



RÖVID ÖSSZEFOGLALÓ

**az országos könyvtárstatisztikai adatbázis összeállításáról,
az adattisztítási feladatokról és az adatok megjelenítéséről**

2021-ben a Könyvtári Intézet Kutatási és Elemző Osztályának egyik kiemelt és nagy ráfordítást igénylő feladata volt a 2015 és 2020 közötti időszakra vonatkozó országos könyvtári statisztika idősoros egységesítése és az adatok pontosabbá tétele. Ennek érdekében adatbázis-egyesítést, -tisztítást és adatjavítást végeztünk. Az idősoros összehasonlíthatóság érdekében a meglévő adatokat közel 700 változóban egységesítettük. Évente nagyjából 4000 adatszolgáltató 700 oszlopnyi adatot töltött fel, így mi a feldolgozott öt év során keletkezett 16 millió cellányi adattal dolgoztunk. Az új adatbázist elérhetővé tettük Excel és SPSS fájlformátumban is. A fájlok közvetlenül beolvashatók a legtöbb statisztikai szoftverbe. Bővebben az adatbázisról és a változókról a Könyvtári Intézet honlapján [▶](#) lehet tájékozódni.

Az adatbázissal és a benne lévő adattömeggel az alábbi munkafolyamatokat végeztük el:

- fix változókat hoztunk létre, és ezek alá 2015-től visszamenőlegesen besoroltuk az adott változóhoz tartozó értéket. Ez a legtöbb esetben egyértelműen megtehető volt, mert az adott adat létezett minden évben, egyes esetekben viszont háttérszámításokat kellett végezni, hogy folytatható legyen az idősor;
- töröltük azoknak a celláknak a tartalmait, amelyek oda nem illő adatot tartalmaztak, jellemzően azért, mert olyan könyvtár is megadta az értéket, akire nem vonatkozott a kérdés;

- vizsgáltunk összefüggéseket a változók között és extrém értékeket, és ahol az egyértelműen megtehető volt, ott elvégeztük a javítást, ez ügyben több esetben kapcsolatba léptünk az adott adatszolgáltatóval (összefüggések vizsgálatára példa: ha nincs weblapja, akkor nem mondhatja róla, hogy akadálymentes; extrém értékre példa: egy adatszolgáltató nagyságrendekkel eltérő értéket adott meg, mint más, hasonló paraméterekkel rendelkező adatszolgáltatók, pl. az ezer forintban megadandó összeget nem osztotta ezerrel);
- felülvizsgáltuk az adatszolgáltató könyvtárak besorolását, elsősorban az „egyéb” típusba soroltakat, amelyeket egyenként besoroltunk egy relevánsabb, meglévő vagy újonnan létrehozott kategóriába;
- javítottunk elgépeléseket.

A megszerzett tapasztalatok és az adatszolgáltatói visszajelzések alapján folyamatosan bővítjük az adatlap-szintű ellenőrzőképletek sorát; immár több mint 420 ellenőrzőképlet felel azért az elektronikus beküldőfelületen, hogy a lehető legkevesebb pontatlan adat kerüljön be az adatbázisba. Ezenfelül a kitöltési útmutatót is jelentősen bővítettük és egyértelműsítettük. Az útmutató elérhető egyben, PDF-formátumban is, de a KultStatban kitölthető adatlapban az egyes cellákba kattintva közvetlenül is elolvasható az adott kérdéshez tartozó segítség.

Céljaink között nemcsak az új adatbázis létrehozása, az adatok tisztítása és a jövőbeni adatszolgáltatás adatminőségének javítása szerepelt, hanem a létrejött, harmonizált adatok bemutatása és könnyen áttekinthetővé tétele is, amelyet adatvizualizációs eszközök segítségével valósítottunk meg.² 

² Forrás:
ki.oszk.hu 