

# **Térinformatika 6.**

## **A térinformatikai adatbázis hibái**

**Végső, Ferenc**

---

## **Térinformatika 6.: A térinformatikai adatbázis hibái**

Végső, Ferenc

Lektor: Detrekői, Ákos

Ez a modul a TÁMOP - 4.1.2-08/1/A-2009-0027 „Tananyagfejlesztéssel a GEO-ért” projekt keretében készült. A projektet az Európai Unió és a Magyar Állam 44 706 488 Ft összegben támogatta.

v 1.0

Publication date 2010

Szerzői jog © 2010 Nyugat-magyarországi Egyetem Geoinformatikai Kar

### **Kivonat**

A térinformatikai elemzés eredményének megbízhatósága sarkalatos kérdés. Ezért tisztában kell lennünk a térinformatikai adatbázis megbízhatóságával és hibáival. Ismernünk kell minden lépés lehetséges hibaforrásait: térinformatikai modell, az adatgyűjtés hibái, az adatintegráció hibái, topológiai hibák, az elemzés hibái (ez utóbbival a második félévben foglalkozunk részletesen). Ki kell térni egy ritkán taglalt problémára is: az idő múlásának hatása a térinformatikai adatbázisra.

Jelen szellemi terméket a szerzői jogról szóló 1999. évi LXXVI. törvény védi. Egészének vagy részeinek másolása, felhasználás kizárólag a szerző írásos engedélyével lehetséges.

---

# Tartalom

6. A térinformatikai adatbázis hibái .....	1
1. 6.1 Bevezetés .....	1
2. 6.2 Adatminőség a térinformatikában .....	1
2.1. 6.2.1 A hiba fogalomköre .....	2
3. 6.3 A hibák és a GIS .....	5
3.1. 6.3.1 Belső GIS hibák – a fejlesztők problémái .....	6
3.2. 6.3.2 Külső GIS hibák – az alkalmazás problémái .....	6
3.3. 6.3.3 A hibák és a GIS használata .....	7
4. 6.4 A GIS hibáinak típusai .....	8
4.1. 6.4.1 Az adatintegráció hibái .....	9
4.2. 6.4.2 A valós világ modellezésének hibái .....	11
4.3. 6.4.3 Az egyedek definiálása .....	12
4.4. 6.4.4 A méretarány hatása az egyed definiálására .....	12
4.5. 6.4.5 Mintavételi és mérési hibák .....	13
5. 6.5 Az adatok előkészítése .....	13
5.1. 6.5.1 Az adatgyűjtés .....	13
5.1.1. 6.5.1.1 Manuális digitalizálás .....	13
5.1.2. 6.5.1.2 Automatikus digitalizálás .....	14
5.1.3. 6.5.1.3 Vektor – raszter átalakítás .....	14
5.1.4. 6.5.1.4 Raszter - vektor átalakítás .....	18
5.1.5. 6.5.1.5 A címkézési hibák .....	19
5.2. 6.5.2 Adatbázisok konvertálása és a kódolás .....	20
5.3. 6.5.3 Adatintegráció .....	21
5.3.1. 6.5.3.1 A geometria tisztítása és szerkesztése .....	21
5.3.2. 6.5.3.2 Generalizálás .....	22
6. 6.6 Az adatok származása .....	22
6.1. 6.6.1 Miért van szükség az adatok származásának követésére? .....	23
6.2. 6.6.2 Mit kell tartalmaznia a származási dokumentumnak? .....	24
7. 6.7 Összefoglalás .....	24



---

# 6. fejezet - A térinformatikai adatbázis hibái

## 1. 6.1 Bevezetés

Ha a térinformatika használata során adatot gyűjtünk, feldolgozzuk azokat vagy végterméket (jelentést) állítunk össze, tisztában kell lennünk a lehetséges hibaforrásokkal és a hibák hatásával a végtermékre vonatkozóan. Képesnek kell lennünk az elkészült anyag minőségének megállapítására. A probléma az, hogy nem mindenki érti ugyanazt a hiba és a minőség fogalma alatt.

## 2. 6.2 Adatminőség a térinformatikában

### Definíciók

**Minőség:** tulajdonképpen a használatra való alkalmasságot jelenti. A mi esetünkben azt tükrözi, hogy az adatok, a térkép vagy a GIS alkalmazás mennyire felel meg az azt megítélő személy – általában a felhasználó – igényeinek. Az adatbázis minőségét egyaránt megítélheti a létrehozó vagy a felhasználó. Tudnunk kell ugyanakkor, hogy a minőség követelményei igen eltérőek lehetnek.

Az adat létrehozója a feltételezett használat összefüggésében definiálja az adatminőséget. Vegyük példának az állami földmérést. Az állami földmérés felelős az egész ország térképezéséért bizonyos felbontással és tartalommal (azért nem mondunk itt méretarányt, mert az egyre jobban terjedő digitális térképi adatbázisoknál nem értelmezhető). Az állami földmérés által létrehozott térképekkel szemben a minőségi követelmények: legyen napra kész, legyen ismert és mérhető pontossága. A minőség kérdéskörében az is előfordul, hogy egy adat-előállító szervezet belső célra más adatminőségi fogalmakat használ, mint a külvilág. Például az állami földmérés alkalmazza belső konzisztencia, átfutási idő vagy a megtérülés fogalmát a minőség definiálásában. Az állami földmérés térképeinek felhasználói másként értelmezik a minőséget. A térképi adatbázis naprakésztsége sokkal fontosabb lehet a mentéssel foglalkozóknak (polgári védelem, katasztrófa elhárítás, tűzoltók stb.), mint például az önkormányzati vagyon kezelőinek. A fenti példa is mutatja, hogy milyen nehéz a minőséget definiálni akkor, amikor a felhasználók egészen eltérő elvárásokat támasztanak. A mentéssel foglalkozók esetében a térkép hibája vagy gyenge minősége emberéletekbe kerülhet, míg az ingatlanfejlesztők esetében legfeljebb a fejlesztés költségeit növelheti. A felhasználók minőség felfogása gyakran nem egyezik az adatszolgáltatók minőség felfogásával. Az adatszolgáltatók alapjában véve a felbontást tekintik az egyik fő minőségi mutatónak. Ezzel szemben, a felhasználók fontos minőségi mutatónak tekintik az adatbázis teljességét. Gondoljunk arra, hogy egy település lakosságát az adózók listájából próbáljuk összeadni. Valószínűleg nem fog sikerülni, pedig elvileg mindenki fizet adót. A minőség jelentheti továbbá az adat alkalmazhatóságát a GIS elemzésben. Például van egy műholdas felvételtől levezetett – amúgy korrekt – vegetáció térképünk 1:25 000 méretarányban. Ha valaki nagy méretarányú vektoros alkalmazást készít, annak ez az adat használhatatlan lesz.

Röviden, a minőség megítélésével kapcsolatban két kérdésre kell válaszolnunk:

- az adat hogyan használható fel az én problémám megoldására?
- hogyan igazítható a feladat a meglévő adatokhoz?

**Hiba:** nem más, mint az eltérés a meglévő adat és az igazi adat között. A hiba a minőség fő kérdése. A hiba fogalmát kiterjesztik mindazokra a hatásokra, amelyek a meglévő adatot az igazi adattól eltérítik. A hibát általában egyszerű eltérésként értelmezik. A gyakorlatban sajnos a hiba fogalmát eléggé lazán használják. Célszerű a hiba fogalmát több oldalról is megközelíteni. Az egyik szempont lehet a hiba típusa, a másik szempont lehet a hiba forrása. Mindkét megközelítésnek megvan a maga gyakorlati haszna. Az első megközelítés a hiba felderítését szolgálja, a második pedig a csökkentését vagy a kiküszöbölését. Az alábbiakban megemlítünk néhány hibaforrást:

Emberi hiba – adatrögzítő személy hibája – digitalizálási hiba

Szoftver hiba – programozási hiba – adatmező vesztes

## GIS hiba – szoftver hiba – hibás interpolációs algoritmus

Minden GIS művelet az adatmodell megfogalmazásától kezdve az adatfeldolgozáson át a végtermék elkészítéséig hibákat hozhat létre, és a meglévőket tartalmazhatja, tovább viheti. A felhasználó kiindulhat hibás adatokból, és ezeket más hibás adatokkal kombinálva még több hibát hozhat létre. Az adat kezdeti hibája tovább terjedhet olyan adatokra, amelyek a kiinduló adatot tartalmazzák (hibaterjedés). Az adatokból levezetett információ használhatóságát a benne maradt hibák erősen csökkenthetik. A fentiek banálisnak tűnhetnek, de gondoljunk arra, hogy a számítógépek hatékonysága hibás adatok és információk előállításában is ugyanolyan magas, mint a használható információk előállításában. Ráadásul a számítógépek nagyon látványos és meggyőző tévedéseket tudnak generálni (pl. másodpercek alatt tudunk teljesen rossz de színes térképeket előállítani).

Mindez a GIS felhasználó számára azt jelenti, hogy a folyamatba nehezen követhető és felismerhető hibák épülhetnek be, amelyek a levezetett információk használhatóságát csökkentik.

A minőség-ellenőrzés és a hibafelismerés minden GIS tevékenység alapvető része, ennek ellenére gyakran mellőzik vagy nem is végzik el. Ennek fő oka, hogy az adatminőség javítása minden esetben költségnövekedést jelent, mégpedig nem lineáris összefüggésben. A GIS használóinak gyakran dönteniük kell, hogy az aktuális helyzetben a minőség vagy a költség fontosabb –e. A kicsi piacokon az ár (költség) gyakran a piacnyerés feltétele és ez rejti a csapdát. Ha ellenőrzés nélkül használunk fel adatokat, végzetes hibát követhetünk el. Ha elemezzük a hibákat, nem marad időnk és pénzünk az adatok felhasználására.

## 2.1. 6.2.1 A hiba fogalomköre

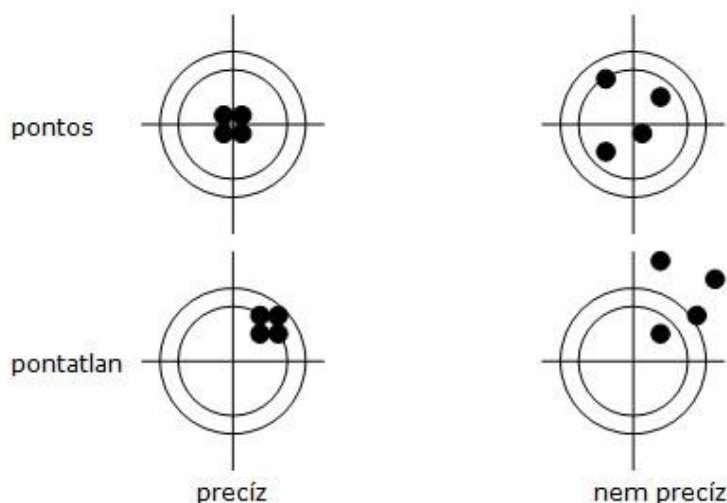
A hiba fogalomkörének és típusainak ismerete rendkívül fontos, mert a különböző szakmák képviselői ezen a területen egy nyelvet kell, hogy beszéljenek. Az alábbi hibatípusokat érdemes közelebbről körüljárni térinformatikai szempontból:

- pontosság
- élesség
- méretarány megválasztásából adódó hiba
- zaj
- teljesség
- időbeli teljesség
- logikai teljesség
- fogalmi (jelentéstani) pontosság
- ismételhetőség

**Pontosság:** a pontosság lehet egy becsült érték, amellyel megközelítjük a valódi értéket (méréseknél ez például nem ismert), más megfogalmazással a zajtól való mentesség foka. A pontosságot úgy is szokták jellemezni, mint egy külső igazsággal való egyezés fokát, amely „igazság” lehet egy definiált állapot vagy számérték. Az adatok esetében ezt arra szokták érteni, mennyire jeleníti meg az adat a valóságot. Az persze, hogy mi az igaz, relatív fogalom. Vegyük példának az idő és a méretarány valódiságát olyan térkép esetén, amely egy nagyváros határát mutatja tíz évvel ezelőtti állapotban. A térképet pontosnak nevezhetjük, hiszen megfelelően ábrázolja a településhatárt a megadott időpontban. Tegyük fel, hogy a szóban forgó térképet a következő tíz évben térinformatikai elemzés céljára használjuk. Ahogy telnek az évek, a térkép (látszólag) egyre pontatlanabb lesz, mert a város határa folyamatosan változik. Ugyanakkor bizonyos feladatok esetében ez a változás nem okozza a pontatlanság érzését. Például 1:100 000 méretarányának megfelelő felbontásban dolgozva, a az említett „pontatlanság” még évtizedekig elfogadható. Itt jegyezzük meg, hogy a helyzeti pontosság és az alak (vagyis az objektum valódi alakjának felismerése) pontossága két különböző kérdés.

**Élesség:** az élesség (precizitás) a részletesség kifejezése. A 4.123456 km-es távolság élesebben van megadva, mint a 4 km. Sajnos sokszor az élesség fogalmát gyakran használják a pontosság helyett. Sokan gondolják, hogy ami élesebben van megadva, az pontosabb is. Ez a káros jelenség főleg a statisztikai gyakorlat és a számítógép használat miatt alakult ki. Gondoljunk arra, hogy amíg manuális módszerek voltak, a mérnökök mindent csak az

optimális élességgel ábrázoltak, míg a szoftvereknél külön le kell tiltani a „végtelen” élességű kijelzést. A fentiek szemléltetésére nagyon jó a célba lövés példája.



A legtöbb gyakorlati alkalmazásban a felhasználó igyekszik optimalizálni a pontosság és az élesség értékét. Ha mindkettőt fokozzuk, nőnek a költségek, ha csak az egyiket, az nem jelenti automatikusan a másik komponens javulását.

**Méretarány és felbontás:** a gyakorlati életben a pontosság és az élesség összekapcsolódik a felbontás fogalmával. A felbontás az egyed legkisebb mérete, amelyet megjeleníthetünk, vagy térképezhetünk. Például 30 méteres felbontással rendelkező úrfelvételen megkülönböztethetjük a lakott és a mezőgazdasági területeket az eltérő visszaverő képesség miatt. Ugyanakkor az egyes fákat nem tudjuk elkülöníteni, mert a műhold érzékelője a pixelen belül integrálja a spektrális visszaverődést. A raszter adatok felbontása egyértelműen definiálható, de a vektoros adatoknak is értelmezhető a felbontása. Ez függ az eredeti térkép méretarányától, az alkalmazott vonalvastagságtól, a digitalizáló tábla minőségétől. Ugyancsak figyelembe kell venni a szkennelés felbontását, ha a vektoros állományt képernyőről digitalizáljuk. Ez az érték általában jobb mint 100 dpi, vagyis a szem felbontóképessége alatt van. Ez általában nem is probléma, mert a vektor adatok rögzítésének pontossága általában messze felülmúlja az átlagos felhasználói igényeket. Az ilyen típusú adatgyűjtés kritikus pontja valószínűleg a digitalizáló személy szemének felbontása és a kéz bizonytalan mozgása, ami legjobb esetben is csak kb. 0.25mm pontosságot tesz lehetővé.

**Zaj:** az adatok esetében a zaj eltérést jelent a normális vagy valódi értéktől. A zaj a legtöbb esetben külső forrásból ered. Például egy digitalizáló személy hajlamos a rajzolt vonal hosszánál rövidebbet digitalizálni. A rosszul kalibrált digitalizáló tábla minden pont helyzetét képes torzítani. Ezek a szubjektív vagy gépi úton keletkezett zajok. A szoftvereknél – programozási hiba miatt – előfordulhat szisztematikus lefelé kerekítés, állandóan alábecsülve térbeli adatokat vagy akár költségeket. Ehhez hasonlóan a raszteres adatokon végzett területszámítási algoritmusok következetesen alábecsülik a valós területet. Ezek a módszerekből eredő zajok. A zajt nehéz tetten érni, mert jelennek meg nyilvánvaló eltérésként. Csak akkor deríthetők fel, ha össze tudjuk hasonlítani a valós értékkel, vagy az adatfeldolgozó rendszer (ember, gép, módszer) elemeit változtatjuk, illetve kicseréljük. A zajokat élesen el kell különítenünk a véletlen hibáktól. A véletlen hiba az összes ismeretlen hiba összegződésének a hatása. Sajnos ezeket sokan a rendszer „zajaként” értelmezik. Pedig a hibaelmélet szerint a véletlen hibák összege a mérhető mennyiségek esetében a nullához tart és mértéke becsülhető a szórással. A személyi hibák, zajok mértéke becsülhető, de általában sem a mértékük és az előjelük nem jelezhető előre.

### Teljesség:

a teljességet két értelemben szoktuk használni:

- egy adott földrajzi terület minden mérhető részére kiterjedő teljesség
- minden olyan időközre kiterjedő teljesség, amikor az adat megmérhető

A teljesség megítélése nehéz és néha szubjektív dolog. Ahol az adatok az egyedek attribútumaiból állnak, a teljesség a különbözőséget jelenti azoktól az egyedektől, amelyeknek nincs attribútumuk (leíró adatuk). A

hiányzó attribútumok felismerése azonban sokszor nehézségeket okozhat, kivéve, ahol tudjuk, hogy minden egyednek kell, hogy legyen attribútuma (pl. minden magyar földrészletnek van helyrajzi száma). Egyéb esetekben – mint a többi hiba esetében – szükségünk van független információkra. A teljességet könnyen megállapíthatjuk a szabályos vektor vagy raszter rácsok esetében. A mintavétel azonban – akár szabályos akár véletlen elrendezésben történik – csak hiányos lehet. Gondoljunk például a domborzatfelmérésre. Ezekben a szituációkban a teljesség fogalma megegyezik a megfelelő mintavételezés fogalmával (mikor mértünk elegendő számban és helyen magasságot a domborzat megfelelő reprodukálásához?). A megfelelőség kérdése a statisztikai elmélethez tartozik, ezért itt nem tárgyaljuk.

Az időfüggő adatok esetében nehéz a teljesség definiálása. Általában az időhöz kapcsolt adatok az idő egy adott – jól elkülöníthető – pontjához kapcsolódnak (pl. dátum). Mivel az idő folyamatos, az adatok nem vonatkozhatnak időintervallumra. Ezért minden mérés, térkép bármely időintervallumban szükségszerűen nem lehet teljes, ami más szóval az állandó változást tükrözi. A népesség változása tíz éves periódusban csak a változás természetét tükrözi, de nem mond semmit a két népszámlálás közötti időpontok népességéről. Az új felméréssel készült térképek tulajdonképpen az elkészítés vége előtt elvesztik teljességüket.

**Időbeli teljesség:** az időbeli teljesség fogalma alatt az adatfeldolgozás ismétlődő elemeinek konzisztenciáját értjük. Elemek alatt az adatgyűjtést, az adattárolást, adatfeldolgozást, elemzést, megjelenítést és értelmezést értünk. A konzisztencia azt jelenti ebben az esetben, hogy az idő előrehaladásának ellenére a rétegeken és azok között ugyanazokat a műveleteket végezzük. Példaként felhozhatjuk a közigazgatási határokat. Ezek a határok – településhatár, választási körzet határa, beiskolázási körzethatár stb. – állandóan változnak. Ez sok problémát okoz az időbeli teljesség fenntartása tekintetében. Az időbeli teljesség fenntartását akadályozhatja az adatgyűjtési módok megváltoztatása is. Itt sok éves vagy sok évtizedes változásokra kell gondolni.

**Logikai teljesség:** lényegében azt jelenti, hogy az adatoknak és a műveleteknek logikailag összeegyeztethetőnek kell lenniük. Ez fajta teljesség sok problémát vet fel, az egyik legfontosabb az adatok megfelelősége a végrehajtani kívánt művelet vagy elemzés számára. Például nincs értelme kiszámolni az utak rendűségének (elsődrendű, másodrendű, autópálya, autópálya stb.) középértékét, vagy elemezni a lejtőkategória és a választók pártállásának összefüggését. A másik fontos probléma az adat alkalmassága a feladat megoldására. Sarkítva: bárhol dolgozzuk fel a talajminőség adatait, nem tudunk belőle a városi lakosság bevételeire nézve használható következtetéseket levonni. A valódi térinformatikai feladatokban ez a probléma persze nehezebben megfogható módon jelentkezik. Ugyanis hiányzó vagy nem megfelelő adatokat közvetett módon is pótolhatunk, hiszen a magas vagy alacsony bevételekre nem csak a kifizetett jövedelmekből lehet következtetni. Ez a módszer nem feltétlenül helytelen. Arra kell figyelni, hogy hol a határa az adatok helyettesíthetőségének. Ez mindig alkalmazásfüggő és a választ nem annyira a térinformatikus, hanem a terület szakértője tudja megadni. Végül foglalkoznunk kell az azonos földrajzi területre vagy azonos időszakra vonatkozó adatok megfelelőségi szintjének problémájával. Az adatok lehetnek elfogadhatóak, ugyanakkor nem megfelelőek. Az 1988-as földhasználati adatok és a 2001 évi mezőgazdasági bevételi adatok lehetnek helyesek, de nem elfogadhatóak egy adott elemzés számára, hiszen más időpontra vonatkoznak.

**Jelentéstani helyesség:** a jelentéstani helyesség annak a fokmérője, hogy a kiválasztott térinformatikai modellhez képest az alkalmazott egyedek tulajdonságait, kapcsolataikat és leíró adataikat (attribútumaikat) mennyire helyesen tudjuk megjeleníteni. A probléma abban rejlik, hogy a valóság elmeit általában földrajzi objektumként látjuk, és nem úgy, mint a valóság geometriai megfelelőjét.

**Ismételhetőség:** az ismételhetőség abban nyilvánul meg, hogy független felhasználó mennyire tudnak ugyanolyan adatokat vagy eredményeket előállítani. Az ismételhetőség fontos tényezője a megbízható eredmények előállításának. Az ismételhetőség általában feltételezett tulajdonság. Az adatok előállítása például túl költséges ahhoz, hogy kétszer végezzük el az adatgyűjtést. El kell hinnünk például, hogy a földmérési alaptérkép alapján végzett elemzés többször is ugyanarra az eredményre vezet. Az állami földmérést jogszabály és belső utasítások, szabályzatok kötelezik a megbízható adatok előállítására. Ha vízművek azt állítja, hogy az ötven évvel ezelőtt lefektetett vezetékek helyesen lettek bemérve és ábrázolva, optimális esetben elhiszük anélkül, hogy ma újra bemérnénk őket. Néhány térbeli (földrajzi) adat esetében nem tudjuk felmérni az ismételhetőség korlátait. Nézzük például a talajtérképeket. Ezeket is állami keretek között készítik, Magyarországon a Növény – és Talajvédelmi Szolgálat a Földhivatalokkal karöltve, szabványok és szabályzatok alapján. Mégis, az eredmény szubjektív és ezért nehezen ismételhető. Ennek az a magyarázata, hogy a formai követelményeken belül a talaj minősítése egy szakember feladata. A talaj típusa vagy egy talajtípus határa azonban sosem jelenik meg „vegytisztán” a természetben. Ezért ha ugyanazt a területet ketten dolgozzák fel egymástól függetlenül, az eredmény nem lesz ugyanaz. Ráadásul a talajtérképek méretaránya sokkal kisebb, mint a mintavételkor használt alaptérképeké, ezért generalizálásra is sor kerül, ami az ismételhetőséget tovább rontja. Röviden, nagyon sok adat, amivel találkozunk nem közvetlenül mért adat, hanem szakértők által



előállított adat. Ezek a szakértőt kivéve lényegében nem reprodukálhatók. Sokszor a szakértők által használt algoritmusok és módszerek sem nyilvánosak.

Az ismételhetőséget érthetjük a műveletek követhetőségére is. A követhetőség megadja az adaton végrehajtott műveletek és feladatok történetét. A követhetőség lehetőséget ad a felhasználónak megismerni, hogy mi minden történt az adattal amíg egyik állapotából a másikba eljutott. Ha egy technológiába beépítjük a követhetőséget, akkor egy másik felhasználó meg tudja ismételni a műveleteket, és véleményt tud formálni az elemzés minőségéről. A követhetőség egyszerűen is megoldható: egy papírra vagy egy szöveges állományba leírjuk, hogy milyen műveleteket hajtottunk végre. A mai GIS szoftverek általában tartalmaznak egy olyan funkciót, amely automatikusan rögzíti egy állományban a felhasználó által végrehajtott parancsokat (history vagy log fájl).

**A hibák és az értékkála:** nem fordul elő az eddig felsorolt mindegyik hiba az adatokban. Az előforduló hibák típusa függ az értékkála típusától. Minden érték a következő skálatípusok valamelyikében jelenik meg: névleges (fiktív), sorrend, értékköz és arány. Az értékkála legalsó szintje a névleges kategória, ahol az egyedek osztályokhoz – pl. adott megnevezésekhez – vannak rendelve. A következő a sorrend, ahol az egyedek tulajdonságuk alapján egymáshoz képest sorba vannak rendezve. Az értékköz és az arány esetében az egyedek tulajdonságaik alapján értékhez vagy számskálához vannak kötve. Mint látni fogjuk, néhány értékkálának nincs valódi nulla értéke, ami korlátozza a felhasználásukat néhány művelet típus esetében. Az eddig leírtakat szemlélteti az alábbi táblázat néhány példán keresztül:

Értékkála	Példa	Tulajdonságok
NÉVLEGES	Talajtípus	Az egyedek egyenrangúak, gyakoriság számítható.
	Közigazgatási egység neve	
	Tulajdonos neve	
SORREND	Életkor	Egyik kisebb, mint a másik, középérték számítható
ÉRTÉKKÖZ	Dátum	A mértékegység ugyanaz, számok összeadhatók és középérték számítható. Szórás számítható
	Hőmérséklet	
ARÁNY	Hossz	A számsornak van valódi nullapontja, arányok számíthatók
	Feszültség	
	Magasság	
	Népszerűség	

A táblázatban a fő határvonal az első két és a második két kategória között van. Az első két kategória értékeit nem vethetjük alá számtani műveleteknek, mint összeadás, kivonás, szorzás és osztás. A vízminőségi mutatókat nincs értelme összeadni vagy osztani. A tulajdonosok vagy a talajtípusok középértékének kiszámítása ugyancsak értelmetlen, tehát fel kell ismernünk, hogy az értékkála bizonyos elmeire akkor sem alkalmazhatunk számtani műveletet, ha azt számértékek formájában tároljuk (pl. személyazonosító számok összeadása).

### 3. 6.3 A hibák és a GIS

A GIS szereplői két nagy csoportra oszthatók: felhasználók és létrehozók. A GIS felhasználói a problémáikat szeretnék megoldani a térinformatika segítségével, és nem érdekli őket vagy nincs energiájuk utána járni, hol jelenhetnek meg hibák, és mi biztosítja az adatok jó minőségét. Az alkalmazás használója csak azt szeretné tudni, mit kezdhet a hibákkal. A GIS létrehozói, fejlesztői másként közelítenek a hibákhoz. Ők szintén problémákat akarnak megoldani a térinformatika segítségével, de sokkal jobban érdekli őket a hibák eredete és a hibák megelőzésének módjai. A GIS folyamatának állomásai során megjelenő hibák létrejöttének megértéséhez el kell mélyedni az algoritmusok, az adatformátumok és a hibák becsléséhez, kezeléséhez szükséges statisztikai eszközök rejtelmében. A legtöbb esetben arra koncentrálnak, hogy az adott alkalmazás mennyire érzékeny a hibákra. Szükséges tehát megérteni a hibák alapjait és azt, hogyan közelítsük meg a hiba problematikáját az alkalmazásokon belül és a hibák halmozódása területén. Első megközelítésben hasznos a hibákat két forrásra visszavezetni:

- **belső GIS hibák**
- **külső GIS hibák**

### **3.1. 6.3.1 Belső GIS hibák – a fejlesztők problémái**

A belső GIS hibáknak azokat tekintjük, amelyek a GIS használatából adódnak és a felhasználó által kevéssé vagy egyáltalán nem befolyásolhatók. Például a hulladék poligonok létrejötte, amelyek a poligon fedvények átlapolásakor keletkeznek, a felhasználó által nem szabályozható hibák. A felhasználó megpróbálhatja a poligonokat figyelmesebben digitalizálni, a hulladék poligonokat kiküszöbölni mégsem tudja. Az ilyen hibák elemzése elsősorban a fejlesztők és nem a felhasználók feladata. Ettől függetlenül jó, ha a felhasználónak van tudomása az ilyen típusú hibák létezéséről és szokásos nagyságáról. Nem szükséges tudni a kiküszöbölés módját, mert a mindennapi munka során ez nem oldható meg, illetve túl sok munkát okozna a hiba okozta bizonytalansághoz képest. A GIS által okozott belső hibák ismertek és előre jelezhetők. Ezeket a hibákat tartalmaznia kell a GIS által létrehozott eredmények értékelésének, vagy bizonyos GIS műveletek előtt figyelmeztetni kell a felhasználót a hiba megjelenésének lehetőségére.

### **3.2. 6.3.2 Külső GIS hibák – az alkalmazás problémái**

A „külső” jelző ebben az esetben kissé félrevezető. A külső hiba is a GIS rendszeren belül fordul elő. Olyan hibára utal, amely akkor keletkezik, ha a felhasználó rossz adatot vagy GIS funkciót használ. Ebből eredően a felhasználó tudja befolyásolni a hiba hatását és nagyságát. A GIS – és általában a szoftverek – óvatos használatával minimalizálni lehet a figyelmetlen munkával okozott hibák hatását. A felhasználónak fontos tudnia, hogy a kezdeti kis külső hibák összeadódnak a belső hibákkal a hibahalmozódás során. A GIS szóhasználatban különbséget tesznek a rejtett hibák és a műveleti hibák között. Rejtett hibának tekintik az adatgyűjtés és az adatintegráció (adatfeldolgozás) során keletkező hibákat, műveleti hibának tekintik a GIS elemzés során keletkező hibákat. Említenek még „használati” hibát is, vagyis GIS által létrehozott adatok használatából eredő hibát. A GIS rendszerek létrehozhatnak egy sajátos, csak rájuk jellemző hibát. Ez a hiba a GIS rendszerek alapját képező egységes koordináta rendszer használatából ered. Mivel a GIS adatbázisban minden adat egy meghatározott helyhez (koordináta) kapcsolódik, megvan a lehetőség és az algoritmus az adatok koordináta alapú összekapcsolására. Olyan kombinációkat alkalmazhatunk ész és elővigyázat nélkül, amelyek darabolhatják, keverhetik az adatokat, vagy új adatokat állíthatnak elő a meglévőkből. Például minden probléma nélkül átfedésbe hozhatjuk a talajtípus, a választási körzet és az egyházmegyék poligonjait azon az alapon, hogy közös koordináta rendszerben vannak ábrázolva az adatbázisban. Ráadásul az adatok összekapcsolásakor megjelenik a hibák sokszorozódásának hatása. Képzeld el, hogy elkészítenek nekünk egy településtérképet három olyan térkép kombinációjából, amelyek mindegyikén van információ a településről. Nem tudjuk, vagy kevés információ van az eredeti térképekről és az eredeti adatok koráról. Az új, kompozit térképen minden egyed egyenrangúnak és egyforma idősnek látszik. A felhasználó azt fogja tapasztalni, hogy az új térképen a neki szükséges információk nagy része megvan, de találni fog sok egyedet, aminek nincs sok kapcsolata a valósággal.



1. ábra Zavaros térkép<sup>1</sup>

Ha a hibaterjedésre gondolunk, vagy a GIS technológiai folyamatára az adatgyűjtéstől a feldolgozáson át az eredmény megjelenéséig, fel kell tételeznünk, hogy az adatok, amelyekből építkezünk, hasonló folyamaton mentek át. Ehhez hasonlóan a GIS által előállított adatok forrásul szolgálhatnak további feldolgozáshoz. Ez a feldolgozási ciklus érvényes minden adatra, tehát ha például adatot gyűjtünk egy térképről a GIS számára, a felhasználó általában már nem vizsgálja a térképről örökölhető hibákat.

Ezek a hibák lehetnek:

- generalizálás
- helyzeti hibák
- attribútum hibák
- a méretarány okozta eltolás a vonalak helyzetében és szélességében

A térképet a GIS tisztán adatforrásnak tekinti, amely nem tartalmaz hibát. Ha GIS-t használunk, a hibák az adatforrások sokkal szélesebb köréből származhatnak: topologikus térképek, úrfelvételek, népszámlálás, légifényképek, mérési eredmények, fotogrammetria stb. Mindegyiknek megvan a maga sajátos hibája, amelyek hajlamosak halmozódni az adatintegráció során.

### 3.3. 6.3.3 A hibák és a GIS használata

A GIS használata közben keletkező hibák két nézőpontból közelíthetők meg: számítógép használat és GIS használat. Mi természetesen az utóbbira koncentrálnunk ebben a tantárgyban, de fontos tudnunk, hogy milyen hibák származhatnak magának a számítógépnek a használatából.

#### Számítógép használat

Ha GIS-re mint eszköze gondolunk, a számítógépet hajlamosak vagyunk gondolatban kihagyni a folyamatból. A baj csak az, hogy a GIS használata ma még nem olyan ösztönös tevékenység, mint az evés vagy a sétálás. Az emberek más szemlélettel közelítenek a számítógépen végrehajtott műveletekhez, mint a hagyományos, kézzel végrehajtott műveletekhez. Jó példa erre a rétegtechnikán alapuló kiválasztás példája. Képzeljük el, hogy több szempontnak megfelelő ideális területet kell kiválasztanunk tematikus térképek alapján. A hagyományos eljárásnál az egyes szempontoknak megfelelő területeket átlátszó anyagra rajzolták, ezeket egymásra fektették és kiválasztották a megfelelő területet. Ha GIS-t használunk, a folyamat a következő: elindítjuk a rétegelemző (overlay) parancsot, megadjuk a térképállományok nevét és várunk az eredményre. A két eljárás a kritikus pontokon nem különbözik. Ettől függetlenül a manuális eljárás során a felhasználó sokkal könnyebben fedezi fel a lehetséges hibákat. Ha például az egyik térképen lévő földrészlet a másik térképen lévő tóba esik, a felhasználó korrigálja a térkép hibáját és megismétli az eljárást. Ha számítógépes módszert használunk, ezt a hibát a parancs végrehajtása előtt nem látja a felhasználó (hiszen csak állományneveket kellett megadnia) és csak a végeredményből, közvetett úton jöhet rá a hibára. A számítógép úgy működik, mint egy fekete doboz.

<sup>1</sup><http://www.ultimatepearland.com/2009/10/confused-over-county-commission-precincts-city-read>

Ezért alakul ki az a téves nézet, hogy a számítógép által szolgáltatott eredmény hitelesnek és pontosnak tekinthető.

### A GIS használat

A számítógépből vagy a GIS –ből származó eredmények kritikátlan elfogadása felveti a kérdést, hogyan kezelhető ez a probléma? Nézzünk két példát erre.

**Első példa:** egy építési hatóság ügyintézőjének értesítenie kell az ügyfél szomszédjait a tervezett építkezésről, hogy élhessenek az ezzel kapcsolatos jogaikkal. Az ügyintéző a megfelelő paranccsal kikeresi az adatbázisból a szomszédok adatait és automatikusan generálja a levelet. Amit az ügyintéző nem tett meg: ellenőriznie kellett volna, hogy a címek között nincs-e logikailag téves (nem szomszéd). Ha egy földrészlet rosszul lett digitalizálva, rossz a topológiai (szomszédsági) információ, a valódi szomszéd nem kap értesítést és kimarad joggyakorlásból. Ez a példa elég gyakori a valóságban. A megoldás itt az, hogy olyan helyzetet teremtsünk az ügymenetben, ahol az eredményeinket össze tudjuk hasonlítani a valósággal.

**Második példa:** a GIS felhasználó feladata az, hogy elemezze egy Balaton parti településen a földrészletek és a part kapcsolatát. Ezért digitalizálja a topográfiai térképről (1:10 000) a Balaton partjának egy szakaszát, és a parton lévő város földrészleteit a földmérési alaptérképről (1:1 000). A probléma nyilvánvalóan a nagyon eltérő méretarányú térképek tartalmi felbontóképességének különbségéből ered. Ez a példa felveti az adatok elővigyázatos alkalmazásának kérdését. A GIS rendszer nem tájékoztatja, vagy figyelmezteti a felhasználót a durva méretarány különbségre és az integráció miatt bekövetkező súlyos pontatlanságra.

A fenti két példa szemlélteti azt a két megközelítési módot, amit a GIS felhasználónak át kell gondolni:

- az adat megfelelő a végrehajtani kívánt feladathoz?
- a végrehajtani kívánt feladat megfelelő a rendelkezésre álló adathoz?

A GIS használata során ezeket a kérdéseket olyan gyakorlatias kérdésekké kell átalakítanunk, amelynek az a célja, hogy megelőzzük a nem megfelelő adat vagy parancs használatát. Szükség van olyan módszertanra, amely kézben tartja a hibákat.

Tágabb értelemben a kérdések a következőképpen merülnek fel:

- megtudhatom-e, hogy a GIS mit csinál az adatokkal, és megtudhatom-e, hogy az eredmény a kívánt minőségű (mi fog történni)?
- mi történhet rosszul a GIS használata során és ennek milyen hatása lesz a kimenetre (mi történhetett)?
- hogyan tudom megszüntetni vagy ellenőrizni a hibákat (mit tehetek velük)?

A megfelelő válasz megadásához három feltételnek kell teljesülni:

- megérteni, hogy mi az adat és a GIS mit csinál az adatokkal
- megérteni a hibákat és azok hatásait
- módszert kialakítani a hibák és hatásaik felderítésére

## 4. 6.4 A GIS hibáinak típusai

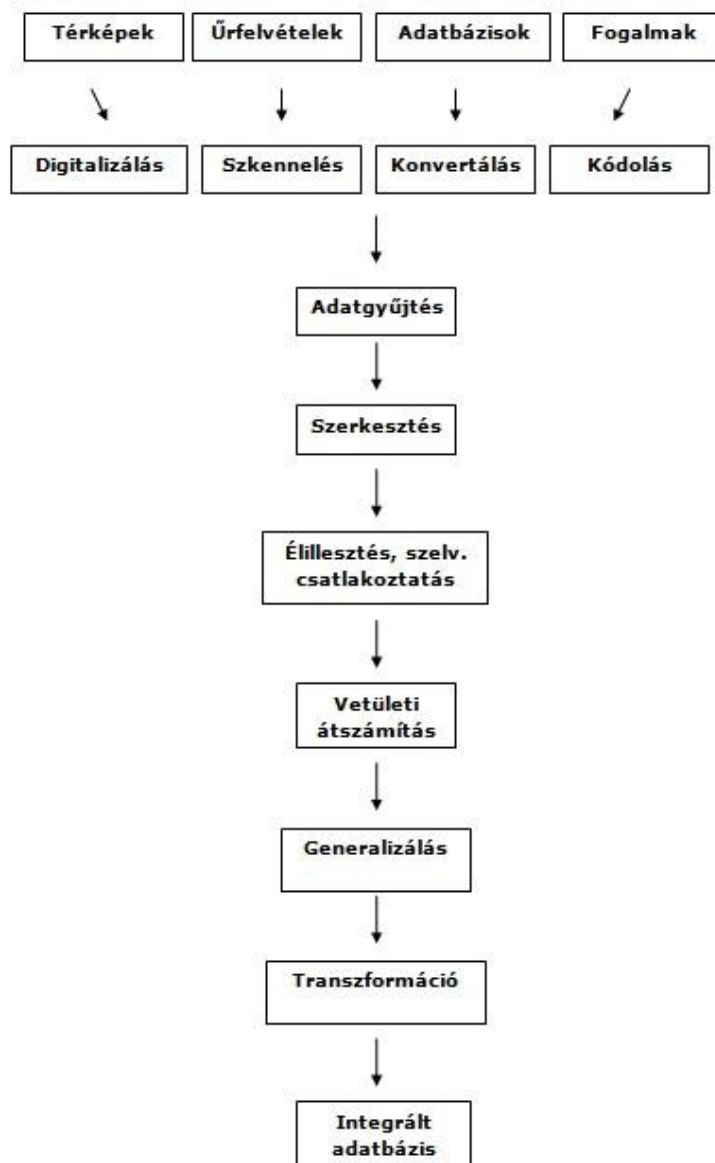
Eddig a hibák osztályozásával foglalkoztunk. Említettünk belső és külső hibákat, öröklött és műveleti hibákat. Megvizsgáltuk a hibákat a fejlesztők és a felhasználók szemszögéből. Most az osztályozáson túllépve megpróbáljuk definiálni a hibák természetét. Ebben a részben a hibák két fő típusát tárgyaljuk meg: a *helyzeti* hibákat és az *attribútum* hibákat. A *helyzeti* hiba alapvetően abból ered, hogy a térbeli egyedeket (pont, vonal, poligon) a térben rossz koordinátákkal ábrázoljuk. A hiba mértéke lehet kicsi vagy nagy, de a legtöbb esetben kis eltérésként jelenik meg az adatgyűjtés felbontásának méretarányában értelmezve. Az *attribútum* hiba azt jelenti, hogy a GIS adatbázisban az egyedeket helytelen leíró adattal jellemeztük. Például egy magyar helyrajzi szám nem lehet 145ac/6. A következőkben ezeknek a hibáknak a forrásait, okait és hatásait fogjuk megvizsgálni.

Egy másik megjelenési hibája a helyzeti és attribútum hibának a *fogalmi* vagy modellezési hiba. Ebben az esetben a hibák abból erednek, hogy különböző emberek különbözőképpen látják a valós világot. Másképp fogalmazva, a hiba oka a nem megfelelő típusú egyedek definiálása illetve a valóság nem megfelelő modellezése.

Ezeket a hibákat a legnehezebb megtalálni, és kezelni, viszont a legnagyobb hatással vannak a térinformatikai alkalmazás használhatóságára.

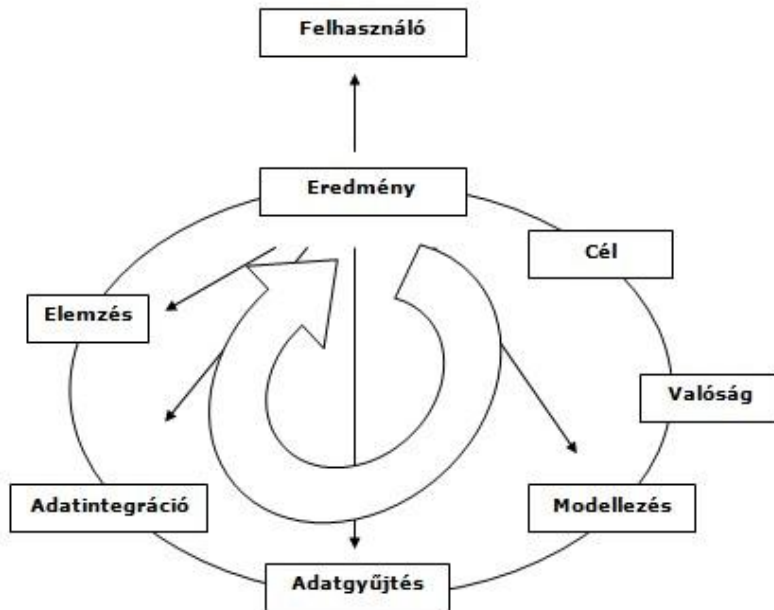
#### 4.1. 6.4.1 Az adatintegráció hibái

Az adatintegráció ebben az esetben azt a folyamatot jelenti, amelynek során kiindulunk a különböző adatformátumokból (térképek, űrfelvételek, adatbázisok, fogalmak stb.), végrehajtjuk az adat átalakítást (digitalizálás, szkennelés, adatformátum váltás, kódolás stb.), a szerkesztést (digitalizálás javítása, szelvénycsatlakoztatás, generalizálás stb.) és a végén eljutunk egy integrált, konzisztens digitális térképi adatbázishoz. Ezt a folyamatot szemlélteti a következő ábra.



Az ábrából az olvasható ki, hogy az adatintegráció egy egyirányú folyamat az adatforrásoktól az integrált adatbázisig. A hibák az integrációs folyamat bármely pontján megjelenhetnek. Ráadásul a hibák nem csak az adatintegráció során keletkezhetnek. Hibák lehetnek a forrásadatokban, felléphetnek az adatátalakítás során, megjelenhetnek a kimeneti oldal készítése során és a GIS-ben való felhasználás során is (elemzés). Ezért nem

elég csak az adatintegráció folyamatát ismerni, be kell ágyazni azt a GIS teljes folyamatába, amely a valóság modellezéséből kiindulva jut el egy felhasználónak szánt végeredméig. Csak ilyen felfogásban tudjuk felmérni a GIS használata során előforduló összes lehetséges hibaforrást. Az adatintegráció GIS folyamatba való beágyazódását mutatja a következő ábra.



Az ábráról leolvasható, hogy az adatáramlás az eredményhez kapcsolódó ciklikus folyamat, másrészt tartalmaz visszacsatolásokat a modellezéshez, az adatgyűjtéshez, az adatintegrációhoz és az elemzéshez. Ha kapott eredmény a felhasználónak valamilyen okból nem megfelelő, a visszacsatolás megváltoztathatja a felsorolt fázisok egy vagy több elemét mindaddig, amíg a felhasználó számára megfelelő eredmény nem születik.

Egy idealizált világban, amelyet a GIS eszközei teljesen lefednek, a valós világot helyettesíthetnénk a térinformatikai modellel és adatbázissal. Például ha a GIS-t egy zsúfolt város közlekedésének modellezésére használjuk, hiánytalanul megválaszolhatjuk a „Mi lenne ha?” típusú kérdéseket, amelyek valóban segítenek a közlekedési problémák megoldásában. Ez a szituáció egyben megvilágítja a probléma lényegét. Röviden arról van szó, hogy a világról alkotott fogalmaink csak töredékesek lehetnek, a világról gyűjtött adataink (térképek, könyvek, listák) nem lehetnek teljes körűek. Hibákat követünk el az adatgyűjtés során, az adatintegráció során. További hibákat követünk a módszerek, szoftverek, hardverek helytelen megválasztásakor. A GIS által szolgáltatott eredmény halmozottan tartalmazhatja a folyamat során elkövetett hibákat. Csak maga a valóság tekinthető tökéletesnek, illetve hibátlannak. Ezzel csak azt fejezzük ki, hogy ismereteink a világról töredékesek, nem is tudjuk biztosan mi a valóság. Ezért a hibák szintje és természete a GIS adatbázisokban és az eredményekben általában ismeretlen. A GIS sok felhasználója és alkalmazója nem tud róla, vagy nem néz szembe ezekkel a hibákkal. Mindez azt is jelenti, hogy lehetetlen a GIS segítségével a valós világot száz százalékos hűséggel leképezni. Ettől azonban nem kell elkeseredni, mert lehetőségünk van a hibaforrások felismerésére, a következményeik felmérésére, hatásuk minimalizálására vagy néhány esetben megelőzésükre. Továbbá képesek vagyunk az adataink és eredményeink minőségének becslésére és ez segít megítélni az adatok és eredmények használhatóságát.

Az eddig leírtak fényében a hibákról a GIS folyamat összefüggéseiben a következőket foglalhatjuk össze:

- sokféle hiba létezik, némelyik nyilvánvaló, némelyik nem. A két fő hiba, amire oda kell figyelnünk: a helyzeti hiba és az attribútum hiba. A többi hiba is fontos, de hatásuk többnyire a felsorolt két hibában nyilvánul meg.
- a hiba a GIS minden fázisában halmozódik. Ezért minden egyes GIS fázishoz tartozó hibát ismernünk kell, ha meg akarjuk tudni a végeredmény megbízhatóságát.

A továbbiakban sorra vesszük a GIS folyamatának állomásait és mindegyiknél megvizsgáljuk a lehetséges hibaforrásokat.

## 4.2. 6.4.2 A valós világ modellezésének hibái

A valós világ körülvesz minket és nagyon nehéz definiálni. A lexikonok definíciója szerint a valóság az, ami tény, ami történik, ami jelenleg van és nem utolsó sorban ami igaz. A valóságnak ezt a megközelítését a filozófusok használják, csak a megközelítési mód változó (pozitivizmus, realizmus). Más filozófusok egészen más fényben látják a valóságot. Az egzisztencialisták szerint az általunk érzékelt világ nem létezik, csak tudatunk terméke. Ezek a példák jól szemléltetik a világ szemléletének különböző módjait. A térinformatika számára a világot leginkább úgy képzeljük el, mint egy valóságot, amelyet hallunk, látunk, tapinthatunk magunk körül.

A valós világ a GIS-nek egyetlen állomása, amely nem tartalmaz hibát. Az első hiba amit beviszünk a GIS-be, az a mód, ahogyan a világot szemléljük, illetve koncepciót alakítunk ki róla. Ezek a hibák akkor merülnek fel, amikor leegyszerűsítés (absztrakció) útján megalkotjuk a saját modellünket, részekre bontjuk a valóságot, vagy csak egyszerűen másként látjuk a világot, mint mások. A modell megalkotását ráadásul komplikálják azok a lehetőségek, amelyek a modell megalkotásakor rendelkezésünkre állnak (meglévő térképek, adatbázisok). Általánosságban is elmondhatjuk, hogy napjainkban az informatika (hardver, szoftver, adatbázis kezelés) lehetőségei nagyon korlátozottak a valós világ hű leképezésére.

Feltételezhetjük, hogy a valóság annyira összetett, hogy lehetetlen teljes részletességgel modellezni és leírni. A második modulban megismertük a valóság részleteinek modellezésének módjait vektorokkal, raszterrel, rétegszemlélettel vagy objektum orientált megközelítéssel. Mivel a világ nagyon bonyolult, kénytelenek vagyunk leegyszerűsíteni, hogy beszélni tudjunk róla. Az ember sokszor fizikai modellt készít, mint például egy város makettje vagy a vasútmodell. A GIS-ben elsősorban fogalmi modellt és matematikai modellt használunk a valóság, a valóság elemei és a közöttük fennálló kapcsolatok szemléltetésére. A modell „a valóság leegyszerűsített struktúrája amely feltehetően fontos egyedeket és kapcsolataikat tartalmazza generalizált formában” (Chorley and Hagget, 1965). A modellezés legismertebb formái a földgömb, a térképek, rajzok. A modellekkel az a gond, hogy amennyire jók a valóság leírására, annyira sok hibát tartalmaznak. Mindenki ismeri az iskolákban található gömbre ragasztott papírtérképet, ez a földgömb. Ezt elfogadjuk a Föld modelljéül, bár tudjuk, hogy a Föld nem gömb alakú és a felszíne nem sima, hanem mély óceánok és magas hegyek tarkítják. Elmondhatjuk, hogy a földgömb, mint a világ modellje hibákat tartalmaz.

A világ modellje sok szintből állhat és különböző mennyiségű hibát tartalmazhat. Fontos megjegyezni, hogy a bonyolultabb modell nem szükségszerűen jelenti a pontosság vagy a valósághoz való hűség növekedését (vagy kevesebb hibát). Az egyszerű modellek néha jobbak.

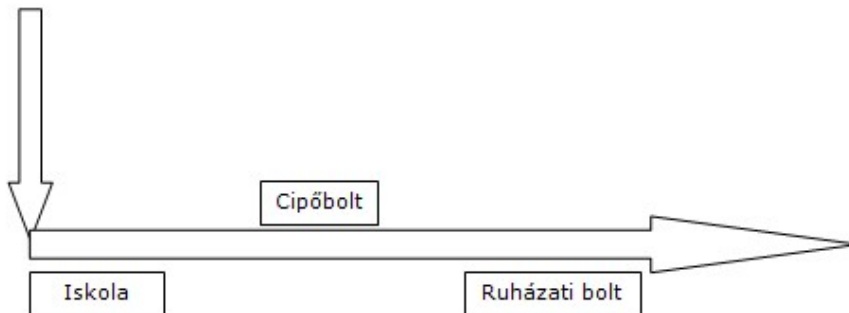
A legegyszerűbb modellezési megközelítés a fekete doboz modell: bemegy valami, történik vele valami, kijön valami. Például a víz körforgását lehet úgy modellezni, hogy leesik az eső, majd elpárolog, illetve elfolyik valahová. Hogy közben mi történik, ez a modell nem vizsgálja, és előfordulhat, hogy erre nincs is mindig szükség. Ennél finomabb megközelítés, a tárolási modell, amikor azt is ábrázoljuk, hogy hová kerül, és hol tartózkodik a víz, amíg elpárolog vagy elfolyik. Még differenciáltabb lesz a modellünk, ha a tárolási modellt földrajzilag kiterjesztjük (más történik ugyanannyi vízzel, ha Skandináviában, vagy a sivatagban esik). Ez a modell figyelembe veszi a vegetáció, az éghajlat, a felszín stb. változásait.

A modellezés másik hibájának oka a szubjektum. A valóság átszűrődik a tudatunkon, amit befolyásol a nemi hovatartozásunk, a kultúránk, a tanult ismereteink és a környezet, amiben felnőttünk. Ha két ember ugyanazon az úton jár munkába, egészen másként érzékeli ugyanazt a valóságot. Mindenkinek más a fontossági sorrendje. Ezt szokták nevezni mentális térképnek is.

Egy férfi mentális térképe:



Egy nő mentális térképe:



Világos, hogy a világról alkotott képünk és az kép alapján készülő modell befolyásolja az információk jövőbeni használatát. Bár ez nem jelenti konkrét hibák jelenlétét, a különböző emberek világgépe különböző, és akár ugyanarról a dologról is eltérő térképet alkothatnak. Jó példák erre a középkori felfedezők által készített földrész – és partvonal térképek, hiszen akkor még lefektetett szabályok sem korlátozták az alkotókat.

### 4.3. 6.4.3 Az egyedek definiálása

A szubjektív világszemlélet miatt a GIS-t érintő legfőbb tényező az, hogyan definiáljuk a földrajzi egyedeket, leíró adataikat és kapcsolataikat. A földrajzi egyedek definiálásának módja és tárolása a térinformatikai adatbázisban hatással van a hibák kialakulására a GIS használatának későbbi fázisaiban. A megfogalmazás fázisában rá vagyunk szorítva geometriai jelképrendszer használatára. Az egyedtípusok korlátozott választéka (pl. pont, vonal, poligon, raszter cella) egyes jelenségeknél problémákat okoz, különösen azoknál, amelyek folyamatos természetűek. Például a magasságokat az analóg térképeken szintvonalakkal és megírt magasságokkal ábrázoljuk. Ezzel a módszerrel egy folytonos felületet csak bizonyos helyeken ismerünk kellő pontossággal. A feltüntetett helyek közé eső felszínről lényegében nincs információnk, közelítő módszerekkel tudhatunk meg valamit róluk. Másik jó példa a növényzet határa a térképeken. Ezeket a határokat a térképen éles vonalak jelzik, holott a természetben az átmenet két növényborítás között folyamatos. Az ilyen átmenetek nagyon gyakoriak a természetben, de jobb híján, a térképeken és a GIS adatmodellben is éles határokat használunk. Ebben az esetben ott követjük el a hibát, hogy az objektum természetét figyelmen kívül hagyjuk. Ez a példa jól szemlélteti, hogy a modellalkotás hogyan vezet helyzeti hibákhoz vagy legalább is bizonytalanságokhoz az analóg és a digitális modellben egyaránt.

### 4.4. 6.4.4 A méretarány hatása az egyed definiálására

Az előző fejezetben említett probléma (az objektum természetének figyelmen kívül hagyása) sokkal többet számít, mint az, hogy az objektumot milyen méretarányban térképezzük, vagy szemléljük a képernyőn. Ha talajfoltok határait nagy méretarányban térképezzük, (1:2 000) a fokozatos átmenet figyelmen kívül hagyása nagy helyzeti hibát okoz. Ha ugyanezeket a határokat sokkal kisebb méretarányban értelmezzük, a hiba ugyanúgy megvan, csak arányaiban kevésbé jelentős (1: 100 000-es méretarányban a 0.5 mm-es térképi vonal a terepen 50 métert jelent, ami valószínűleg nagyobb mint a talajok közötti átmenet).

Hasonló a helyzet a határozott körvonallal rendelkező jelenségeknél. Egymilliós méretarányban egy város csak egy pont. Ötvenezres méretarányban egy poligon, amely csak a beépített terület határát ábrázolja. Huszonötzres



méretarányánál már megjelennek az építési tömbök. Ötezres méretarányánál már önálló épületek is vannak, ezes méretarányánál pedig minden részlet megjelenik pontok, vonalak, foltok formájában.

Az eddigieket úgy is nevezik, mint méretarány függő generalizálás és lényegében a forrástérkép helyzeti pontosságát befolyásolja. A vonalas létesítményeket bizonyos méretarányon túl a helyükről eltolva ábrázolják. A bonyolult vonalakat, mint folyópart, tömeder az áttekinthetőség és az adatmennyiség csökkentése érdekében leegyszerűsítik. A méretarány csökkenésével a generalizálás foka növekszik.

## 4.5. 6.4.5 Mintavételi és mérési hibák

A generalizálás tárgyalásával bevezettük a mintavételezési hiba fogalmát. Vannak ennél fontosabb hibák, amelyeket az adatgyűjtésnél követünk el. Például a digitális magassági modell (DEM) a kérdéses területet borító cellák magassági értékén alapul. Ha a rács sűrű, sokkal több mérésünk lesz és a valósághoz jobban közelítő felszínt kapunk, mintha a rács ritkább. Ez a példa arra a hibára mutat példát, amit a mintavételezés technikájának és részleteinek megválasztásával követhetünk el.

A mérési hibákkal kapcsolatban figyelemmel kell lennünk a mérés pontosságára. A terepi méréseknek mindig korlátozott pontossága van az emberi – és a műszerhibák miatt. Az emberi hibák szubjektív tényezőkre vezethetők vissza, mint az elírás, vagy a mérési eredmény helytelen leolvasása. A műszerhibák közül megemlíthetjük a mérőműszer mérési képességeit, igazíthatóságát, stb.

## 5. 6.5 Az adatok előkészítése

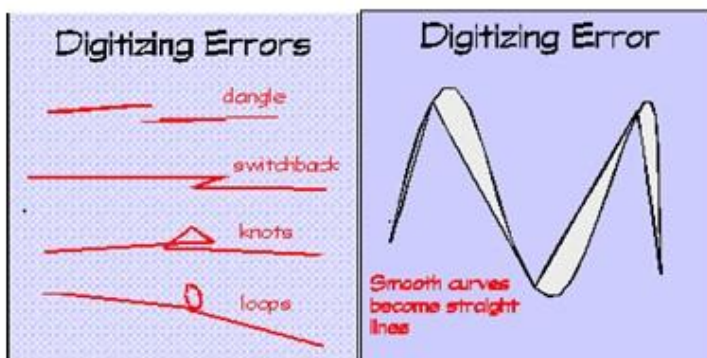
A GIS használatának következő három fázisa az adatgyűjtés, adat integráció és a generalizálás. Az adatgyűjtés ebben az esetben az adat átalakítását jelenti számítógép által értelmezhető formába. Az adatintegráció az a folyamat, amikor az adatot beépítjük a GIS adatbázisba, a generalizálás során kiválogatjuk és egyszerűsítjük a részleteket az adatbázis céljának megfelelően. A továbbiakban az adatelőkészítés során fellépő hibákra koncentrálnunk.

### 5.1. 6.5.1 Az adatgyűjtés

Az adatgyűjtés az adatforrással együtt értelmezhető. A papírtérképet digitalizáljuk manuálisan vagy automatikusan; az űrfelvételeket szkenneljük vagy állomány formájában visszük be, az adatbázisokat és CAD rajzokat konvertáljuk, a fogalmakat kódoljuk. A lényeg az, hogy mindegyik adatgyűjtési forma más – más hibákat rejteget.

#### 5.1.1. 6.5.1.1 Manuális digitalizálás

A manuális digitalizálás messze a leggyakoribb adatbeviteli forma és a helyzeti hibák legismertebb forrása. Nagyon sok körülmény nehezíti a dolgozó helyzetét, amikor követnie kell a térkép vonalait. Ezek között van a kézremegés, hajlam a mellédigitalizálásra, kifáradás, a kéz – szem koordináció hiányosságai. Jenks (1981) két kategóriába sorolta a digitalizálási hibák eredetét: lélektani és fiziológiai. A lélektani hibák között van a digitalizált vonal közepének megítélése, és a szátkereszt pontos vezetése a vonalon. Ezek okozzák azt, hogy a digitalizált vonal a valódi mellett halad, vagy a jobb – és baloldalán kileng. A fiziológiai hibák oka az izmok görcsössége, amely nagyobb kilengések (hurkok, túszerű kiugrások formájában jelenik meg. A digitalizálási hibákat mutatja a következő ábra.



<http://www.colorado.edu/abra/M7P51>

A digitalizálási hibák nagy része viszonylag könnyen felismerhető (hurkok, kiugrások), de a következetes mellédigitalizálást már nehezebb tetten érni. Ennek oka az, hogy maga a digitalizálást végző személy sem érzékeli, hogy mindig egy irányban téved 0.3 - 0.5 mm-t. Annyi lehetőségünk marad, hogy hosszabb megfigyelés után kialakíthatjuk az adott személy „digitalizálási hibaprofilját” és ennek alapján próbálunk korrigálni. A leírt jelenség – azon kívül, hogy hiba – inkább a bizonytalansági tényezők sorába tartozik.

A digitalizálás két módja között is megfigyelhető különbség a helyzeti hibák tekintetében: a pontok digitalizálása vagy a folyamatos digitalizálás. A pontszerű digitalizálás hívei azzal érvelnek, hogy a gondosan kiválasztott és digitalizált pontok helyzetileg jobban ábrázolják a vonalat. Ez azonban újra felveti a mintavétel hibájának problémáját. Az operátor által rögzített több pont részletgazdagabb vonalat eredményez. Bár a digitalizált vonal így jobban követi az eredetit, az egyes digitalizált pontok helyzeti hibája megmarad.

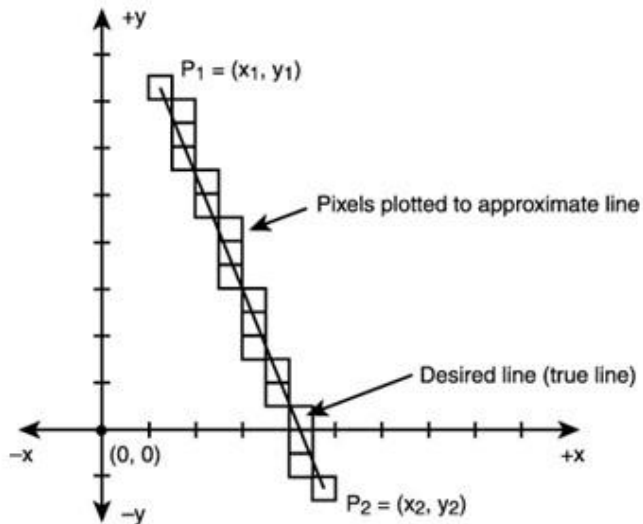
Meg kell még emlékeznünk arról a hibáról, amely a térképnek a digitalizáló tábla (illetve a vetület) koordináta rendszerébe való illesztésekor (regisztrálásakor) keletkezik. Ha ez hanyagul vagy helytelenül történik (kevés vagy nem megfelelően elhelyezkedő illesztőpont, rosszul megválasztott transzformáció), akkor a digitalizált térkép összes koordinátáját szabályos hiba terheli.

### **5.1.2. 6.5.1.2 Automatikus digitalizálás**

Az automatikus digitalizálás hibái mások mint a manuális digitalizálás hibái, egyedül a regisztrálás és a mintavétel hibái mutatnak hasonlóságot. Az automatikus digitalizálásnak két formája van: a szoftveres vonalkövetés és a szkennelés, melyek vektoros illetve raszteres kimenetet eredményeznek. A korábban megismert helyzeti hibák csökkennek (csak a megadott tűrést érhetik el, a program nem fárad el, stb.), viszont a térkép, fénykép minősége döntően befolyásolja az eredményt. A rajzolt vonalakon lévő szakadások (elmosódott rajz, magassági megírással megszakított szintvonal stb.) tönkreteszik a vektorizálás konzisztenciáját és sok utómunkát eredményeznek. A szkenneléssel kapcsolatos hibák főleg a felbontás (pixelméret) megválasztására vezethetők vissza. A nagyobb felbontás (kisebb pixelméret) csökkenti a raszter hibáit, de a tárolandó állomány mérete négyzetesen nő. Kétszer jobb felbontás kb. négyszer akkora állományméretet jelent. Ahogyan már említettük, a helyzeti hibák csökkennek az automatikus digitalizálásnál, az eredményt még sokáig kell javíthatni, hogy beilleszthető legyen a térinformatikai adatbázisba. Utaltunk már a szakadásokra, de problémás a geometriai törvényszerűségek kezelése (épületek derékszögesítése) és a feliratok megkülönböztetése az egyéb rajzi elemektől. További gond, hogy a vektorizáló szoftver, illetve a szkennel nem képes az egyedeket rétegekbe rendezni, ezt is utólag kell megtenni. A vektorizáló szoftverek árát és hardver igényét, valamint az utómunkákat figyelembe véve, gyakran döntenek a felhasználók a manuális digitalizálás mellett.

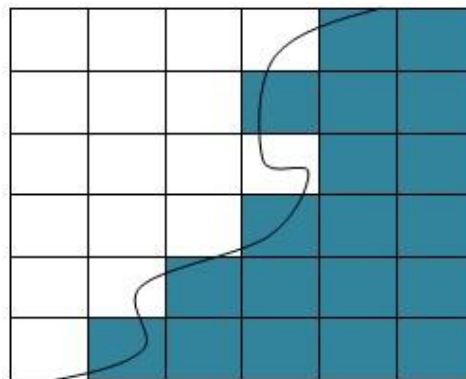
### **5.1.3. 6.5.1.3 Vektor – raszter átalakítás**

A szkennelt légifényképek és az űrfelvételek használata elvezet minket a vektor – raszter konverzió általános problémaköréhez. A digitalizáló táblák vektorokat állítanak elő Y,X koordináta párok sorozataként. Ha raszteres GIS-ben dolgozunk (pl. IDRISI), szükség van a vektorok raszterre történő átalakítására. Az átalakítás módja és a raszter felbontása hatással van a helyzeti hibákra és a leíró adatokra is. A következő ábra a raszterizálás hibáját mutatja.



<http://www.dpfiles.com>

Az ábrán látható, hogyan tolódik el az eredeti vektoros vonal (true line) amikor raszteres formába alakítjuk át. Több olyan cella is van amelyben az eredeti vonal nincs is benne. Újra megállapíthatjuk, hogy ezt a fajta hibát csak a raszter felbontásának növelése csökkentheti. A raszterizálás módszere hatással van a raszterizálás hibáira. A raszterizálásnak két módszere terjedt el, a középpontok módszere és a dominancia módszere. Az utóbbit alkalmazzák szélesebb körben gyakorlatban. A középpontok módszerénél a cella középpontjának a vektorhoz való viszonya határozza meg a cella értékét. A dominancia módszerénél abban a cellában jelenik meg a vektor értéke, amelyikben nagyobb helyet foglal el. A két módszer eredményét szemlélteti az alábbi ábra.



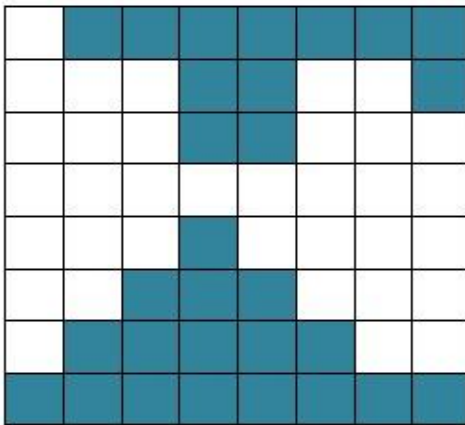
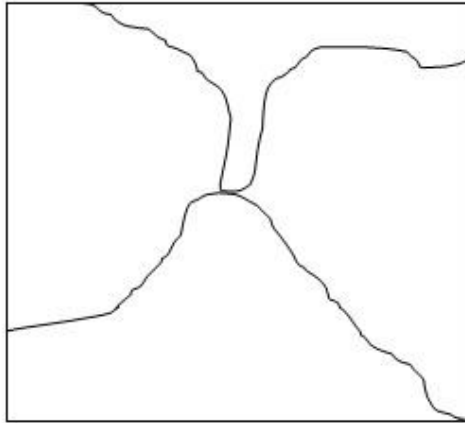
3. ábra A dominancia módszer

Láthatjuk, hogy a két módszer alkalmazásával némileg eltérő rasztert kaptunk. Az eltérés abban jelentkezik, hogy a raszter széle „lépcsős” és helyzetileg is eltolódik.

Ennél az adatelőkészítési eljárásnál jelentkezik a hibák halmozódása illetve terjedése. Ezek a hibák jelen esetben a következőkből adódnak:

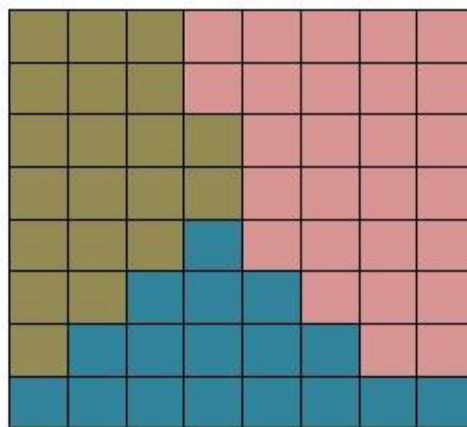
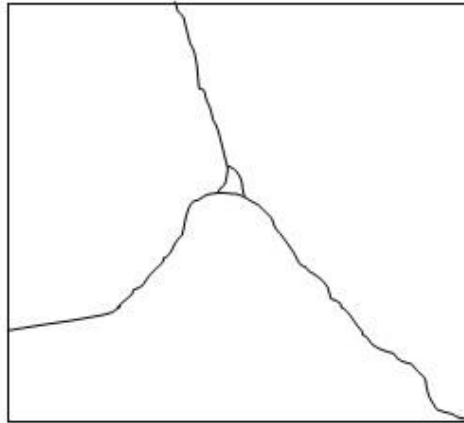
- a forrás térkép hibái (mintavételezési és helyzeti hibák)
- a digitalizálás hibái (szabályos és szabálytalan hibák)
- a raszterizálás hibái (a módszerből eredő hibák)

Egy másik topológiai probléma ami feltétlenül hibához vezet, a kapcsolódás problémája. Ez ott jelentkezik, ahol a vektor poligonok kisebbek, mint a cella méret fele. A poligonok közötti kapcsolat megszűnhet, vagy rossz helyre kerülhet, ahogyan az alábbi ábra mutatja:



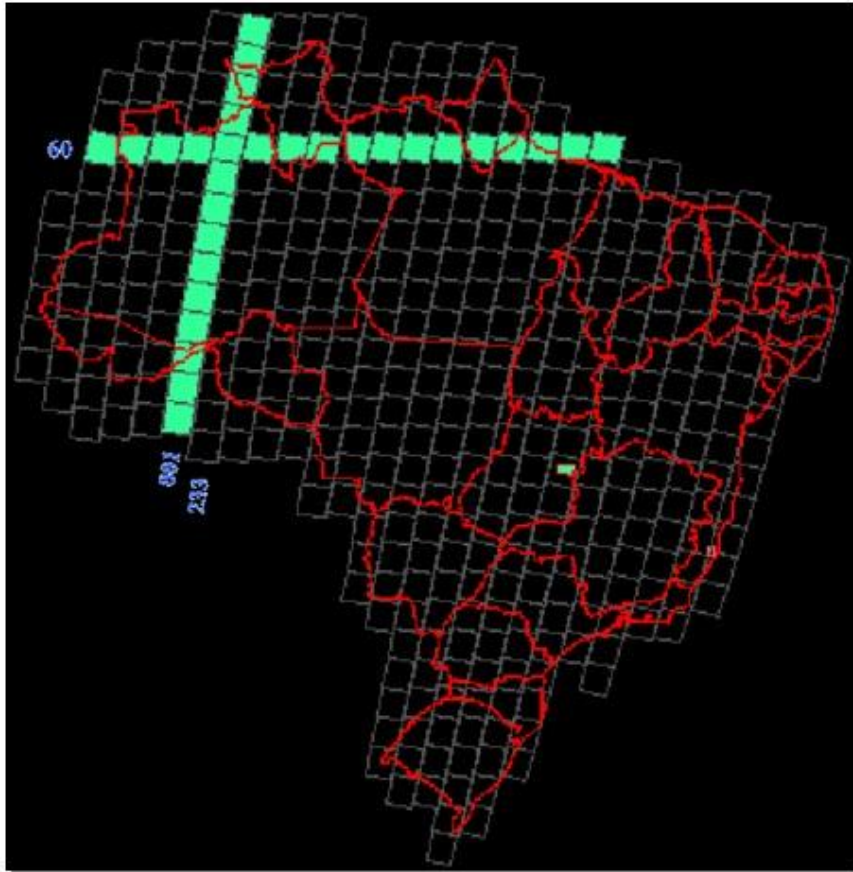
4. ábra A raszterizálás hatása a kapcsolatokra

Hasonló a eredmény akkor, ha a poligon túl kicsi a cella méretéhez képest, ilyenkor a teljes poligon eltűnik.



5. ábra A középső kis poligon eltűnt

A raszter tájolása (elforgatása) szintén hatással van a hibákra. Ha két rács egymással nem a 90 fok többszörösét zárja be, az eredmény minden elforgatásnál különbözni fog. Ez általános eset, amikor a rasztert párhuzamossá tesszük a referencia koordináta rendszerrel. Különösen akkor kell vigyázni, amikor eltérő vetületekben lévő rasztereket (pl. űrfelvétel) próbálunk összeilleszteni. A Landsat űrfelvételek például 9 fokot zárnak be a hosszúsági körökkel.

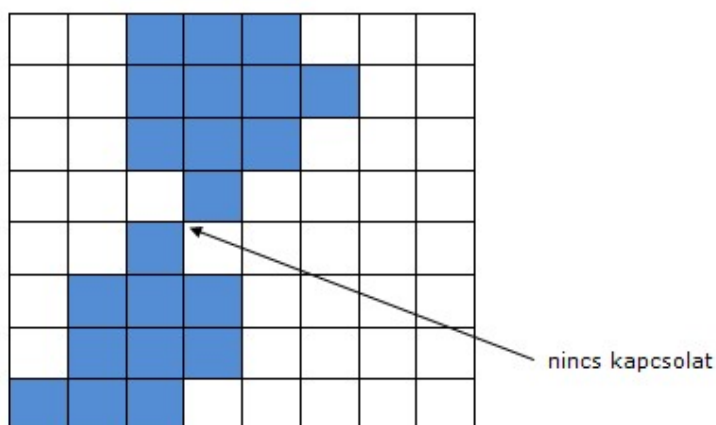
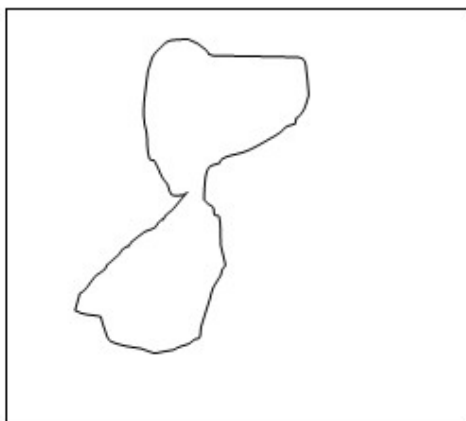


6. ábra A Landsat sorai és oszlopai

<http://www.dgi.inpe.br>

#### 5.1.4. 6.5.1.4 Raszter - vektor átalakítás

A raszter – vektor átalakítás nem rejteget annyi hibalehetőséget, mint a raszterizálás. Topológiai bizonytalanság keletkezik, ha különböző értékű rasztercellák érintkeznek a vektor „sarkain”. Ebben az esetben az eredeti térképhez való visszatérés nélkül lehetetlen megmondani, hogy az eredeti poligonon összetartoztak-e vagy sem? Ezt az információt akkor sem kapjuk vissza, ha újra vektorizáljuk a rasztert. A leírtakat szemlélteti az alábbi ábra.



További figyelmet érdemel az a kérdés, hogy ha a raszter esetleg állományként van megadva, akkor az állományban szereplő értékek a cella mely pontjára vonatkoznak? Ez lehet a cella közepe, bal alsó sarka, jobb felső sarka stb. Vannak olyan esetek, amikor a cella terepi mérete több méter (vagy még több), és ilyenkor az érték vízszintes helyzetében hiba léphet fel.

### 5.1.5. 6.5.1.5 A címkézési hibák

A digitalizálás után következő lépés a digitalizált egyedek (pont, vonal, poligon) címkepontjainak elhelyezése és leíró adatokkal való feltöltése. Az adatelőkészítési folyamatnak során ekkor történik a legtöbb attribútum hiba. Az egyedeket általában tematikusan digitalizáljuk (egyik rétegbe az utakat, másikba a vizeket, harmadikba a földrésztleteket stb.), és a címkek elhelyezésére két módot használunk: a címkeket a digitalizálás közben helyezzük el, vagy mindent digitalizálunk és utólag címkézzük. Bármelyik eljárást követjük, az eredmény az lesz, hogy hiba esetén a leíró adat nem a megfelelő objektumhoz fog kapcsolódni. A poligonok esetében a címke elhelyezése történhet manuálisan vagy automatikusan. Ha címkepontot a dolgozó vagy a program rossz helyre teszi (például a poligonon kívülre), a fentebb leírt hiba lép fel. A címke rossz elhelyezésének oka néha a címkéző szoftver eljárásának a hiányossága. A régebbi programok a poligon súlypontját számították és az lett a címkepont helye. Korszerűbb eljárás a pont – a – poligonban szituáció eldöntése. Tudjuk például, hogy a kifli alakú poligonok súlypontja kívül esik a poligonon. Az alábbi ábra bemutat néhány komplikált esetet.



A jelenlegi programoknál ezek a hibák nem fordulnak elő, de gondolni kell rá, ha régebben készült geoadatbázisokat használnak.

## 5.2. 6.5.2 Adatbázisok konvertálása és a kódolás

Az adatok külső adatbázisból való átvétele és a fogalmak kódolása sok különböző hiba forrása lehet. Ezek nagyrészt attribútum hibához vezetnek, de ha a konvertált adatbázis tartalmaz topológiai információkat (pl. CAD rajz), akkor helyzeti hiba is előfordulhat.

A kész adatbázisok konvertálásakor leggyakrabban fellépő hiba a mezők megváltozása illetve a változók típusának megfeleltetési problémája. Elterjedt gyakorlat volt az adatok kiolvasása a GIS-ből ASCII formában, átformázás valamilyen házilag írt programmal és visszaolvasás a másik GIS-be ASCII formátumban. Ha ebben a folyamatban figyelmetlenek vagyunk, hatalmas adatbázisokat tehetünk tönkre. Gyakori (de a legkönnyebben felismerhető) hiba a koordináták élességének megváltozása a konvertálás során, és az új rajzunk teljesen máshová fog esni a síkban. A leíró adatokkal ugyanilyen könnyű elkövetni hasonló hibákat (mezőszélesség elírása), de annál nehezebb kinyomozni őket. Ha a nyolc karakter széles text mezőt átírjuk hat karakter széles mezőbe, biztos, hogy információt veszünk, sőt az adatok értelmezhetetlenné válhatnak.

Az elírási hibák a geoadatbázisban további hibalehetőségek és helyzeti, valamint attribútum hibákhoz vezethetnek. Ha egy koordinátát írunk el, viszonylag könnyű felismerni a hibát, ha a képernyőre vagy papírra kirajzoljuk az adatbázis tartalmát. Az ember hamar rájön, hogy a felsőrendű alappont nem eshet a tóba, vagy a Balaton középhőmérséklete nem lehet +40 Celsius fok.



Hiba léphet fel komplett állományok átvitele közben, főleg ha fizikai adathordozóról van szó (CD, DVD, flash memória). A hálózaton történő adatátvitel sokkal ritkábban hibás (inkább csak megszakad), mert a korszerű hálózati protokollok beépített hibajavító rutinokkal rendelkeznek.

A fogalmak kódolásakor az a probléma, hogyan alakítsuk át a fogalmat a szoftver által érthető formába, ugyanakkor emlékeztessen valamennyire az eredeti fogalomra (pl. A=jobb, B=ugyanolyan, C=rosszabb; vagy 1=férfi, 2=nő). Amíg a kódok szigorúan következetesek, addig ez nem jelent problémát. Nem szabad azonban elfelejtenünk a kód eredeti jelentéséről. Ha kód kategóriát jelöl, akkor a kategóriákhoz illeszkedő elemző funkciókat kell alkalmazni rá. A fenti példában nincs értelme összeadni a férfiakat és a nőket jelölő kódokat.

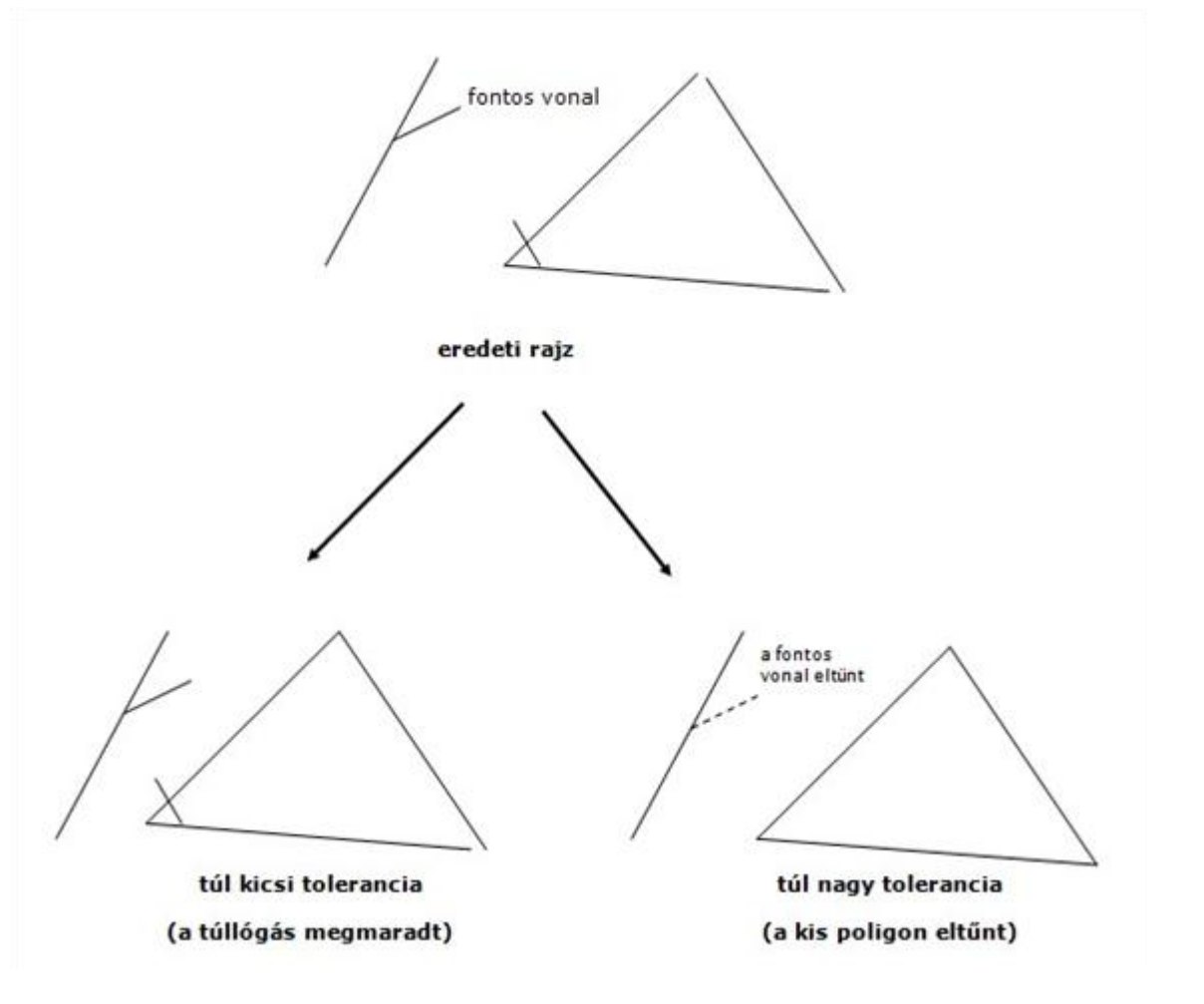
### 5.3. 6.5.3 Adatintegráció

Az adatintegráció során előforduló műveletek: geometria tisztítás, szerkesztés, szűrés, élillesztés, vetületi átszámítás, transzformációk, generalizálás, részekre bontás stb. Ezekről a műveletekkel a jegyzet 5. fejezete foglalkozik.

#### 5.3.1. 6.5.3.1 A geometria tisztítása és szerkesztése

A hibaterjedés során a geometria tisztítása és szerkesztése jelenti a védőfalat, amivel megakadályozhatjuk, hogy az előző lépések hibái tovább terjedjenek a GIS adatbázis felé és meghamisítsák, eltorzítsák elemzéseink eredményét. Természetesen nem lehet minden hibát csapdába ejteni, de a tisztítás és szerkesztés sok nyilvánvaló, vagy kevésbé nyilvánvaló hibát segít kiküszöbölni.

A tisztítás és szerkesztés nem számít hibaforrásnak. Ennek ellenére főleg az automatikus topológia építés okozhat meglepetéseket a geometriában. A poligonok záródását, a túllógó vonalak levágását a felhasználó csak tolerancia megadásával szabályozhatja. A helytelenül megválasztott tolerancia nem kívánt eredményeket hozhat az egész geometriára nézve. Ha túl kicsi, a topológiai hibák nem tűnnek el, ha túl nagy, egész egyedek tűnhetnek el az adatbázisból. A leírtakat szemlélteti az alábbi ábra.



A topológia tisztítását elvégezhetjük manuálisan is ami megelőzi a hasonló hibákat, de nagy adatbázisoknál ez hosszú és fárasztó munka ami magában hordozza az operátos tévedésének lehetőségét.

### 5.3.2. 6.5.3.2 Generalizálás

A generalizálás az adatminőség sok komponensét érinti, beleértve a helyzeti pontosságot, a leíró adatok pontosságát, az adatbázis konzisztenciáját és teljességét. Az objektumok áthelyeződése okozza a helyzeti hibát, a teljesség a kiválasztó és elemző műveletek eredményét befolyásolja; leíró adatok veszhetnek el az átosztályozás miatt, az adatbázis konzisztenciája sérülhet a térbeli és időbeli absztrakció hibás alkalmazása miatt. A generalizálás néha megjósolhatatlan következményeket okozhat a helyzeti pontosságban, a topológiai és a térkép jelentéstani rendszerében. Kimutatható, hogy például a vonalak hossza megváltozik generalizáláskor és ezek a változások meghiúsítják a rétegek közötti műveletek eredményét.

Széles körben elfogadott nézet, hogy a GIS szoftverek generalizáló funkciói hasznos kiegészítések, de számos gyakorlati oka van, hogy a legtöbb még kísérleti stádiumban van. A kartográfusok szerint a teljes körű generalizáló szoftver megvalósítása nem lehetséges, az emberi asszociációs képességet még sokáig nem lehet helyettesíteni szoftverekkel. A legtöbb állami térképészeti szolgálat továbbra is fenntartja a különböző generalizáltsági fokú térképsorozatokat: nagy méretarányú kataszteri térképek, közepes és kis méretarányú topográfiai térképek és földrajzi atlaszok.

*Érvek a generalizálás mellett (Spiess, 1995)*

- A generalizálás segítségével lehetőségünk van a valóság leírására az absztrakció különböző fokain, koncentrálhatunk a felhasználói csoportok számára fontos információkra.
- A generalizálás teszi lehetővé a térbeli modellezést.
- A geoadatbázis megjelenítésénél a generalizálás valamilyen foka megkerülhetetlen; a generalizálás javíthat a megjelenítés minőségén, a lényeg kiemelésén.
- A generalizálás eszköz az információk közlésére egy adott méretarány mellett.
- A generalizálás segít megőrizni az adatok optimális mennyiségét a geoadatbázisban.
- A generalizálás segít eltávolítani a „zajt” a képekről és kiemelni a fontos részleteket (itt nem csak a geometriai zajokra gondolunk).

A fejezet végén megemlítjük még az adatminőség kezelésének és ellenőrzésének kérdését. Ennek a kérdésnek három vonatkozása van:

- az adatok származásának nyomonkövetése
- a szabványok és metaadatok szerepe az adatminőségben
- a hibák megkeresése és megjelenítése a térinformatikai elemzés eszköztárával

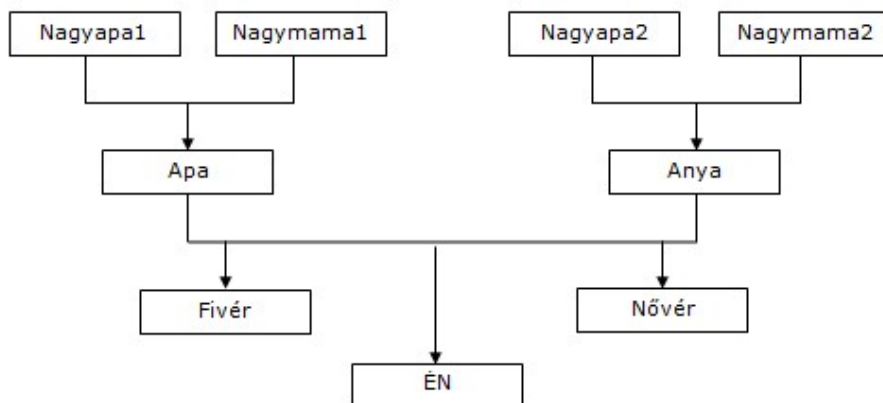
Ebben a modulban csak az adatok származásával foglalkozunk, a másik két tényezőhöz olyan ismeretek szükségesek, amelyek nem tartoznak a térinformatika első részéhez.

## 6. 6.6 Az adatok származása

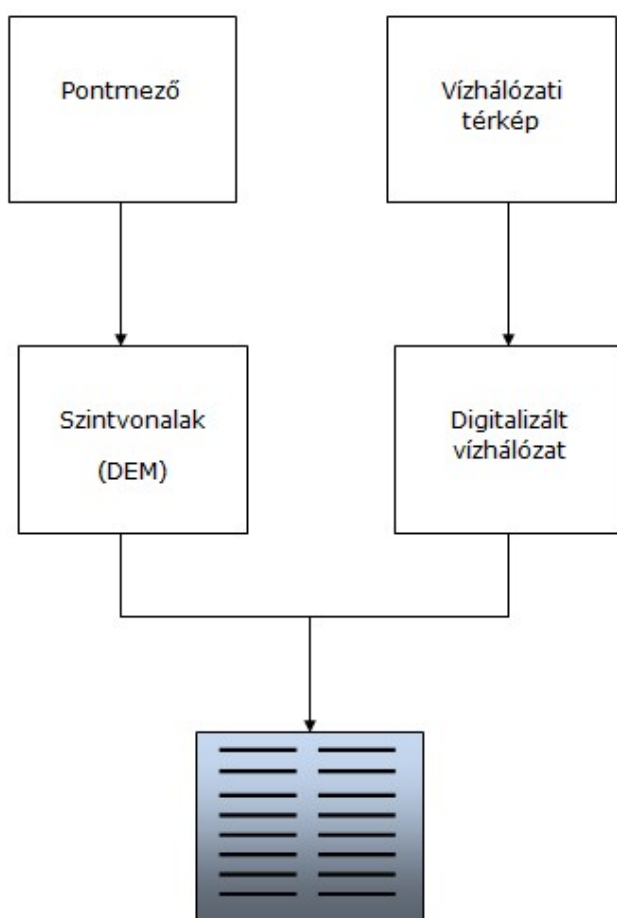
Az adatok származása az adatokkal történt események leírása családfa szerűen megjelenítve. A családfa szerkezete és a benne rögzített adatok két a származás két fontos összetevőjét tartalmazza:

- információkat az egyedekről és csoportjaikról
- a születési és házassági kapcsolatokat

A családfa szerkezete:



Az adat leszármazása:



Vízhozam számítás

### 6.1. 6.6.1 Miért van szükség az adatok származásának követésére?

Az adatok származásának követése nem látszik fontosnak, amíg egyedül dolgozunk egy adatbázison. Általában emlékszünk arra, hogy melyik adat honnan való, mit csináltunk vele. Ha azonban csoportban dolgozunk, vagy munkánkat az utókor is használni fogja, szükség van az adatok származásának ismeretére. A származás megmutatja, hogy honnan vettük az adatokat, milyen átalakításon estek át, míg bekerültek a GIS adatbázisba. A származás követésének rögzítenie kell a változásokat is. A döntéstámogató térinformatikai rendszerekben – a GIS alkalmazások többsége ilyen – rögzíteni kell a valós időt és a GIS adatbázis vonatkozási idejét (adatbázis

idő) (Langran 1992). Látnunk kell az adatbázisnak azt az állapotát, amikor a döntés megszületett. Ha közben megváltozott az adatbázis, ugyanaz az elemzési modell más döntéshez vezethet.

Például van egy térképünk 1988-ból a Szovjetunióról, és van egy, amit 1992-ben készített a Független Államok Szövetsége. Ha nem tudunk a két időpont között bekövetkezett politikai változásokról, helytelen következtetésekre jutunk a két térkép egyidejű használatával.

## 6.2. 6.6.2 Mit kell tartalmaznia a származási dokumentumnak?

Az AGI (Association for Geographical Information = Egyesület a földrajzi információkért) az USA – ban és Angliában a következőképpen határozta meg a származási dokumentum tartalmát:

- **Az adatok eredetének leírása.** Ez a leírás tartalmazza a fontos paramétereket: az adat neve, a létrehozás időpontja és módszere, az utolsó módosítás időpontja, készítője, méretaránya (felbontása), vetülete stb. Informálni kell a felhasználót arról is, milyen technológiával készült az adat.
- **Az átalakítások dokumentálása.** Ezek az információk részletesen leírják az adaton végrehajtott műveleteket, amelyek az adat jelenlegi formájához vezettek. Ez lehet parancsleírás, paraméterek és használatuk szoftveres környezete.
- **Input/Output leírás.** Lényegében a digitális formában létező adat jellemzőinek leírása. Tartalmazza a fájlformátumot, adatátviteli formátumot, a hordozó média leírását és minden beavatkozást ami az állománnyal kapcsolatban történt.
- **ADI (Application Dependent Information = Alkalmazásfüggő információ).** Ez a rész rögzíti, hogy az adatok milyen céllal készültek. Ez az adat segíti a felhasználót annak megítélésében, hogy a saját céljaira alkalmas lesz-e az adatforrás?

## 7. 6.7 Összefoglalás

A térinformatika használata során tisztában kell lennünk a térinformatikai adatbázis megbízhatóságával és hibáival. Ismernünk kell minden lépés lehetséges hibaforrásait: térinformatikai modell, az adatgyűjtés hibái, az adatintegráció hibái, topológiai hibák, az elemzés hibái (ez utóbbival a második félévben foglalkozunk részletesen). Foglalkozni kell egy ritkán taglalt problémával is: az idő múlásának hatása a térinformatikai adatbázisra.

### Önellenőrző kérdések

Mit tekintünk hibának a térinformatikában?

Milyen belső és külső hibái vannak a GIS-nek?

Ismertesse a GIS hibáinak típusait!

Milyen hibái vannak az adatok előkészítésének?

Mit jelent az adatok származása és mire használható?

## Irodalomjegyzék

Blakemore M.: *Generalisation and error in spatial databases*. Cartographica, 21. , 1984

Aronoff S.: *Geographic Information Systems*. WDL publications. , Ottawa, 1989

Goodchild M.F. : *The accuracy of Spatial Databases*, London: Taylor and Francis, 1989

Nigel Trodd : *Data quality*, MMU. 1996, <http://www.colorado.edu>