

## Adattisztítás *Tartalomtípus* generálásának előkészítéséhez

Az RDA egyik látványos újítása, hogy az általános anyagmegjelölés helyett három szabályozott szótár segítségével jellemezhetjük a forrásokat. Ez a három szótár: a *Tartalomtípus*, a *Médiatípus* és a *Hordozótípus*. A rekordokban bizonyos formában eddig is rögzítettünk ilyen típusú információkat, ezeket tehát felhasználhatjuk a *Tartalomtípus* automatikus kitöltéséhez. Természetesen csak abban az esetben, ha az adatok eddigi rögzítése során pontosan és következetesen jártunk el. A továbbiakban azt nézzük végig, hogy milyen adatokból mit lehet generálni, valamint igyekszünk ahhoz is némi támpontot adni, hogy esetlegesen mit kell javítanunk a rekordjainkban, hogy a legenerált adatok hűen tükrözzék a forrás természetét.

Mivel a tényleges rögzítés esetleges szabványoktól eltérő volta erősen könyvtárfüggő, egy ilyen segédanyagnak nem lehet feladata az összes létező helyzetre kész megoldásokkal szolgálni, viszont célunk, hogy egyfajta sorvezetőt adjunk a könyvtárak kezébe a saját megfeleltetéseik, illetve az esetlegesen szükséges adattisztítások kidolgozásához.

A távmunkaajánlás csak az automatikus kitöltés előkészítésére vonatkozik, kérjük, még ne generálják le a kifejezéseket és ne kezdjék el aktívan használni őket!

### Tartalomtípus

A kommunikációs forma kategorizálása aszerint, hogy hogyan jut kifejezésre a tartalom és azt milyen módon érzékeljük. A tartalomtípus tükrözi a térbeli dimenziók számát és a mozgás jelenlétét vagy hiányát is, amennyiben a tartalmat kép vagy képek formájában fejezik ki.

A szótár a következő elemeket tartalmazza:

térképészeti adathalmaz	<i>Tartalomtípus</i> , ami olyan kartográfiai tartalomból áll, amely számítógépes feldolgozásra szánt, digitálisan kódolt adathalmazként jut kifejezésre.
kartográfiai kép	<i>Tartalomtípus</i> , ami olyan kartográfiai tartalomból áll, amelyet vonallal, alakkal, árnyékolással stb. fejeznek ki, amit állóképként vagy kétdimenziós képként való vizuális érzékelésre szántak.
kartográfiai mozgókép	<i>Tartalomtípus</i> , ami olyan kartográfiai tartalomból áll, amelyet mozgó, kétdimenziós képként történő érzékelésre szolgáló képekkel fejeznek ki.
tapintható térkép	<i>Tartalomtípus</i> , ami olyan kartográfiai tartalomból áll, amelyet tapintással, kétdimenziós állóképként történő érzékelésre szolgáló vonallal, alakkal és/vagy egyéb formával fejeznek ki.
háromdimenziós tapintható kartográfiai forma	<i>Tartalomtípus</i> , ami olyan kartográfiai tartalomból áll, amelyet tapintással, három dimenzióban történő érzékelésre szolgáló formával vagy formákkal fejeznek ki.
háromdimenziós kartográfiai forma	<i>Tartalomtípus</i> , ami olyan kartográfiai tartalomból áll, amelyet vizuálisan, három dimenzióban történő érzékelésre szolgáló formával vagy formákkal fejeznek ki.
számítógépes adathalmaz	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amely számítógépes feldolgozásra szánt, digitálisan kódolt adathalmazként jut kifejezésre.
számítógépes program	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet számítógépes feldolgozásra és végrehajtásra szánt, digitálisan kódolt utasításként fejeznek ki.

mozgásírás	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet vizuális érzékelésre szánt mozgáslejegyzés valamely formájával fejeznek ki.
zenei lejegyzés	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet vizuális érzékelésre szánt zenei lejegyzés valamely formájával fejeznek ki.
előadott zene	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet hallható formájú zenén keresztül fejeznek ki.
hangok	<i>Tartalomtípus</i> , ami olyan a nyelvtől és a zenétől különböző tartalomból áll, amelyet hallható formában fejeznek ki.
beszéd	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet hallható formájú nyelven keresztül fejeznek ki.
állókép	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet vonallal, alakzattal, árnyékolással stb. fejeznek ki, amit állóképként vagy kétdimenziós képként való vizuális érzékelésre szántak.
tapintható kép	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet vonallal, alakzattal és/vagy egyéb formákkal fejeznek ki, amelyeket arra szánnak, hogy tapintással kétdimenziós állóképként érzékeljék őket.
tapintható zenei lejegyzés	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet olyan zenei lejegyzési formával fejeznek ki, amelyet tapintásos érzékelésre szánnak.
tapintható mozgásírás	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet olyan mozgáslejegyzési formával fejeznek ki, amelyet tapintásos érzékelésre szánnak.
tapintható szöveg	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet olyan nyelvi lejegyzési formával fejeznek ki, amelyet tapintásos érzékelésre szánnak.
tapintható háromdimenziós forma	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet egy vagy több formával fejeznek ki, amelyeket arra szánnak, hogy tapintással háromdimenziós formaként vagy formákként érzékeljenek.
szöveg	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet egy nyelvi lejegyzési formával fejeznek ki, amelyet vizuális érzékelésre szánnak.
háromdimenziós forma	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet egy vagy több formával fejeznek ki, és ezeket a formákat arra szánják, hogy vizuálisan három dimenzióban megjelenőként érzékeljék őket.
háromdimenziós mozgókép	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet képekkel fejeznek ki, és ezeket a képeket arra szánják, hogy mozgásban lévőként és három dimenzióban megjelenőként érzékeljék őket.
kétdimenziós mozgókép	<i>Tartalomtípus</i> , ami olyan tartalomból áll, amelyet képekkel fejeznek ki, és ezeket a képeket arra szánják, hogy mozgásban lévőként és két dimenzióban megjelenőként érzékeljék őket.

1. táblázat

Elképzelhető, hogy a forrás egynél több kategóriába is besorolható. Ez esetben (a *leader*ben rögzítéssel ellentétben) lehetőségünk van a leírásban minden releváns tartalomtípust megadni.

A *Tartalomtípus* tehát ismételhető, viszont a *leader* 06 értékét felhasználva csak egyféle kifejezést tudunk generálni. A vegyes tartalmú források részletezését felmenő rendszerben, illetve manuális javítással (szükség esetén a forrás kézbevételel) lehet majd kihasználni.

Tartalomtípus-jellegű adatot eddig – minden Magyarországon használt MARC formátumban – a *leader* 06-os pozícióján rögzítettünk a következő formában:

MARC 21	MARC 21	HUNMARC	HUNMARC
nyelvi anyag	a	a	nyomtatott nyelvi anyag
		b	kézírt nyelvi anyag

lejegyzett zene	c	c	nyomtatott zenemű
kéziratos zene	d	d	kéziratos zenemű
kartográfiai anyag	e	e	nyomtatott térkép
kéziratos kartográfiai anyag	f	f	kéziratos térkép
kivetített médium	g	g	audiovizuális anyag
nem zenei hangfelvétel	i	i	nem zenei hangzó anyag
zenei hangfelvétel	j	j	zenei hangzó anyag
kétdimenziós nem kivetíthető grafika	k	k	kétdimenziós (nem kivetíthető) ábrázolások (korábban: grafika)
		l	számítógépes anyag
számítógép fájl	m		
		m	multimédia
		n	különleges oktató anyag
készlet	o	o	vegyes dokumentum csomag (korábban: készlet)
vegyes anyag	p	p	vegyes anyag
háromdimenziós alkotások vagy természeti tárgyak	r	r	háromdimenziós alkotások
kéziratos nyelvi anyag	t		

2. táblázat

Amint az látható, a kétféle MARC formátum nagyrészt ugyanazokat az értékeket használja. A különbségekre és azok megfeleltetésére a MARC 21 formátumra való áttérés támogatására készült [HUNMARC-MARC 21 megfeleltetési táblázat](#) részletesen kitér, ebben a segédanyagban nem foglalkozunk a kérdéssel. A táblázatokat és segédanyagokat igyekeztünk úgy összeállítani, hogy a felhasználó kollégák a számukra irreleváns, esetleg zavaró MARC részeket egyszerűen törölhessék.

A következő táblázatban azt láthatjuk, hogy a különböző *Tartalomtípus*-kifejezéseket milyen kód jelölheti. Amint az látható, egy kifejezésnek több kód is megfelel. Ez abból adódik, hogy a MARC *leader* bizonyos forrásoknál megkülönböztet nyomtatott és kéziratos változatot, míg az RDA *Tartalomtípus*-szótára ezzel a megkülönböztetéssel nem él.

Amennyiben van olyan tartalomtípus, amelyet meg szeretnénk jeleníteni a leírásban, de valamilyen oknál fogva nem a megadott (egyik) MARC kóddal rögzítettük, ezt javítanunk kell, különben ezek a leírások kimaradnak az általunk preferált halmazból, és esetleg belekeverednek egy másik halmazba.

<b>RDA Tartalomtípus</b>	<b>MARC 21 leader/06</b>	<b>HUNMARC leader/06</b>
térképészeti adathalmaz	e vagy f	e vagy f
kartográfiai kép	e vagy f	e vagy f
kartográfiai mozgóképek	e vagy f	e vagy f
tapintható térkép	e vagy f	e vagy f
háromdimenziós tapintható kartográfiai forma	e vagy f	e vagy f
háromdimenziós kartográfiai	e vagy f	e vagy f

forma		
számítógépes adathalmaz	m	l
számítógépes program	m	l
mozgásírás	a vagy t	a vagy b
zenei lejegyzés	c vagy d	c vagy d
előadott zene	j	j
hangok	i	i
beszéd	i	i
állókép	k	k
tapintható kép	k	k
tapintható zenei lejegyzés	c vagy d	c vagy d
tapintható mozgásírás	a vagy t	a vagy b
tapintható szöveg	a vagy t	a vagy b
tapintható háromdimenziós forma	r	r
szöveg	a vagy t	a vagy b
háromdimenziós forma	r	r
háromdimenziós mozgókép	g	g
kétdimenziós mozgókép	g	g
egyéb *	o vagy p	m, n, o, p

3. táblázat

\* Az „egyéb” nem eleme az RDA Tartalomtípus szótárnak. Itt annak jelzésére szolgál, hogy az adott kód nem feleltethető meg egyértelműen az RDA *Tartalomtípus* szótár egyik értékének sem.

A következő táblázat azt mutatja, hogy egy adott MARC kódértékhez milyen *Tartalomtípus*-kifejezések tartoznak. Ez a táblázat az alapja a saját megfeleltetésünk kidolgozásának.

MARC 21	RDA <i>Tartalomtípus</i>	HUNMARC	RDA <i>Tartalomtípus</i>
a	mozgásírás	a	mozgásírás
	tapintható mozgásírás		tapintható mozgásírás
	tapintható szöveg		tapintható szöveg
	szöveg		szöveg
c	zenei lejegyzés	b	mozgásírás

	tapintható zenei lejegyzés		tapintható mozgásírás
d	zenei lejegyzés		tapintható szöveg
	tapintható zenei lejegyzés		szöveg
e	térképészeti adathalmaz	c	zenei lejegyzés
	kartográfiai kép		tapintható zenei lejegyzés
	kartográfiai mozgókép	d	zenei lejegyzés
	tapintható térkép		tapintható zenei lejegyzés
	háromdimenziós tapintható kartográfiai forma	e	térképészeti adathalmaz
	háromdimenziós kartográfiai forma		kartográfiai kép
f	térképészeti adathalmaz		kartográfiai mozgókép
	kartográfiai kép		tapintható térkép
	kartográfiai mozgókép		háromdimenziós tapintható kartográfiai forma
	tapintható térkép		háromdimenziós kartográfiai forma
	háromdimenziós tapintható kartográfiai forma	f	térképészeti adathalmaz
	háromdimenziós kartográfiai forma		kartográfiai kép
g	háromdimenziós mozgókép		kartográfiai mozgókép
	kétdimenziós mozgókép		tapintható térkép
i	hangok		háromdimenziós tapintható kartográfiai forma
	beszéd		háromdimenziós kartográfiai forma
j	előadott zene	g	háromdimenziós mozgókép
k	állókép		kétdimenziós mozgókép
	tapintható kép	i	hangok
m	számítógépes adathalmaz		beszéd
	számítógépes program	j	előadott zene
o	egyéb*	k	állókép
p	egyéb*		tapintható kép
r	tapintható háromdimenziós forma	l	számítógépes adathalmaz
	háromdimenziós forma		számítógépes program
t	mozgásírás	m	egyéb*
	tapintható mozgásírás	n	egyéb*
	tapintható szöveg	o	egyéb*
	szöveg	p	egyéb*
		r	tapintható háromdimenziós forma
			háromdimenziós forma

4. táblázat

\* Az „egyéb” nem eleme az RDA Tartalomtípus szótárnak. Itt annak jelzésére szolgál, hogy az adott kód nem feleltethető meg egyértelműen az RDA *Tartalomtípus* szótár egyik értékének sem.

Csak felületesen megnézve a táblázatot, az a benyomásunk támadhat, hogy egyértelmű megfeleltetés csak a 'j' érték, vagyis az előadott zene esetében lehetséges, hiszen a többi kódhoz egynél több érték tartozik. A legtöbb könyvtár ugyanakkor nem rendelkezik az összes RDA *Tartalomtípus* szótárban felsorolt típusú forrással. Például az 'a' értékhez 4 tartalomtípus-érték tartozik: szöveg, tapintható szöveg, mozgásírás, tapintható mozgásírás. Viszont a könyvtárak

túlnyomó többségének ezekből csak szöveg-típusú forrása van, ezért a legtöbb könyvtár a saját megfeleltetése kidolgozásakor az 'a'-t egyértelműen megfeleltetheti a szövegnek.

*Ezen a ponton szeretnénk leszögezni, hogy ez a segédanyag azért született, hogy a járványhelyzet miatt kialakult távmunkakényszerben ötletet adjunk a kollégák számára, hogy milyen típusú adattisztítással / adatgazdagítással szolgálják legjobban a katalógusuk jövőbeni érdekeit. Eredeti terveink szerint a Tartalom-, Média- és Hordozótípusok generálására és használatára vonatkozó segédanyagot akkor tettük volna közzé, amikor az összes speciális típusokra is kidolgoztuk az ajánlásokat. Ezt (az adott típusokkal gyakorlati tapasztalattal rendelkező) külső szakértők bevonásával szerettük volna megvalósítani. Erre még nem került sor. Úgy gondoltuk ugyanakkor, hogy a legtöbb könyvtár számára a hagyományosabb forrásokra koncentráló anyag is hasznos lesz, mivel a speciálisabb típusok természetüknél fogva csak kevés könyvtár gyűjtőkörébe tartoznak bele.*

Ha a számunkra irreleváns tartalomtípusok figyelmen kívül hagyása után még mindig van olyan kód, amelyhez egynél több érték tartozik, akkor nézzük meg, hogy van-e a leírásban olyan kiegészítő információ, amelynek segítségével meg tudjuk különböztetni egymástól a két (vagy több) típust.

Mivel ennek kiváltására hozták létre, mind a tartalom-, a média- és a hordozótípus esetén a legkézenfekvőbb kiegészítő információ a 245 \$h almezőben található kifejezés.

A tartalomtípus független attól, hogy a forrás kéziratos, nyomtatott, esetleg elektronikus formában érhető-e el. Tehát egy kéziratos szöveg, egy könyv és egy e-könyv tartalomtípusa egyaránt szöveg.

Számítógépes anyagnak/számítógépes fájlak kódolt forrás elvileg csak olyan lehetne, amely megfelel az *RDA Tartalomtípus* szótár számítógépes adathalmaz és számítógépes program értékeinél található definícióknak. Az egyéb elektronikus forrásokat a bennük kifejezett tartalomnak megfelelően kell kódolni. Amennyiben ezt a múltban nem így kódoltuk, javítsuk, amennyiben ez lehetséges. Ebben segítségünkre lehetnek: 008/elektronikus dokumentumok/26. pozíció, esetleg a 538 tartalma vagy éppenséggel a hiánya.

Az *RDA Tartalomtípus* szótárban nincsenek összetett tartalomtípusok, így a multimédia, a különleges oktató anyag, a vegyes dokumentumcsomag (vagy -készlet) és a vegyes anyag nem feleltethető meg egy kifejezésnek sem. Ha a forrás többféle tartalomtípussal leírható, az RDA három opciót javasol:

1. Rögzítsünk minden releváns tartalomtípust.
2. Csak a legjellemzőbb tartalomtípust rögzítsük.
3. Csak a forrás legjelentősebb tartalmi egységét jellemző tartalomtípusokat rögzítsük.

Bármely opciót választjuk is, az ilyen típusú források tartalomtípusát(-típusait) csak a forrás újbóli kézbevételeivel tudjuk majd megfelelően rögzíteni.

## Áttekintés a teendőkről

1. Tanulmányozzuk át az 1. táblázatban található *Tartalomtípus*-szótárat!
2. Válasszuk ki, melyek azok az értékek, amelyek a gyűjtőkörünk szempontjából relevánsak!
3. Gyűjtsük össze, hogy a különböző tartalomtípusokat hogyan katalogizáljuk jelenleg, illetve hogy bántunk velük a múltban!
  - a. Fordítsunk kiemelt figyelmet a *leader* 06-as pozíciója kódolásának vizsgálatára!
4. A 3. táblázatot alapul véve vizsgáljuk meg, hogy egy adott tartalomtípust képviselő összes forrást megfelelően kódoltuk-e! Amennyiben a források egy csoportját nem a táblázatban látható módon kódoltuk, akkor mit rögzítettünk helyette? Le tudjuk keresni ezt a csoportot egy vagy több keresőkérdés segítségével?
  - a. Ha igen: mi ez a keresőkérdés?
  - b. Ha nem: akkor az a nagyobb halmaz, amibe keresőkérdéssel szétválaszthatatlanul belekeveredett, mekkora elemszámú? Reális opció-e ezt a halmazt egyesével átnézni és manuálisan szétválogatni?

Ahol lehet, javítsuk. Ha maradt olyan rossz kódolás, amit semmilyen eszközzel nem tudunk elkülöníteni és javítani: mi helyett mit kódoltunk? Nagyságrendileg hány rekordot érint?
5. A 4. táblázatot alapul véve vizsgáljuk meg, hogy egy adott *leader*ben szereplő kódot egyértelműen meg tudunk-e feleltetni egy és csak egy tartalomtípusnak! Ezen a ponton ne tegyünk különbséget aközött, hogy azért nem csak egy tartalomtípus felel meg, mert ez van a táblázatban, vagy azért, mert a feldolgozó gyakorlatunk miatt vegyes az azonos módon kódolt források tartalomtípusa. Amennyiben az egy kódhoz többféle tartalomtípus tartozik, melyek ezek? Le tudjuk keresni a különböző tartalomtípusokat külön-külön is egy vagy több keresőkérdés segítségével?
  - a. Ha igen: mi ez a kérdés?
    - i. Ha a 4. táblázatban azonos kóddal szerepelnek a tartalomtípusok: nincs több teendők.
    - ii. Ha a mi gyakorlatunk miatt vegyes a halmaz: javítsuk!
  - b. Ha nem: akkor a kódhoz tartozó rekordhalmaz mekkora elemszámú? Reális opció-e ezt a halmazt egyesével átnézni és manuálisan szétválogatni?

Ahol lehet, javítsuk! Ha maradt olyan rossz kódolás, amit semmilyen eszközzel nem tudunk elkülöníteni és javítani: mi helyett mit kódoltunk? Nagyságrendileg hány rekordot érint?
6. Dokumentáljuk javításainkat és a javítás utáni helyzetet a nem javítható esetekkel együtt! A nem javítható esetek feltárása azért is kiemelten fontos, mivel ezekben az esetekben erősen megfontolandó, hogy érdemes-e visszamenőlegesen legeneráltatni a *Tartalomtípust*. Általánosságban azt mondhatjuk, hogy ahol túl sok esetben képződnének félrevezető tartalomtípusok, ott tartózkodjunk a visszamenőleges változtatásoktól.

(kész.: Ilácsa Szabina)