

A SEKÉLYFŰRÁSI ADATOK BEILLESZTÉSE A NATÉR-ADATBÁZISBA

HALUPKA GÁBOR – OROSZ LÁSZLÓ

INTEGRATION OF DATA OF SHALLOW DRILLINGS INTO
A NAGIS DATABASE

Abstract

The authors intend to present a revision of data of shallow drillings that were carried out in the 1960s and the 1990s in Hungary. Based on this process, the revised data could be incorporated into a NATÉR (NAGIS) database, which is established to support decision making in the context of climate change and adaptation. The primary objective of this work was to render the old datasets compatible with the new database. Since the old data were generated decades ago, and without a digital database in the background, an essential task of our work was to make proper, error-free data communication between the old data and the new environment. Review of this work can shed light on the structure of the database, which can also foster and facilitate the understanding and usage of the database.

Keywords: database, shallow drillings, NAGIS, data-compatibility, error-free operation

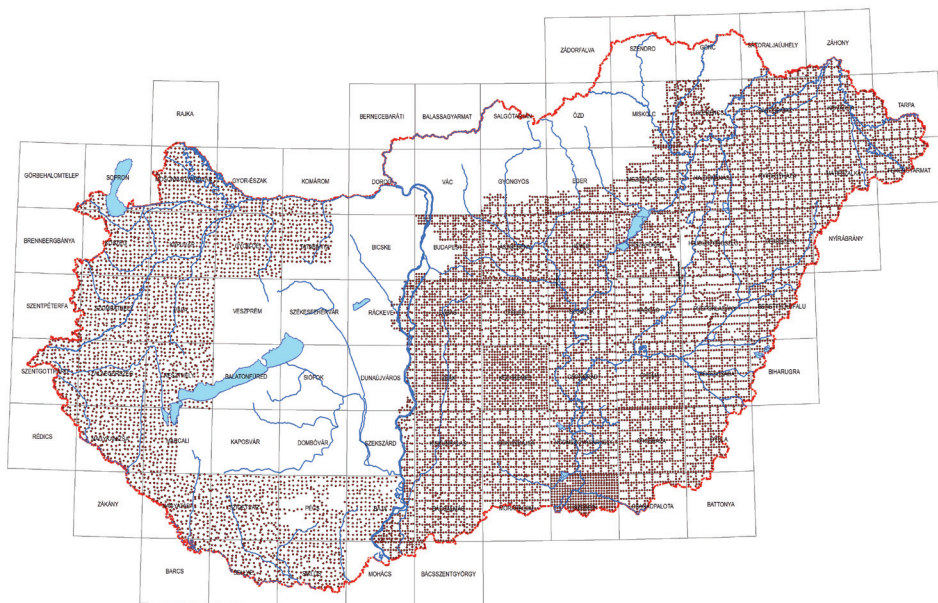
Bevezetés

Az 1960-as években, az akkori Magyar Állami Földtani Intézet Síkvidéki Osztálya nagy ívű sekélyföldtani térképezési programot indított, amelynek célja a nem-hegyvidéki területek felszíni és felszínközeli üledékeinek feltárása volt. Az előirányzottan 10 m talpmélységű fúrások az 1960–80-as években az Alföld feltárásával kezdődtek, majd az ország más tájain is folytatódtak. Az alkalmazott szabályos fúrási hálóterv és a kialakított saját laboratóriumi háttér biztosította a feladat optimális megoldását (RÓNAI A. 1985; KUTI L. [szerk.] 2009) (*l. ábra*). A felvételek eredményeként képződött hatalmas adatállomány nemcsak a korszerű síkvidéki földtani térképsorozatok megszerkesztését tette lehetővé, hanem támogatta az agrogeológiai, környezetföldtani, településgeológiai kutatásokat is.

A több tízezer fúrás feldolgozása jelentékeny időt és erőforrást vett igénybe. A fúrásokhoz vízkémiai és szedimentológiai elemzések, terepen felvett fúrásleírások is kapcsolódtak.

Az adatmennyiség ugyan jól volt tárolható papír adathordozón, ám nehézkesen volt kereshető. Ez, valamint a számítógépes adatfeldolgozás fejlődése volt a motorja az újabb adatfeldolgozási munkáknak, amelyek során a laboratóriumi eredmények Excel-táblázatokba kerültek. A terepi (makroszkópos) leírásokat beszkenelték, ezzel létrehozták az első digitális sekélyfúrás adatbázist (MÜLLER T. 2006).

Az adatbázis építése azonban elszakadt az informatika felgyorsult fejlődésétől: a szoftverkövetés elmaradása miatt az eredetileg sokrétű keresési lehetőségeket biztosító adatbázist már nem lehetett használni. 2004-ben a MÁFI életre hívta a GeoBank nevű digitális fúrás adatbázist (GYALOG L. et al. 2005). A szándék az volt, hogy a Földtani Intézetben megtalálható valamennyi fúrás egy adatbázisba integrálják, és nyilvános webfelületen kereshetővé tegyék. A GeoBank megszületése vetette fel annak igényét, hogy a sekélyfúrás adatok is helyet kapjanak az új környezetben. Az ehhez vezető első konkrét lépések a 2010-es évek elején történtek meg. 2014-ben, immár a Magyar Földtani és Geofizikai Intézet (MFGI) berkein belül, a Nemzeti Alkalmazkodási Központban (NAK) javában



1. ábra Az 1960–90-es években mélyült sekélyfúrások elhelyezkedése
 Figure 1 Shallow drillings in different regions of Hungary were carried out in the 1960–90s

zajlott az a munka, amelynek leglátványosabb eredménye a NATÉR, az éghajlatváltozás magyarországi hatásait térinformatikai rendszerben kezelő, multi-tematikájú, döntéstámogató adatbázis megszületése volt. E feladat igényelte a sekélyfúrási adatok beillesztését az éghajlatváltozást leíró modellekbe. A lépést sokrétű adat-kompatibilitási vizsgálat előzte meg az eredeti sekélyfúrási adattáblázatok és a fogadó adatbázis struktúrája között. A GeoBank szerepe az volt, hogy strukturálisan mintaként, adatbeillesztési szempontból pedig egyfajta közvetítő felületként szolgált a NATÉR adatrendszere felé. A cél eléréséhez lehetővé kellett tenni az alap-, kémiai, és szedimentológiai adattípusokat tartalmazó sekélyfúrási adathalmaz egységes szemléletű kezelését, valamint az egyértelmű adatkapcsolatokat. Ennek hiányában az adatkapcsolatok nem működhettek, a NATÉR-ba történő adatintegrálás nem lehetett sikeres.

Cikkünk az adatbázis struktúráját, az azt kialakító megfontolásokat mutatja be. Az elvégzett munkával kapcsolatban örömmel vennénk megjegyzéseket, véleményeket, hiszen a célunk a felhasználói igényeket kielégítő, működő, jól kereshető adatbank létrehozása.

Az adatbázis revideálása, módszertani megfontolások

Az adatbázis 15 575 sekélyfúrást tartalmaz (2016 januári állapot), amelyhez több százezer laboratóriumi adat kapcsolódik. Ezek NATÉR-ba integrálása számos kérdést vetett fel, amelyek alapvetően a fogadó adatbázis adatstruktúrájába való beilleszthetőséget érintették. Azt kellett ugyanis elérni, hogy a nem egységes adatbázis hibamentesen illeszkedjék a NATÉR adatszerkezetéhez. Ennek jegyében a 2014-es év második felében zajlott le a sekélyfúrási adatok SQL-alapú tesztelése, az adatkapcsolatok működését gátló hibák feltárása, a hibák javítási módszerainak kimunkálása, valamint maga a hibajavítás.

Az adatok nagy mennyisége számos adatkapcsolati hibát generált, annak tulajdoníthatóan, hogy a beemelendő, alapadat/kémiai adat/szedimentológiai adat osztatú adathalmaz részei között fennálló tartalmi kapcsolat az adatbázis-kezelő program számára nem minden esetben volt felismerhető, mert ebben valamilyen látszólagos vagy tényleges hiba megakadályozta. Ilyen látszólagos hibák voltak a nevezéktant érintő (szintaksziális) eltérések a három adathalmaz között, míg tényleges hibát az adatsorok számítógépre vitele során kimaradt/hibásan bevitt rekordok okoztak. *A tesztek során előfordult hibák döntő többsége a nevezéktani hibatípusba volt sorolható.* Az előfordult hibák jellege, mennyisége informatikai alapú, „tömeges” hibafeloldást kívánt. A hibákat első lépésben feloldani, majd második lépésben ténylegesen megoldani volt szükséges. Az első lépés az adatbázis működését (kereshetőségét, lekérdezhetőségét) állította helyre, míg a második az adatbázis szakmai tartalmán javított.

A kitűzött célt többkörös, SQL-támogatott, adatkapcsolat-lekérdezésekre támaszkodó hibaazonosítással, majd a feltárt hibák szakmai döntéssel nyugvó, szintén SQL-alapú tömeges/egyedi javításával értük el, kérdéses esetekben ellenőrizve a papír alapú archív jegyzőkönyveket.

Az adatbázis hibajavítása során a legfontosabb megoldandó feladat az volt, hogy az adatbázis kezelőprogramja egy adott keresés végén ne küldjön hibüzenetet – ugyanakkor nem a hibüzenet kiiktatása volt a cél! A hibafeloldás bizonyos esetekben nem volt egyenlő – sokszor a hiányos archív információ miatt nem is lehetett egyenlő! – a hiba tényleges megoldásával. Vagyis ezekben az esetekben csak a kezelőprogram számára szüntettük meg a hibát, tényleges javítás nélkül. Ha az archív dokumentáció ezt megengedi, a későbbiekben a hiány jelentős manuális munkával ugyan, de megszüntethető. A hibás adatok illetően státuszát a megfelelő helyen jeleztük.

A korrekció során első lépésben szükséges volt olyan alapelvek lefektetése, amelyek révén azonos szemlélettel lehetett az azonos típusú hibákat feloldani. A feltárt hibákat összesen 6 hibaosztályba soroltuk, s ezzel nyílt meg az út a hibák tömeges kezelése, azonos szemléletű feloldása előtt, ami többkörös korrekciót követő, ismételt lekérdezéseken alapult. Ez a munkamódszer nem csupán a hatalmas adatmennyiség miatt volt indokolt, hanem így olyan hibatípusokat is feltártunk, amelyek a munka kezdetén még nem voltak ismertek, lévén e hibák nem befolyásolták érdemben a fő adattáblák adatkapcsolatait, nem okoztak közvetlenül hibüzenetet sem, s rejtve maradhattak.

A javítási gyakorlat alapján megfogalmazott módszertani megfontolásokban azonban azt is feltételeztük, hogy minden egyes fűrés egyedileg, egyértelműen és tartósan azonosítható. Mivel az adatkapcsolati hibák jelentős részének nevezéktani oka volt, ezért egyértelművé vált, hogy a fűrés saját neve/jele a vártnál sokkal kevésbé stabil jellemző, tehát nem ez biztosítja a fűrés egyértelmű azonosíthatóságát. Emiatt minden egyes fűrés olyan egyedi azonosító számot kapott, amely szakmai (földtani) jelentéssel nem rendelkezett. Az ilyen típusú azonosító szám bevezetésével úgy tudtuk egyedivé tenni a fűrés/adatot, hogy teljesen kikerülhetővé vált egy későbbi, bármilyen szempontú nevezéktani változtatásokból eredő látszólagos/tényleges adatvesztés vagy adatbizonytalanság.

Az alábbiakban sorra vesszük a javítási alapelveket, a hibaosztályokat, majd a munka során feltárt hibákat.

A feltárt hibák és a javítás alapelvei

A javítások adattábla-, illetve hiba-specifikusak voltak. Az első esetben a hibák megjelenése jól köthető egy adott típusú adattáblához, míg a második esetben hat, leginkább

az eredeti adatbevitellel kapcsolatos problémát kellett kezelnünk. Emellett megemlíten-
dők azok a hibák, amelyek épp a javítás során fokozatosan helyreálló adatkapcsolatok és
hivatkozások révén válhattak ismertté. E hibákkal kapcsolatban is néhány, jól definiálható
és kezelhető problémába ütköztünk.

Adattábla-specifikus hibák

Ebben az esetben voltaképp három fő adattábla (alap-, kémiai, szedimentológiai) „rend-
ellenességeit” kellett kezelnünk. Az alapadati táblában a hibák vagy az adatok kiindulási
Excel-táblázataiban végzett műveleteihez kötődtek vagy magából az adatállományból
fakadtak. Az előbbi az Excel működési sajátosságainak következménye volt, például
egy dátumbejegyzést elvégzendő műveletként értelmezett a program. Az utóbbi esetben
pedig a hibák döntően a hiányos adatbevitelre voltak visszavezethetők, ún. nem egységes
szemléletű fűrasszám-elnevezések, koordináta-hiányok vagy a fűrás mélyítési dátumá-
nak, helyszínének hiánya. A feltárt hibák döntően az érintett adatok ismételt adatbázisba
emelésével voltak kiküszöbölhetők.

A szedimentológiai adattáblában két fő hibatípus mutatkozott. Az ún. „alsó rétegha-
tár” és „felső réteghatár” adatoszlopa nem egységes sorrendben jelent meg; valamint
a mintára vonatkozó általános megjegyzések (pl. „tőzeges minta”) minden egyes betűje
külön-külön cellába került.

A kémiai adattáblák felülvizsgálatakor abból indulhattunk ki, hogy az archív kémiai
laborjegyzőkönyvek két fő adattípust tartalmaztak: vízszint-adatokat és a vízminták labo-
ratóriumi elemzési eredményeit. Így a revideálás is ezt a két adatféléseget érintette. Az
eltérő időben, több laboratóriumban keletkezett adatok között számos eltérés volt tapasztal-
ható, ami szintén nehezítette az egységes adatkezelést és értelmezést. Ilyen probléma
volt a nem-szám típusú bejegyzés szám-definiált cellába, a fűrás alapadataiban megadottól
eltérő dátum megjelenése, a kimutatási határral vagy kémhatással kapcsolatos, nehezen
értelmezhető bejegyzések megjelenése, illetve az egyszerű elütés.

Az adatbevitelre visszavezethető hibák, hibaosztályok

Amikor SQL-környezetben elkezdtük futtatni az első lekérdezéseket, hamar nyilvánva-
lóvá vált, hogy néhány, jól körülhatárolt, az eredeti adatbevitel hiányosságaiból, hibáiból
fakadó hibaféléseget találtunk. Ezek nagy darabszáma jelentős mértékű, ám informati-
kai szempontból hatékony hibakezelési folyamatot vetített előre. Ezeket a hibákat összes-
ségében hat jelentős hibaosztályba sorolhattuk: nem egységes fűrásazonosítók, hiányos
koordináták; hibás vagy nem egyértelmű dátumok; duplumok; nem-szám típusú bejegy-
zések szám-definiált cellába; tartalmi hibák (jellemzően a szedimentológiai adatsorokat
illetően); végül hiányos településnevek.

A nem egységes fűrásazonosítók, illetve hiányos koordináták problémaköre mintegy
2000 fűrást érintett. Talán érzékelhető, hogy egy fűrás ilyen adathiánya alapvető gondot
jelent a fűrás adatainak felhasználhatóságában. Nézzük, mi ennek a háttere? A sekélyfű-
rást a törzsadatok (alapadatok), a kapcsolódó vízkémiai és szedimentológiai laboratóriu-
mi elemzési eredmények, valamint a terepi jegyzőkönyvek együttesen reprezentálják. Így
a NATÉR-ba nem pusztán egy koordinátahalmazt kellett beilleszteni, hanem egy három-
osztatú adattábla-rendszert, az adott fűráshoz kapcsolódó terepi jegyzőkönyvekkel együtt.
A helyzetet bonyolította, hogy a nem egységes nevezéktan miatt sokszor az egy fűrás-
hoz tartozó adattáblák közötti kölcsönös hivatkozás sem volt hibamentes. Égetően szük-
séges volt tehát kialakítanunk egy olyan egységes nevezéktant, amelyben a fűrásnevek

valamennyi adattábla-típusban ugyanolyan elvek szerint épültek fel. Ez a gyakorlatban „szám/szám” vagy „szám/betű/szám” jellegű elnevezés-szintaktikát jelentett az alföldi sekélyfúrások esetében. Az előbbit a korabeli kutatási program kidolgozásakor betervezett fúrások kapták, míg az utóbbi a valamilyen okból terven felül mélyített, pontosító vagy pótfúrásokhoz volt rendelhető. A kisalföldi fúrások tisztán hat számjegyes fúrásazonosító rendszere nem kívánt változtatást, hiszen az alföldi fúrásoknál látott szintaktikai hibák nem fordulhattak elő. A makroszkópos leírások elnevezésének egységesítése is szükséges volt, mert számos esetben a papír alapú leírás beszkenelése több részletben történt, több digitalizált leírás-részt eredményezve. Ezek mellett a koordinátával beazonosított laboratóriumi mérési jegyzőkönyvek és a vonatkozó fúrások közötti adatkapcsolat ténylegesen nem mindig állt fenn. Ezt a problémát azzal tudtuk feloldani, hogy sikerült megtalálnunk az adott labor-adatsorhoz tartozó fúrást és helyazonosítóját (koordinátáját). Ugyanakkor bizonyos esetekben ténylegesen hiányoztak a koordináták. Ez több okra volt visszavezethető. Vagy csak szöveges leírások kaptak helyet a terepi jegyzőkönyvben, vagy a koordinátákat nem EOV-rendszerben rögzítették. Tekintettel az akár 40–50 évvel ezelőtti terepmunkára, ma már nem vagy csak nehezen lehet egyértelműen reprodukálni e fúrások helyzetét. Megjegyezzük emellett, hogy mivel munkánk a digitálisan már rendelkezésre álló adatállományt tekintette át, ezért ez a hiány később pótolható, pótolandó. Ugyanakkor külön kellett kezelnünk azokat a fúrásokat, amelyek esetében a koordináták ugyan elérhetőek voltak, ám a kapcsolódó kémiai/vízföldtani/szedimentológiai adat nem állt rendelkezésre.

A tapasztalt koordináta rendelkezésre-állás vagy hiány alapján adatkezelési döntéseket fogalmaztunk meg (1. táblázat). Hangsúlyozni kívánjuk ezzel kapcsolatban, hogy az azonosíthatatlanság miatt az adatbázisból kikerülő (vagyis: a NATÉR-ba beemelésre nem kerülő) adatokat nem töröltük, adatvesztést nem engedtünk meg. A be nem emelt adatok később, pontosítást követően gazdagíthatják az adatbázist.

1. táblázat – Table 1

Adatkezelési döntések az egyes adattáblákban tapasztalt koordináta
rendelkezésre-állás vagy hiány alapján
Decisions in data handling process according to availability of coordin

Koordináta van (1) nincs (0)	Vízszint-adat van (1) nincs (0)	Kémiai adat van (1) nincs (0)	Szedimentológiai adat van (1) nincs (0)	Adat státusza	Adat sorsa (tárolási helye)
0	1 / 0	1 / 0	1 / 0	Adatbázisból kikerül	Külön tárolás
1	0	0	0	Adatbázisban marad	Adatbázisban tárolás
1	1	0	0	Adatbázisban marad	Adatbázisban tárolás
1	1 / 0	1	0	Adatbázisban marad	Adatbázisban tárolás
1	1 / 0	0	1	Adatbázisban marad	Adatbázisban tárolás
1	1 / 0	1	1	Adatbázisban marad	Adatbázisban tárolás

A hibás vagy nem egyértelmű mélyítési dátummal rendelkező fúrások száma mintegy 1700 volt. Dátumhiány a fúrás alap- és kémiai adatait érinthette. Az alapadati dátum-

hiányt pótoltuk, ha az a papíralapú fúrásdokumentációban elérhető volt. Amennyiben az alapadatok nem tartalmaztak dátumot, az esetlegesen rendelkezésre álló kémiai adatsor dátumát rendeltük a fúráshoz. Ez azt is jelentette, hogy a fúrás mélyítési dátumát követő, akár több héttel későbbi laborelemzési időpont megjelenését fogadtuk el. Ezt az utólag kideríthetetlen hosszúságú „időelcsúszást” az adott fúrással kapcsolatban jeleztük. Ha ezen a módon sem volt lehetőség a dátum konkretizálására, ott ennek ténye egy ún. köztes, a hibát a keresőmotor látóköréből kiemelő, de listaszerűen tartalmazó adattáblában rögzítettük, biztosítva az adatbázis fő adattábláinak zökkenőmentes kereshetőségét. A nem egyértelmű dátum problémájával elsősorban a kémiai elemzések dátumbejegyzései között találkoztunk. Ebben az esetben jellemzően csak az év és a hónap került bejegyzésre. Mivel ezt az információt pótolni szintén nem volt lehetőségünk, ezért a „hiba” létét szintén köztes adattáblában rögzítettük.

A duplikálások problémájával két esetben találkoztunk, jellemzően és legvalószínűbben adatbevitelre visszavezethetően. Az egyik eset a fúrásazonosítókat (2 db), a másik a kémiai adatsorokat (28 db) érintette, a hiba egyedi lekereséssel megszüntethető volt.

Az egyik legnagyobb fúrászámot – szám szerint 9100-at – érintő hiba a „nem-szám típusú bejegyzés szám típusú cellába” módon volt definiálható. A kémiai, valamint a szedimentológiai adattáblában megjelenő hibák közös jellemzője az volt, hogy a szám-definiált cellába nem-szám típusú bejegyzés került. Tulajdonképpen a labor így rögzített olyan paramétereket, amelyek bár a mintát jellemezték (pl. a minta szervesanyag-tartalmára vagy egy adott komponens koncentrációjára vonatkozóan), mégsem voltak számszerűsíthetők. Ezek a problémák az érintett fúrások jelentős száma ellenére meglepően egységes szemlélettel voltak kezelhetők. A kialakított megoldási séma a következő volt. Először a tartalmilag egyező paraméterek egységes elnevezésének kialakítására volt szükség. Ezt követte az így „véges számúra” csökkent szöveges bejegyzés-féleségek kódszámmal történő helyettesítése. A kódszámokkal kiváltott bejegyzés-féleségek megőrzésére köztes adattáblákat hoztunk létre, amelyekben a kódszámokat a vonatkozó szöveges bejegyzésekkel definiáltuk, összekapcsoltuk. Emellett szükségesnek tartottuk az archív laborbejegyzések megőrzését a fő adattáblába egy adatmezőjében. A kódokkal helyettesített bejegyzéseket így hozzárendelhattük a fúrás azonosítószámához („id”-jéhez). E módszerrel valamennyi olyan hibát fel tudtuk oldani, ahol nem volt lehetőség a bejegyzést ténylegesen numerikus értékkel kiváltani. A numerikus értékek alkalmazhatatlanságának oka az archív információ hiánya vagy hiányossága volt.

Az említett megoldási sémán mint általános hibafeloldási munkafolyamaton túl néhány, sajátosan a kémiai adattáblához köthető jellemzőt érdemes megemlíteni. Ebben az adattáblában a fúrási munkákat követő terepi észlelési (vízszint-) adatok, valamint a kémiai laboratóriumi elemzési adatok találhatóak. Az utóbbi adattípusba tartozott egy olyan laborbejegyzés, amely hibaként jelent meg az SQL-alapú lekérdezések során, és léte jól reprezentálja az archív adatok utólagos értelmezésének nehézségeit. A bejegyzés az adott komponens koncentrációjára vonatkozott (a vizsgált paraméter „nyomokban” jelenik meg, amely kifejezés további árnyalatokat is kapott, a „nagyon gyenge” nyomtól a „nagyon erős”nyomig). Mivel ez a szöveges bejegyzés szám-definiált cellában informatikai szempontból nem volt kezelhető, továbbá a korabeli, empirikus laborbejegyzések tényleges szakmai tartalma már nem rekonstruálható, ezért a kimutatási határ felével (esetünkben: 0,05 mg/l) mint numerikus értékkel helyettesítettük a szöveges bejegyzést. Ugyanakkor magát a szöveges bejegyzést is megőriztük, köztes adattáblában.

Az egykori szedimentológiai laborvizsgálatokhoz kötődött, illetve az eredmények későbbi feldolgozásából származott a mintegy 2100 fúrás esetében megjelenő tartalmi hiba. Mit is jelentett ez?

A vizsgálatok során a mintát leszitálták, így szemcseméret-frakciókat állítottak elő. A frakciók összesített tömegének ideális esetben egyenlőnek kellett lennie a kezdeti minta tömegével. Az egyes frakciók közötti tömegmegoszlást százalékosan fejezték ki, azaz a frakciók tömegszázalékban kifejezett összegének 100%-nak kellett lennie. A laboratóriumi gyakorlatban azonban több olyan tényező is árnyalta a vázolt, elvi helyzetet, amelyek miatt a frakciók tömegszázalék-összege nem érte, nem érthette el a 100%-ot; attól általában $\pm 5\%$ -ban tért el. Ilyen tényező volt a szitasoron fennakadó mintaveszteség, vagy a minták nem tökéletes dezaggregálásából visszamaradó konkréciók előfordulása. Amennyiben azonban e jelzett $\pm 5\%$ -os sávból kilépett az eltérés, úgy nagy valószínűséggel a laborjegyzőkönyvek számítógépre vitele során fordult elő adatbeviteli hiba (elütés). Mivel az archív, papír alapú jegyzőkönyvekben történő visszakeresés jelentős emberi erőforrást és/vagy időt igényel, ezért elsődleges célunk a kapott hibák feloldása lehetett, szintén köztes adat-táblában rögzítve a hiba tényét. A tényleges hibajavítás egy jövőbeli korrekció része lehet.

Az utolsó, szinte valamennyi – több mint 15 ezer fúrászt érintő – hibaosztályban a fúrásokhoz tartozó településneveket kellett pótolnunk. A nagy hibamennyiség arra volt visszavezethető, hogy a helymeghatározás során sokszor csupán a térképlap nevének hozzárendelése történt meg a fúrásokhoz, a ténylegesen érintett települések nevei nem. A problémát térinformatikai úton, ArcGIS segítségével tudtuk hatékonyan orvosolni.

Külön szükséges kitérni azokra a hibákra, amelyek a javítások során helyreálló adatkapcsolatok miatt kerültek, kerülhettek felszínre; azaz a hibajavítási folyamat megkezdése előtt ezekről nem lehetett tudomásunk. E hibák láthatóvá válása és kezelése a teljes javítási folyamat hosszát előre nem tervezett módon tovább növelte. E folyamatban három fő hibaféleség jelentkezett: az adatmezők elcsúszása; bizonyos adatmezők adatbázisba emelésének hiánya; nem következetes mintaszám-elnevezések. Bár maga a hibajavítás nem követelte meg kifejezetten az okok kiderítését, egy ilyen irányú „nyomozás” segíthette a hiba keletkezésének megértését is.

Az adatmezők elcsúszása akkor jelentkezett, amikor a beemelt nyersadat-file-ok fejlécstruktúrája nem egyezett meg a fogadó adatbázis fejléckiosztásával, s ennek következtében nem a megfelelő helyre kerültek a vonatkozó adatok. Így bár az adatbázis formálisan kereshető maradt, ám tartalmi (szakmai) szempontból nem valós eredményt kaptunk. Voltaképp a hiba épp e formálisan megfelelő programműködés miatt maradt rejtett az első hibaaazonosítási körben. A hibát mezőeltolással oldottuk (szüntettük) meg, ami a gyakorlatban mezőbeszúrászt, majd a fejléc-definíciók egyenkénti eltolását jelentette, a fejléc-cellák átnevezésével. Az adatmezők hiánya az eredeti adatsorok betöltésének hiányosságából fakadt. Megoldása során a hiányzó mezőket kellett pótlólag beemelni az adatbázis-struktúrába. A fúráselnevezés és a mintaszámok együttes használatával sikerült egyedivé, s így egyértelműen meghatározhatóvá, majd a vonatkozó fúrásához kapcsolhatóvá tenni az addig hiányzó adatmezőket. A mintaszámok nem következetes elnevezésének oka az volt, hogy a fúrások egyes fő adattáblái közötti adatkapcsolat létrehozásához e jellemzőre nem volt szükség, így közvetlenül nem jelentkezett hibaként. Ebben az esetben a kérdéses mintaszámok helyét egyedi vizsgálattal lehetett megtalálni.

A javított adatállomány beemelése a NATÉR-adatbázisba

Mint fentebb jeleztük, a sekélyfúrás adatkör revideálását a NATÉR fejlesztése gyorsította fel azzal, hogy igényelte e fúrás adatok beépítését is az összetett, éghajlatváltozást leíró modellekbe. Mivel a sekélyfúrás adatok NATÉR-ba illesztését a GeoBank mint adatstrukturális minta, továbbá mint adatbeillesztési közvetítő felület tette lehetővé, lényeges volt a GeoBank struktúrájával kommunikálni képes sekélyfúrás adatszerkezet kialakítása.

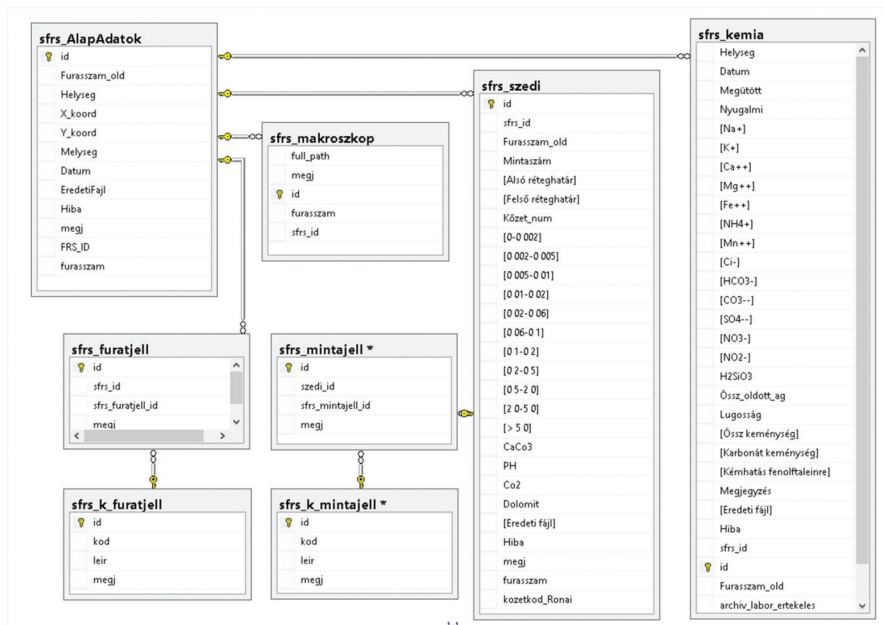
A NATÉR és a GeoBank adatszerkezete közötti eltérés főleg az idősoros adatok tárolásában mutatkozik meg: a NATÉR-adatkörök jelentős része idősoros jellegű (pl. vízállás, csapadék), ami a földtani adatkörre kevésbé jellemző. Ugyanakkor épp a sekélyfúrási adatkörnél fordul elő, hogy egy fúráshoz például eltérő időpontból is tartozhat vízkémiai elemzés, ami így már idősoros adatnak tekinthető.

Az adatok elsődleges feldolgozása helye egy MsSQL 2008 R2-es környezetben levő w_nater munka adatbázis volt. Ide kerültek betöltésre a nyers adatok. A korábban leírt adatrendezési, adatjavítási folyamatokat már ebben az adatbázisban végeztük el. A w_nater adatbázis szerkezete (2. táblázat) és a táblák közötti kapcsolat (2. ábra) a tematika által megkövetelt ideális felépítésű volt.

2. táblázat – Table 2

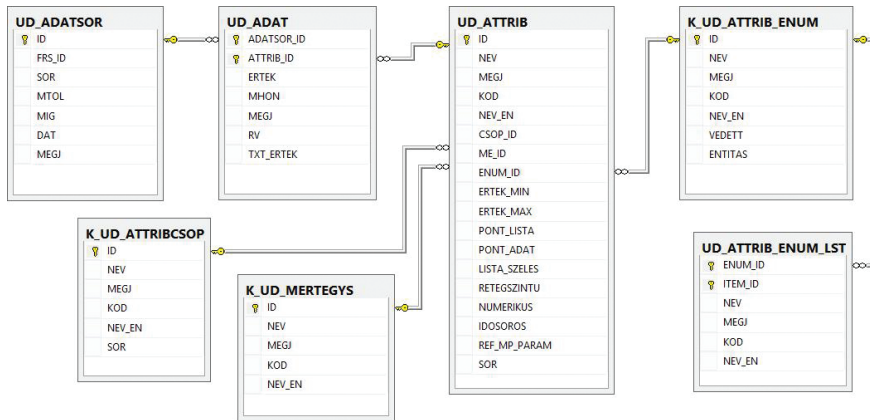
A w_nater munkaadatbázis táblái és azok tartalmának rövid leírása
Tables of w_nater test database and brief description of their content

Táblanév	Tartalom
sfrs_Alapadat	fúrások törzsadata (azonosító, település, koordináta, talpmélység stb.)
sfrs_szedi	a szedimentológiai vizsgálatok adatai
sfrs_kemia	a vízkémiai vizsgálatok adatai
sfrs_makroszkop	a szkennelt (pdf) makroszkópos jegyzőkönyvek elérési útja
sfrs_furatjel	furatjellemzők
sfrs_k_furatjel	az sfrs_furatjel táblában használt szótárértékek
sfrs_mintajel	mintajellemzők
sfrs_k_mintajel	az sfrs_mintajel táblában használt szótárértékek



2. ábra A w_nater munkaadatbázis szerkezeti diagramja
Figure 2 Diagram of structure of w_nater test database

A NATÉR adatszerkezetét azonban ennél általánosabbra, tágabbra kellett tervezni, hiszen a rendszernek szinte bármilyen adatkör befogadására alkalmasnak kell lennie, amit attribútumos tárolási renddel valósítottuk meg. Ennek lényege, hogy az adathoz 4 paramétert – objektumazonosító; időpont; mélységadat; attribútum – kapcsolunk, amiktől egyedivé válik. Az objektumazonosító bármilyen egyedi azonosító lehet, amelyet utána az egész rendszer egységesen használ. Az időpont az adat saját időjellemezője (amennyiben releváns). A mélységadat teszi lehetővé a harmadik dimenzió (mélység, magasság) tárolását; ez esetünkben elengedhetetlen. Az attribútum a tárolni kívánt adatkör definiálását jelenti (pl. vízállás, pH stb.). Mindennek megfelelően a NATÉR-adattárolás szerkezetét a 3. ábra mutatja be. Az objektumok törzsadata továbbra is külön táblában kerül tárolásra. A tábla egy sora egy objektumot definiál.



3. ábra A NATÉR-adatbázis attribútumadat-tároló részének szerkezeti diagramja
 Figure 3 Structure diagram of data attribute storage sector of the NATÉR database

De miképpen is valósul meg a gyakorlatban az attribútum alapú adattárolási rend? Az objektumokhoz adatsorokat rendelünk (UD_ADATSOR). Egy adatsorhoz tartozik az egy időben és egy mélységben mért adatok teljessége. Az adatsorhoz tartoznak az adatok (UD_ADAT), amelyek magukat a mért vagy rögzített értékeket jelentik. Az adat az attribútum definiálásával értelmezhető. Külön táblában (UD_ATTRIB) kerül rögzítésre az attribútumok listája. Az attribútumokat csoportokba soroltuk (K_UD_ATTRIBCSOP), hogy könnyebb legyen az áttekintés, illetve a felhasználás. Továbbá az attribútumok többségéhez tartozik mértékegység, amelyeket szintén külön táblából (K_UD_MERTEGYS) lehet az adatkörhöz hozzácsatolni.

Az (UD_ADAT táblában tárolt) adat lehet numerikus vagy alfanumerikus (betűkből álló). A kéttípusú adatot két külön mezőben – ERTEK, TXT_ERTEK – kell tárolni. Ezek közül egyszerre csak az egyik mező lehet kitöltve.

Meg kellett oldani még a szó-listás (kódszótáras) adatok tárolását is. Ezt további két tábla beillesztésével tudtuk megtenni. Az első tábla (K_UD_ATTRIB_ENUM) definiálja a kódszótárat, míg a másik tábla (UD_ATTRIB_ENUM_LST) tartalmazza a szótár elemeit. Egy kódszótáras elem esetén tehát az UD_ADAT tábla ERTEK mezője van kitöltve és az UD_ATTRIB_ENUM_LST tábla valamelyik ITEM_ID mezőjére utal.

Bemutatunk egy konkrét példát is, a Karcag 573/18-as fúrásra vonatkozóan. A fúrás 10 m mély, 1970-ben mélyült, koordinátája ismert. A hozzá tartozó sekélyfúrás adatkör tárolása a következő logikai sor mentén áll fel (szögletes zárójelben az érintett tábla neve):

- 13 mintát vettek a fúrásból – 13 adatsor [UD_ADATSOR]. Ezek mélység-adata tér el.
- Mindegyik adatsorhoz 12 adat [UD_ADAT] tartozik: 9 szemcseméret-frakció értéke, továbbá CaCO₃-, dolomit-, és pH-adatok.
- A 12 adatot egy-egy attribútumhoz [UD_ATTRIB] rendeljük, amelyek az adat értelmezhetőségét biztosítják. Például a szedimentológia attribútumcsoport [K_UD_ATTRIBCSOP] tagja a '0,02–0,06 mm' nevű attribútum (mértékegysége vol% [K_UD_MERTEGYS]), amelynek értéke az 573/18-as fúrás 7. mintájában (mélység 3–4 m között) 30,2.
- A fúráshoz tartozik még egy fúráskörülményeket leíró adat: 'nem volt víz'. Ez a 'Fúráskörülmények' attribútumhoz [UD_ATTRIB] tartozó 'A fúrás körülményeit rögzítő megjegyzések' nevű kódtábla [K_UD_ATTRIB_ENUM] értékeinek egyike [UD_ATTRIB_ENUM_LST].

Az adatok végső tárolási helyére – a NATÉR-adatbázisba – töltését több lépéssel értük el. A w_nater adatbázisban tárolt adatkor a NATÉR adattárolási rendszer szerint a következő részekre bontható: fúrási adatok; attribútum-adatok; s ez utóbbiak kiegészítő adatai. A fúrási törzsadatok (sfrs_Alapadatok) a NATÉR-adatbázis FURAS táblájába kerültek.

A Bevezetésben említett, első adatbázissal szemben ez a verzió már lehetővé teszi, hogy az egyes fúrásokhoz közvetlenül csatoljunk dokumentumokat. Így például a makroszkópos leírások szkennelt állományai az érintett fúrás „mellől” közvetlenül elérhetők. Ugyanakkor a mintajellemző adatok tárolása egy, a NATÉR-adatbázisban bemutatott, attribútum alapú tároláson túlmutató igényt is támasztott. Ugyanis a mintajellemzők egy adott mintához további – kódszótáras – részletes adatokat tartalmaznak. Ez egy második szintű attribútumkört jelentett. Azonban e második szint informatikai szempontból egzakt kezelése nem volt reálisan elérhető cél, ezért ezeket az adatokat is „felemeltük” hagyományos attribútum-szintre, újabb 10 attribútumot definiálva (3. táblázat). Adatot így nem veszítettünk, ugyanakkor szakmailag helytálló az adatok tárolása. Az adatok értelmezésekor szükséges a felhasználónak szem előtt tartania, hogy ezeket az adatokat külön ki kell kérni.

3. táblázat – Table 3

A mintajellemzők a NATÉR-adatbázisban történő tároláshoz szükséges attribútumai.

- id – egyedi azonosító; csop – az attribútumcsoport neve; nev – az attribútum neve;
- MEGJ – megjegyzés; me – mértékegység

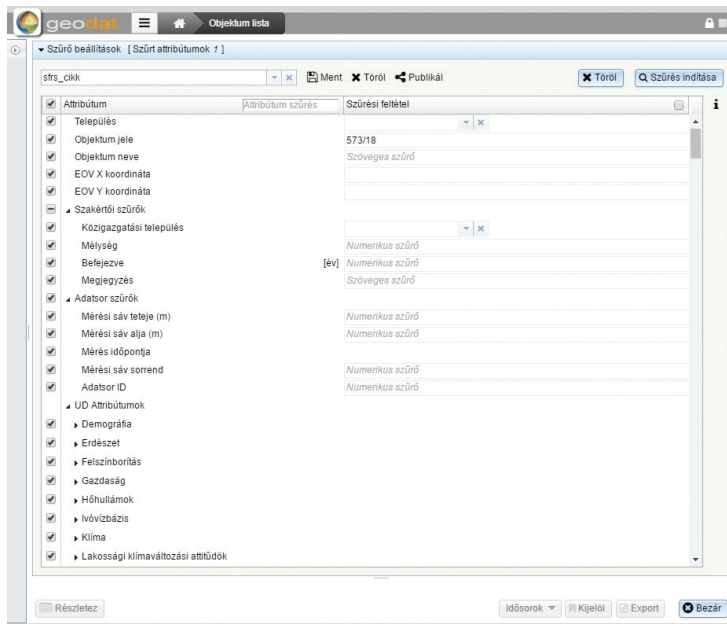
The necessary attributes of data features for storage in the NATÉR database.

- id – unique identification number; csop – name of attribute group; nev – name of the attribute; MEGJ – notice; me – unit

id	csop	nev	MEGJ	me
4007	szedimentológia	konkréción, átszámolva	A mintában előfordult konkréción tömegét levonták a teljes mintatömegből, az így kapott maradék mintatömeget tekintették 100%-nak	–
4008	szedimentológia	csigahéj	Csigahéj- (törmelék-) tartalmú minta	–
4009	szedimentológia	tőzeges	A minta valamely mérettartományában tőzeg fordult elő	–
4010	szedimentológia	vasas kiválás	A minta vasas kiválást tartalmazott	–
4011	szedimentológia	kavics	Az adott mérettartományban előfordult homogén szemcse (kavics)	–
4012	szedimentológia	nyom	A komponens tömegszázalékosan nem volt kifejezhető	–

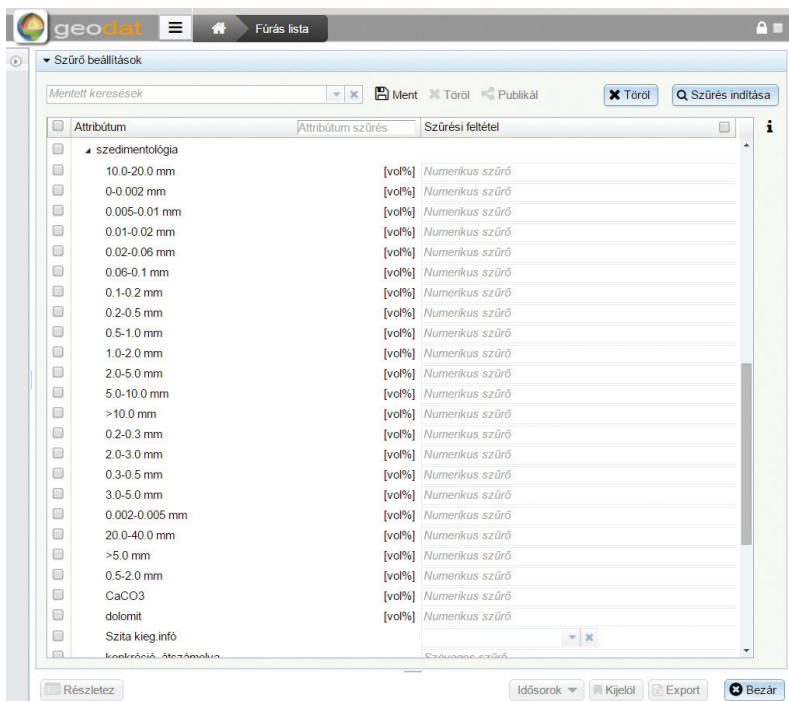
id	csop	nev	MEGJ	me
4013	szedimentológia	tőzeges átszivárgás	A mintának a vártnál nagyobb volt a szerves anyag tartalma, amely a felsőbb szintek tőzeges rétegeiből juthatott le a vizsgált mintába	–
4014	szedimentológia	gipsz	A minta gipsztartalmú	–
4015	szedimentológia	konkrecio	A mintában előfordult konkreció	–
4016	szedimentológia	Megjegyzés	Szedi attribútumhoz tartozó megjegyzés	–

Nem várható el azonban a széles szakmai közönségtől, hogy otthonosan mozogjon a NATÉR-adatbázist kezelő és a háttérből támogató, 50–60 táblás SQL környezetben. A NATÉR-adatbázisnak – és benne a sekélyfűrési adatkörnek is – elérhetőnek és könnyen értelmezhetőnek kell lennie az informatikában nem járatos földtani szakemberek számára is. Ehhez egy olyan kezelőfelületet kellett létrehozni, amely megfelel a kor kihívásainak és él a koradta lehetőségekkel. A NATÉR-adatbázist kezelő alkalmazás a GeoDat nevet kapta. A GeoDat az ismert webes böngészőkben futtatható, a megfelelő elérési út ismeretében (<http://nagis.hu/geodat> – regisztráció után érhető el). Az alkalmazás több szempontú lekérdezéseket tesz lehetővé (4–5. ábra), táblázatos formában megjelenítve az eredményeket (6. ábra).



4. ábra A GeoDat alkalmazás keresőfelületének törzsadat- és egyéb alapadat-kereső része
Figure 4 Main data (coordinate, settlement, name, etc.) search interface of the GeoDat application

Az eredménytáblázat tetszőlegesen sorba rendezhető és exportálható lesz xls-formátumban. Megfelelő jogosultság esetén az adatok helyben javíthatók, valamint új adatok a felületen keresztül is bevihetők. A felületen kiadott minden módosítás naplózásra kerül, így a korábbi állapotok visszatekinthetőkké válnak.



5. ábra A GeoDat alkalmazás keresőfelületének szedimentológiai kereső rész
 Figure 5 Sedimentological search interface of the GeoDat application

The screenshot shows the search results for 'Karcag 573/18'. The window title is 'Szűrő beállítások | Mentett keresések szfz_csk1'. The table has columns for 'Térség', 'Fúrás jelle', 'Mérési sáv kezdete (m)', 'Mérési sáv vége (m)', 'Mérési sáv sorrend', and several size range columns: '0-0.002 mm [val%]', '0.005-0.01 mm [val%]', '0.01-0.02 mm [val%]', '0.02-0.06 mm [val%]', '0.06-0.1 mm [val%]', '0.1-0.2 mm [val%]', '0.2-0.5 mm [val%]', and '0.5-1.0 mm [val%]'. The data shows 15 measurement points for each size range.

Térség	Fúrás jelle	Mérési sáv kezdete (m)	Mérési sáv vége (m)	Mérési sáv sorrend	0-0.002 mm [val%]	0.005-0.01 mm [val%]	0.01-0.02 mm [val%]	0.02-0.06 mm [val%]	0.06-0.1 mm [val%]	0.1-0.2 mm [val%]	0.2-0.5 mm [val%]	0.5-1.0 mm [val%]	Mé
Karcag	573/18	0.00	0.30	1	13.50	14.90	18.50	49.70	2.00	0.30	0.80	0.80	
Karcag	573/18	0.30	0.80	2	13.50	17.00	24.70	29.70	0.50	0.50	0.50	0.80	
Karcag	573/18	0.80	0.90	3	6.50	17.20	26.50	34.70	0.50	0.50	0.50	1.00	
Karcag	573/18	0.90	1.50	4	13.70	18.50	20.90	29.70	0.30	0.30	1.00	1.00	
Karcag	573/18	1.50	2.00	5	11.50	18.30	21.90	31.30	0.50	0.50	1.00	1.00	
Karcag	573/18	2.00	3.00	6	9.50	19.10	24.50	33.50	0.50	0.50	1.00	1.00	
Karcag	573/18	3.00	4.00	7	11.90	18.50	22.70	30.20	0.50	0.50	1.00	1.00	
Karcag	573/18	4.00	5.00	8	12.10	19.10	22.70	24.30	1.00	3.00	1.50	1.50	
Karcag	573/18	5.00	6.00	9	9.70	12.50	20.50	38.30	1.00	0.50	4.50	4.50	
Karcag	573/18	6.00	7.00	10	15.20	18.70	25.50	25.30	0.30	1.00	1.00	1.00	
Karcag	573/18	7.00	8.00	11	10.70	15.10	20.60	25.30	1.00	0.50	1.00	1.00	
Karcag	573/18	8.00	8.50	12	12.90	17.80	20.30	31.30	0.30	0.50	0.80	0.80	
Karcag	573/18	8.50	9.00	13	7.70	19.50	23.10	31.50	2.00	0.50	1.00	1.00	
Karcag	573/18	9.00	9.50	14	8.50	16.00	28.20	33.70	1.50	0.50	1.00	1.00	
Karcag	573/18	9.50	10.00	15	11.90	17.20	18.90	38.10	1.00	1.50	1.50	1.00	

6. ábra Találati eredmény: Karcag 573/18 jelű fúrás szedimentológiai adatai
 Figure 6 A search result: sedimentological data of the shallow drilling of „Karcag 573/18”

Továbblépés

A továbblépés három irányba lehetséges. Az első a használhatóságra, adattartalomra vonatkozó felhasználói visszacsatolás, a második az adatbázis gazdagodása esetleges új fúrási adatok révén, míg a harmadik, rövid távon mindenképp a legnagyobb időigényű tétel a még hiányzó archív adatok lehetőség szerinti pontosítása, pótlása a papír alapú jegyzőkönyvek, térképek alapján. A felsorolt irányokból érzékelhető, hogy az adatbázis

pusztán adatgazdai oldalról korlátok között fejleszhető csak. A szerzők nem titkolt célja az volt, hogy tanulmányukkal felhívják a majdani felhasználók figyelmét tapasztalataik megosztására annak érdekében, hogy az adatbázis, illetve a kezelőfelület felhasználóbarát módon, gazdag, jól kereshető tartalommal feltöltve álljon rendelkezésre.

Köszönetnyilvánítás

Jelen revideálási folyamathoz nélkülözhetetlen segítséget nyújtott KUTI LÁSZLÓ és MÜLLER TAMÁS, akik számos, ma már nehezen nyomozható bejegyzést azonosítottak.

HALUPKA GÁBOR
Magyar Földtani és Geofizikai Intézet, Budapest
halupka.gabor@mfgi.hu

OROSZ LÁSZLÓ
Magyar Földtani és Geofizikai Intézet, Budapest
orosz.laszlo@mfgi.hu

IRODALOM

- GYALOG L.–OROSZ L.–SÍPOS A.–TURCZI G. 2005: A Magyar Állami Földtani Intézet egységes földtani jelkulcsa, fűrési adatbázisa és webes lekérdező felületük. – Magyar Állami Földtani Intézet Évi Jelentése, 2004. pp. 109–124.
- KUTI L. (szerk.) 2009: Agrogeológia. – Dura Stúdió, Budapest. 98 p.
- MÜLLER T. 2006: Környezetföldtani adatbázis, v. 1.01. Részletes kézikönyv. – Kézirat, MÁFI.
- RÓNAI A. 1985: Az Alföld negyedidőszaki földtana. – Geologica Hungarica, 21. 446 p.