

A TÁRSADALOMTUDOMÁNYI ADATFILE-OK ORSZÁGOS INFORMÁCIÓS RENDSZERE FELÉ

WHITE, Howard D.: Toward a national information system for social science data files c. cikkének (Journal of the American Society for Information Science. 28. vol. 1977. 6. no. 313–321. p.) tömörítését DIENES Geneonné készítette.

A számítógépen tárolt társadalomtudományi adat-file-oknak, melyek statisztikai felvételek, tesztek, népszámlálások stb. adatait tartalmazzák fontos szerepük van a kutatás szempontjából. Ezek az adattárak általában szélesebb körben nem is hozzáférhetőek, annak ellenére, hogy több országos intézmény tett már ezirányban erőfeszítést, így például az Interuniversity Consortium for Political and Social Research, a Roper Public Opinion Research Center és a National Technical Information Service. Egységes bibliográfiai leírásuk sem megoldott még.

Jelen cikk két lehetséges megoldást tárgyal a társadalomtudományi adatbázis szélesebb felhasználására: a számadatos file-ok könyvtári katalogizálása országos szövetségi katalógusban és az országos szolgáltatás létrehozásának lehetőségét, mely az adatokat kivonatolná, indexelné és on-line kutatás számára tenné hozzáférhetővé.

Az adat-file-ok általában két részből állnak: a kódkönyvből, mely az ember számára érthető formában adja a bibliográfiai leírást és a gép számára olvasható rész, mely kártyán, lemezen vagy szalagon tárolt számadatokból áll. Ezek a számok változó értékek és a kódkönyv szolgál az egyes számok tartalmának azonosítására. A kódkönyvre tehát mindenkinek szüksége van, aki hozzá akar férni a file-okhoz. Szerencsére ez a könyv bárhova eljuttatható, így tehát indexelésnél és katalogizálásnál is jól használható.

Az adat-file-okat a könyvtárak „nem könyvként” kezelt dokumentumoknak tekintetik és elhelyezhetik címfelvételüket a katalógusban illetve elkészülhet róluk az országos szövetségi katalógus. Kérdés azonban, hogy vajon egy ilyen katalógus elégíti-e ki legjobban a felmerült igényeket. Bizonyos okok miatt ugyanis nem minden adatbázis nyilvános, illetve nem mindegyik terjeszthető könyvtárközi kölcsönzés formájában és ez csökkenti a katalógus értékét.

A másik megoldás lenne az adat-file-ok országos rendszerének kialakítása kombinálva pl. az Országos Műszaki Információs Szolgálattal (NTIS), vagy az Oktatási Források Információs Központjával (ERIC). Ez a rendszer egységes bibliográfiai leírásokat adna, továbbá a file-ok tartalmának összefoglaló kivonatát és a kódkönyvet. Ilyen rendszerrel jelenleg csak a NTIS rendelkezik.

A következőkben e rendszer egy megoldását ismertetjük.

1972-ben jelent meg John BYRUM és Judith ROWE munkája az adat-file-ok bibliográfiai leírásáról: indexeléséről, osztályozásáról és katalogizálásáról. Legfőbb érdemük, hogy meghatározzák azt a dokumentációs szintet, amelyre a legtöbb felhasználónak szüksége van: vannak felhasználók, akiknek csak bibliográfiai címfelvételre: szerző, cím és tárgykör meghatározására, de vannak olyan felhasználók is, akiknek tartalmi kivonatra és kódkönyvre is szükségük van. Végül azoknak a felhasználóknak, akik maguk akarják az adatbázist használni, szükségük van maguknak az adatoknak a leírására is, például egy adott mágnesszalag jellemzőire.

A 70-es években az American Library Association's Catalog Code Revision Committee új angol-amerikai katalogizálási szabályzatot dolgozott ki számadatos adatbázisokra, melynek kiadása első alkalommal teszi lehetővé azt, hogy a könyvek katalógusa kiegészüljön a könyvekhez tartozó számadatos adatbázisok leírásával. Ezek a katalogizálási szabályok azonban nem tartalmazzák a tárgyszavazás szabályait. A Library of Congress tárgyszó-jegyzékének alkalmazása pedig nyilvánvalóan problémákat fog felvetni. Ilyen probléma lehet például az, hogy egy fogalom ill. tárgyszó túlságosan tág, túl általános (ilyen tárgyszó például a „társadalmi helyzet” vagy „társadalmi felmérés”). Az indexelés felhasználói pedig általában speciális tárgyszavakat igényelnek. Ezért nagyon fontos, hogy a könyvtár, amelynek az adat-file-okhoz tárgyszókatalógusa van, kódkönyvvel is rendelkezzen. Ha ugyanis a kutató a tárgyszókatalógustól nem fordulhat egyenesen a kódkönyvhöz, akkor a könyvtár nem érte el a célját a katalógus felállításával: nem biztosította a felhasználó számára az adatokhoz való tényleges hozzáférést.

A Library of Congress címfelvételi szabályzata szerint, ha számítógépes számadatos adattárról van szó, úgy ezt a cím után szögletes zárójelben tartalmazza a címfelvételt. Ez azonban a katalogizálás szempontjából nem látszik elegendőnek, kívánatos lenne, hogy a Library of Congress Tárgyszó-jegyzékébe felvegyék a „Kódkönyv” tárgyszót, mely alatt a számítógépes adattárak egy címszó alatt összegyűjthetők, ellenkező esetben azok a témának megfelelően szétszóródnak a különböző tárgyszavakhoz (pl. Társadalmi helyzet és USA).

Végül még egy lehetséges megoldás: megszervezni, hogy az Országos Szövetségi Katalógus mely eddig csak könyveket, térképeket és sorozatokat ölelt fel, ezentúl tartalmazza a számadatos adatbázisok leírását is. Ez azonban ismét felvet egy fontos problémát: a szövetségi katalógus célja, hogy lehetővé tegye a kutatók, felhasználók számára, hogy megtudják, hogy a keresett könyv mely könyvtárban található, és ha ezt megtudta, úgy könyvtárközi kölcsönzés révén hozzá is tud jutni. Nem ez a helyzet a számadatos adatbázisokkal. Ezek legtöbbje – így például az Interuniversity Consortium for Political and Social Research (ICPSR) és a Roper Public Opinion Research Center – jelentős előfizetési díj ellenében szolgáltatja a felhasználóknak az adatokat és meg is tiltja előfizetőinek az adatok továbbadását nem-előfizetőkhez. Ha találunk is számos olyan egyetemi, intézeti adatbázist, melyeknél a fenti megkötések nem állnak fenn, az adatok kölcsönzése mégsem olyan egyszerű, mint a könyveké és nem is annyira olcsó. Így mind a felhasználó, mind a kölcsönző könyvtár meggondolja, hogy érdemes-e valamely adatbázis adatait kölcsönkérni. Ahhoz, hogy erre jó lelkiismerettel igenlő választ tudjanak adni, valószínű, hogy a szövetségi katalógus vagy valamely könyvtári katalógus tárgyszórendszerét meglehetősen szegényes segédeszköznek fogják tartani. Mindenesetre nem elég speciálisnak, nem elég egyedi-

nek. Nem is cél ezért a számadatos adat-file-oknak a könyvtárközi kölcsönzésbe való bevonása. Vannak azonban az adatbázisokhoz, adat-file-okhoz tartozó, viszonylag olcsón beszerezhető segédeszközök: tartalmi kivonatok és kódkönyvek, ezeket kellene szélesebb körben terjeszteni. A hiba az, hogy ezeket a tartalmi kivonatokat különböző stílusban és hosszúságban írták és a legkülönbözőbb bibliográfiai kiadványokban jelennek meg, melyekhez a hozzáférés ismét esetleges. Ezért javasolni lehet, hogy az adat-file-okról készült kivonatok és a hozzájuk tartozó kódkönyvek leírását építsük be valamely már működő bibliográfiai szolgáltatásba, mint pl. az ERIC-be vagy a NTIS-ba. Az ERIC rendszer erre felépítése, témaköre és formája miatt is nagyon alkalmasnak látszik. Az ERIC lehetőséget nyújthatna továbbá arra, hogy a felhasználók mikrofilmlapon (mikrofiche) megkapják a kért adatbázis kódkönyvét is, de arra is van lehetőség, hogy a kódkönyv átadásának jogát az adatbázisok fenntartják maguk számára.

Az adat-file-ok illetve adatbázisok fenntartói számára is előnyt jelentene egy ERIC-féle bibliográfiai adatbázishoz való kapcsolódás, hiszen így szélesebb felhasználói réteghez jutnának. Másik előnyt jelentene, ha a felhasználók valamely on-line rendszerben szerezhetnének tudomást adat-file-ok létezéséről. Az ERIC rendszer ezt az előnyt is biztosítani tudja. Végül nagy előnyt jelentene, hogy az ERIC rendszer címfelvételi rendszerében a felhasználó részletes tájékoztatást kaphatna az adatbázisokkal kapcsolatban; a rendszer ugyanis tartalmazná a következő adatokat: szerző neve, a file-címe, az adatokat gyűjtő intézmény, a kiadás éve, a kódkönyv oldalszáma és ára, az azt szolgáltató intézmény címe, a file-hoz kapcsolódó kiadványok illetve tanulmányok, a mágnesszalagok adatai. Az ERIC rendszer valamely adat-file leírásához 10–15 indexszót használna, szemben a Library of Congress tárgyszókatalógusával, indexelése pedig lehetővé teszi, hogy a felhasználó megtudja, hogy a keresett adat-file földrajzilag hol található meg. Az ERIC tezaurusz áttekinthetőségéből láthatjuk, hogy a terminológiája nem speciálisan oktatási profillal rendelkezik, indexelő nyelve sokkal általánosabb érvényű.

Számadatos adat-file-oknál azonban nemcsak egy-egy file címét kell tudni visszakeresni, hanem ennél mélyebben, az egyes változókat, adatokat magukat. Ez pedig csak úgy lehetséges, ha megfelelő indexelési technikával rendelkezünk. Erre a Roper Center-nek van egy érdekes technikája, mely valamely felvétel egyes elemeit fogalomláncok segítségével oldja meg. Ez a módszer viszonylag sok változót viszonylag kevés hely felhasználásával, kivonat keretei között is vissza tud adni. Ha ezt a módszert az ERIC on-line rendszerével össze lehetne kapcsolni, akkor megoldható lenne a számadatos file-ok dokumentálása, sőt ezek rendszeres, negyedévi, éves vagy akár rendszertelen szelektív szétsugárzása is. Ehhez azonban az kell, hogy az ERIC rendszer fenntartója, a National Institute of Education részéről ez a döntés megszülessen. Ez nagyban előmozdítaná a társadalomtudományi információs rendszer fejlődését.