keresztnévük is elkülönített már, ORCID-al azonosítottak, intézményi affiliációval ellátottak.

Az első lekérdezést 2017-ben azonosították DOI-val, ez volt az OpenBioMaps Database Query címmel ellátott objektum (<https://commons.datacite.org/doi.org/10.18426/obm.56j5rb3h9lc0>). A korábbi adatkör azzal bővült ki, hogy megjelöltük, hogy melyik adatbázisból történt a lekérdezés. Később informatívabb címeket adtak a lekérdezéseknek, mint pl. Database query for Violet click beetle from Duna-Ipoly National Park species occurrence database (doi: [10.18426/obm.23mj4em6bca0](https://commons.datacite.org/doi.org/10.18426/obm.23mj4em6bca0)) - az adatokat ld. a <https://commons.datacite.org/doi.org/10.18426/obm.36vn3g36r3m0> oldalon. Mivel az adatbázis folyamatosan bővül, gondoskodtak a lekérdezés DOI-val azonosított állapotának elmentéséről is (a nyers adatok ma is elérhetőek, bár sajnos a térképes és táblázatos formátum már nem), valamint az adatok gyűjtésének ideje és a lekérdezés dátuma is bekerült a DOI-hoz kapcsolt adatok közé.

Sajnos ebben a rendkívül kis mintában is azt tapasztaljuk, hogy a megadott adatok köre ingadozó, például a szerzői affiliáció és az ORCID sem minden esetben megadott adat. Sokat kell még azon is dolgozni, hogy az objektumokat leíró metaadatok a publikálás helyén is fel legyenek tüntetve, hacsak lehet hivatkozási minta is szerepeljen az objektumoknál és számos alkalommal egyeztetünk a bővülő

adatbázisok, dokumentumok, módosuló objektumok DOI azonosításáról is, de az OpenBioMaps-n kívül más ügyfelünk nem alakította ki ennek technikai feltételeit.

Az a tény, hogy a DataCite központi adatbázisa csak discovery level – többé-kevésbé DublinCore – metaadatokat tartalmaz, ennél többet csak opcionálisan (ezért esetlegesen), rontja a központi metaadattár felhasználhatóságát a keresés terén. (Természetesen ez a probléma jelentkezik a lokális metaadattárat alkalmazó azonosítóknál is). Viszont a DC adatok megléte a hivatkozás-gyűjtési funkciót támogatja.

Több ügyfelünk is érdeklődött a regisztráció automatizálása iránt, de miután megkapták az API dokumentációt (<https://support.datacite.org/docs/api>), az ötletet sajnos nem követte megvalósítás.

A szolgáltatás igen népszerű volt, amíg ingyenesen lehetett hozzájutni. Jelenleg a legalacsonyabb ár az évi 2000 EUR összegű egyesületi tagdíj, az 500 EUR intézményi díj és a 0,8 EUR / DOI díj megfizetése. Bár az MTA KIK az egyesületi tagdíjat átvállalja, a díjszabás megváltozása óta sajnos alig csatlakozott új ügyfél. Eddig összesen 10.198 DOI-t regisztrált az MTA KIK DOI Irodája a DataCite ügynökségnél, ebből mindössze 1824 azonosít kutatási adatot (nyers adatokat tartalmazó fájlokat, térképeket, képeket, videókat, kutatási projekteket tartalmazó kollektciókat).

Ezeknek az objektumoknak a hosszútávú megőrzése néhány kivételtől eltekintve nem biztosított. Arra is megoldást kell találni, hogy az idő múlásával sok esetben megfelelnek az ügyfelek arról, hogy a DOI-kat nem elég feltüntetni az objektumokon, de regisztrálni is kell őket a központi adatbázisban. Nem tudjuk, hogy hány inaktív DOI szerepel az interneten, melyek regisztrációját ügyfeleink nem rendelték meg.

7.) ARK ÉS DOI

Az ARK és a DOI azonosítók használatának logikáját jól leírja a California Digital Library EZID tájékoztatója (<https://ezid.cdlib.org/learn/#05>):

Can DOIs and ARKs be used together?

ARK identifiers have certain features that can be very useful for:

- Keeping track of many granules of a dataset (ARKs will be able to "pass through" a suffix, so many thousands of items can be referenced on the basis of a single registration);

- Keeping track of data before a decision has been made about whether or not it is going to be retained (ARKs can be deleted).

This can make it attractive to use ARKs during the early part of a dataset's "life" or the early stages of the research process. Then, when the time comes to begin writing up results and it becomes clear which object(s) will be cited, it may appropriate to get DOIs for those objects.

DOIs have the citation-level "reputation" and it is possible to use the ARK of the cited object as the "suffix" for the DOI so that there is a traceable connection between the two. Here is what that means:

Step 1. You assign an ARK to a resource for good management and tracking: ark:/99999/fk4sf2w65j

Step 2. You decide to cite the resource, so you want a DOI.

Step 3. Using either the Advanced Create UI or the API, request a doi with this form:

doi:10.5072/FK2fk4sf2w65j

In this way, the two identifiers have a relationship, so the object can be tracked throughout its life cycle. With EZID, clients get access to both of these identifiers and can take best advantage of both approaches.

8.) DOI ÉS ARK HASZNÁLAT INTEGRÁLÁSA, HIERARCHIKUS AZONOSÍTÁS, VERZIÓK AZONOSÍTÁSA

A verziókezelés megvalósítását tárgyalja az ASAPbio preprintekkel foglalkozó tanulmányának (Beck et al., 2020) "Approaches to versioning" fejezete. A mi javaslatunk szerint a verziójelölést emberi fogyasztásra és az adatrepozitórium belső logikája számára az azonosító végére ponttal elválasztva, "v" jelzővel célszerű megvalósítani (pl. xyz.v1 ; xyz.v2 az első és második változat jelölésére az alapváltozaton túl).

Mind a verziókezelést, mind az esetleges logikai hierarchiát (adathalmaz / részhalmaza) az ARK qualifier elemében lehet megvalósítani.

alap :

ark:/99999/fk4sf2w65j/

részalmazai :

ark:/99999/fk4sf2w65j/abc

ark:/99999/fk4sf2w65j/def

ark:/99999/fk4sf2w65j/def/ghi

verziók :

ark:/99999/fk4sf2w65j.v1

ark:/99999/fk4sf2w65j/def/ghi.v1

bármelyikből lehet DOI azonosítót képezni:

doi:10.78912/fk4sf2w65j/def/ghi.v1

(A DOI suffixben megengedett a "/" és a "." karakter. Megjegyzendő, hogy az ARK-ban a "/" karakternek speciális szerepe a qualifier rész elválasztása.)

A fenti javaslat összhangban van a korábban idézett EZID gyakorlattal.

9.) ADATOK HIVATKOZÁSA

A kérdést tárgyalják a FORCE11 "Joint Declaration of Data Citation Principles", valamint a CODATA "Out of Cite, Out of Mind" tanulmányai.

Lényeges megállapítások, hogy bár az egyedi azonosítóknak (PID) kulcsszerepe van a kutatási adatok hivatkozásában, a hivatkozás azonban nem csak egyedi azonosító kérdés. A tanulmányok megállapítják azt is, hogy nem csupán a DOI alkalmazható egyedi azonosítóként adathivatkozásokban. Mindazonáltal a szerzők egyikének (H.A.) tapasztalata az, hogy néhány évtized óta rögződött, niche azonosítón túl (ilyen a csillagászatban a BIBCODE) szinte kizárólag DOI-kat használnak manapság kutatási adatok hivatkozására. Az általános használatú, népszerű, és az adatrepozitóriumi funkcióknak csupán egy részhalmazát megvalósító szolgáltatások, mint a figshare és a Zenodo szintén DOI-t használnak (automatikusan DOI-t regisztrálnak az elhelyezett tételeknek).

10.) TOVÁBBI EGYEDI AZONOSÍTÓK (KUTATÓK, SZERVEZETEK AZONOSÍTÁSA)

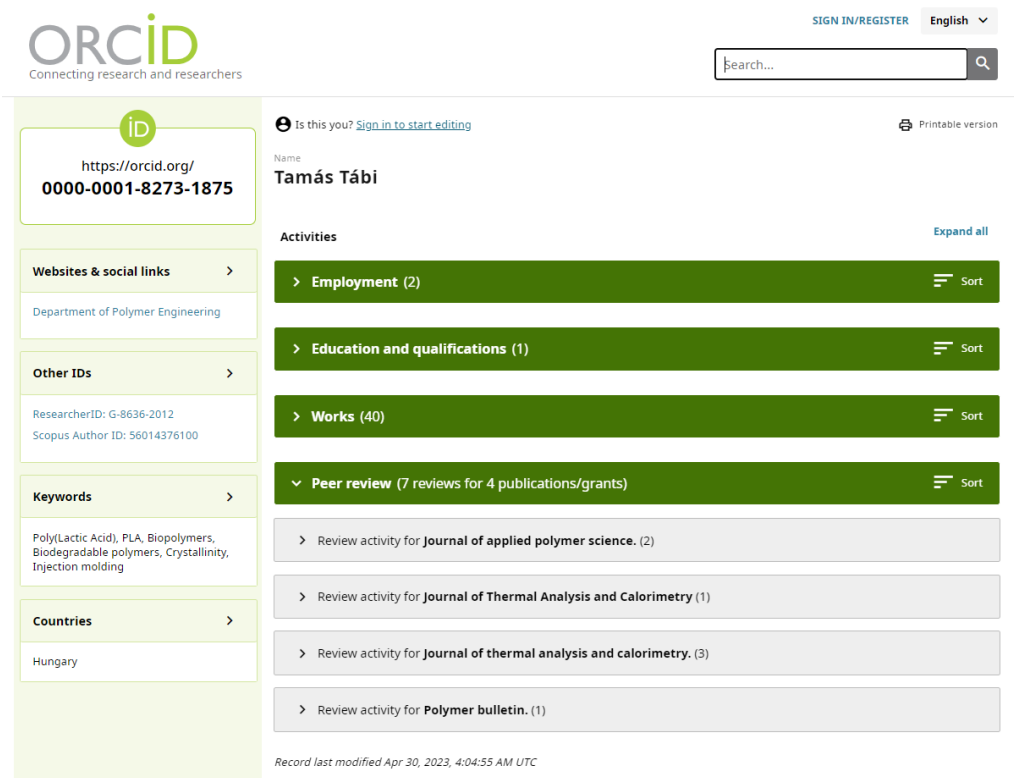
Egyedi, állandó azonosítók léteznek kutatók (szerzők) és (kutatással foglalkozó) intézmények azonosítására is. Ezek használata is ajánlott – adott esetben a kutatási adatok vagy publikációk egyedi azonosítóinak metaadataiban is –, hiszen igen nagy mértékben megkönnyítik az adatcsere során a számítógépes feldolgozást, ezáltal pedig jelentős munkaerő-, azaz költség-megtakarítás érhető el (ld.

a JISC 2021-ben végzett felmérésének²² eredményeit, melynek keretében a DOI, az ORCID, a ROR, a RAiD és a CrossRef Grants egyedi azonosítók használatának megterülését vizsgálták).

Ezeket az azonosítókat ezen tanulmányban nem ismertetjük részletesen.

ORCID

Az Open Researcher and Contributor ID a szerzők egyedi azonosítója (Holl és Bilicsi, 2017). Közösségi szervezésben működik, használata a szerzők számára ingyenes. Érdekes használni minden kutatással összefüggő tevékenység végzése során, legyen az publikálás, kutatási adat közzétevése, konferenciárészvétel, lektorálás, pályázat stb. Így a kutatók – amennyiben ebben az egy adatbázisban karbantartják profiljukat – könnyedén megoszthatják adataikat az általuk használt rendszerekkel, ezáltal minden kutatással összefüggő tevékenységük során azonosíthatóvá válnak, azok összegyűjtése gépesítetten, az esetleges hibák minimalizálásával és az azonos nevű szerzők elkülönítésével valósítható meg. A DataCite és a CrossRef DOI azonosítók metaadat-sémája is tartalmazza.²³ A DOI-k metaadataiban szereplő ORCID azonosítók teszik lehetővé, hogy a kutatók közleményei a gépi adatsere segítségével – azaz manuális beavatkozás nélkül – kerüljenek be a szerzők az ORCID adatbázisában szereplő publikációs-listájába (AutoUpdate ORCID profil bővítés).²⁴



The image shows a screenshot of an ORCID iD profile for Tamás Tábi. The profile is displayed on a public search page. The ORCID iD is https://orcid.org/0000-0001-8273-1875. The name is Tamás Tábi. The profile includes sections for Websites & social links (Department of Polymer Engineering), Other IDs (ResearcherID: G-8636-2012, Scopus Author ID: 56014376100), Keywords (Poly(Lactic Acid), PLA, Biopolymers, Biodegradable polymers, Crystallinity, Injection molding), and Countries (Hungary). The Activities section is expanded, showing Employment (2), Education and qualifications (1), Works (40), and Peer review (7 reviews for 4 publications/grants). The Peer review section is further expanded, showing review activity for Journal of applied polymer science (2), Journal of Thermal Analysis and Calorimetry (1), Journal of thermal analysis and calorimetry (3), and Polymer bulletin (1). The profile was last modified on Apr 30, 2023, at 4:04:55 AM UTC.

1. kép: ORCID adatlap a publikus keresőben

²² Brown, Josh, Jones, Phill, Meadows, Alice, Murphy, Fiona, & Clayton, Paul. (2021). UK PID Consortium: Cost-Benefit Analysis (1.0). Zenodo. <https://doi.org/10.5281/zenodo.4772627>

²³ <https://support.datacite.org/docs/can-i-add-orcid-ids-to-the-datacite-metadata-1>

²⁴ <https://www.crossref.org/community/orcid/>

ROR

A Research Organization Registry (ROR), a kutatást végző intézmények egyedi, állandó azonosítója.²⁵ Lehetővé teszi – többek között –, hogy a szerzőkhöz hasonlóan az intézményeket is kapcsolni lehessen a kutatási adatokhoz azok közreadása során úgy, hogy ez az információ is feldolgozható legyen szoftveresen. Azaz az intézmények kezelésének tekintetében biztosítja az egyértelmű azonosítást olyan adatbázisokban, amelyek rögzítik az intézményi hovatartozást. A rendszer a közösség szervezésében működik Creative Commons Public Domain (CC0) liszensz alatt, a szervezetek számára az azonosító regisztrációja és használata ingyenes. A DOI-hoz hasonlóan a ROR azonosítóhoz is kapcsolódik az intézmények adatait tartalmazó adatbázis, melyben keresni is lehet.²⁶

The screenshot shows the ROR Registry search results for the record <https://ror.org/04ws47v52>. The record is for the "Library and Information Centre of the Hungarian Academy of Sciences". The data is organized into several sections:

- ORGANIZATION TYPE:** Archive
- LOCATION:** Budapest (GeoNames ID 3054643), Hungary
- OTHER NAMES:** Magyar Tudományos Akadémia Könyvtár és Információs Központ
- OTHER IDENTIFIERS:** GRID grid.496758.1, ISNI 0000 0001 2178 9790
- WEBSITE:** http://konyvtar.mta.hu/index_en.php
- RELATIONSHIPS:** Parent Organization(s) Hungarian Academy of Sciences

At the bottom, there is a link to submit a curation request: [Submit a curation request](#).

2. kép: ROR adatlap a publikus keresőben

²⁵ <https://ror.org/> ; <https://www.crossref.org/community/ror/>

²⁶ <https://ror.org/search>

Beck, J., Ferguson, C. A., Funk, K. et al. (2020, July 21).

Building trust in preprints: recommendations for servers and other stakeholders.

<https://doi.org/10.31219/osf.io/8dn4w>

Bilicsi Erika (2015)

DOI regisztráció a Magyar Tudományos Akadémia Könyvtár és Információs Központban.

A szolgáltatás első évének tapasztalatai.

Networkshop 2015

<http://real.mtak.hu/44021/>

Bilicsi Erika (2017)

Egyedi, permanens azonosítók használata a tudományos folyóirat-kiadásban

Közlekedéstudományi szemle 67, 6. 6-8.

Bilicsi Erika (2017)

On-line megjelenő folyóiratcikkek azonosítása és elérésének biztosítása DOI segítségével.

Közlekedéstudományi szemle, 67. 2. 10-11.

COAR - ASAPBio WG (2022)

Ten Recommended Practices for Managing Preprints in Generalist and Institutional Repositories.

<https://www.coar-repositories.org/files/Preprint-Report-Formatted-Version-Dec-12-2022-1.pdf>

CODATA-ICSTI Task Group on Data Citation Standards and Practices (2013)

Out of cite, out of mind: the current state of practice, policy, and technology for the citation of data

Data Science Journal, Volume 12, 13 September 2013

Data Citation Synthesis Group (2014) Joint Declaration of Data Citation Principles.

Martone M. (ed.) San Diego CA: FORCE11;

<https://doi.org/10.25490/a97f-egy>

Holl András, Bilicsi Erika (2017)

ORCID – egy újabb szerző-azonosító tudományos közleményekhez.

KÖNYVTÁRI FIGYELŐ 63 : 3 pp. 346-350. , 5 p.

<http://real.mtak.hu/65517>

Karácsony Gyöngyi, Holl András (2011)

HU-DOI. Tanulmány az MTMT számára.

Simons, Natasha (2012)

Implementing DOIs for research data

D-lib Magazine, May/June 2012, v18, 5-6.

<https://doi.org/10.1045/may2012-simons>

Weibel Stuart, Jul Erik (1995)

PURLs to improve access to the Internet

OCLC Newsletter (November/December)

<https://library.oclc.org/digital/collection/p267701coll28/id/1839>

URL-EK:

MTA KIK OS tájékoztató oldal

<https://openaccess.mtak.hu/doi/>

MTMT dokumentum

<https://www.mtmt.hu/hirek/doi-azonosito-regisztraciojanak-lehetosege-0>

DPC Digital Preservation Handbook

<https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>

Persistent Identifiers - LibGuides tájékoztató oldal

https://transportation.libguides.com/persistent_identifiers

THOR Persistent Identifier Platform

<https://project-thor.readme.io/>

ARK

<https://arks.org/>

http://n2t.net/e/ark_ids.html

EZID

<https://ezid.cdlib.org/>

DOI Citation Formatter

<http://citation.crosscite.org>

Freya

<https://blog.datacite.org/tag/freya/>