

Karol HUSÁR

VARIANTY DIGITALIZÁCIE PRIESTOROVÝCH ÚDAJOV

Husár Karol: Variants of Spatial Data Digitization. Kartografické listy, 1998, 6, 1 fig., 3 refs.

Abstract: The contribution deals with technological variants of digitisation of spatial data as linked to vector and raster data structure. Special attention is given to the trends in the field of digitisation in relation to the preferences of particular data structure. At present it may seem that vector and raster data structures are two mutually complementary data structures while each of them has its pros and contras. In spite of it the trend suggests that the raster is ever more preferred at the cost of vector

Keywords: Digitization, vector data, raster data.

Úvod

Téma digitalizácie nie je v súvislosti s geografickými informačnými systémami (GIS-mi) témou novou. Naopak, patrí k tým, ktoré boli z odborného hľadiska už dostatočne rozpracované a z aplikačného hľadiska sa stali rutinnou procedúrou (Husár 1994). V tomto príspevku by sme sa chceli pokúsiť o zhodnotenie digitalizácie priestorových údajov z hľadiska jej historického vývoja. Budeme mať predovšetkým na zreteli vzťah digitalizácie a údajových štruktúr. Dôraz položíme na základné princípy a trendy v oblasti digitalizácie a poukážeme na stav digitalizácie na Geografickom ústave SAV (GÚ SAV).

Miesto digitalizácie vo funkčnej štruktúre GIS

Vychádzajúc z rôznych definícií, GIS v určitom zjednodušenom chápaní možno charakterizovať pomocou jeho štyroch základných funkcií: vstup, výstup, archivovanie a analýza (Goodchild 1987), pričom digitalizácia je jednou zo súčastí vstupu údajov do počítačového prostredia. Ak sa na GIS pozeráme očami geografa, ako na systém, účelom ktorého je produkovať priestorové analýzy, potom prvé tri z uvedených funkcií slúžia ako predpoklad a podpora analýzy. Z tohto pohľadu by sa mohlo zdať postavenie digitalizácie v rámci štruktúry GIS-u akoby okrajové. Sme však presvedčení, že jeho postavenie osobitne v aplikačnej oblasti nemožno marginalizovať a to prinajmenšom z dvoch dôvodov:

- 1) Digitalizáciu vo všeobecnosti možno z hľadiska spotreby času i z hľadiska ceny ľudskej práce priradiť k najnáročnejším aplikačným procedúram v rámci GIS-u.
- 2) Digitalizácia významnou mierou ovplyvňuje kvalitu údajov, pretože ide o proces, ktorý je náchylný k chybám i k menšiemu/väčšiemu stupňu presnosti. Neodstránené chyby i miera presnosti digitalizovaných údajov sú prítomné vo všetkých ďalších procedúrach ich spracovania. V niektorých prípadoch môžu významnou mierou deformovať celkové výsledky. Je v záujme riešiteľa projektu, aby chyby, ktoré vyplývajú z digitalizácie boli pokiaľ možno úplne eliminované a aby stupeň presnosti digitalizovaných údajov bol pod kontrolou človeka a v súlade s vopred určeným účelom aplikácie.

Vektor versus raster

Problém uprednostnenia vektorovej alebo rastrovej údajovej štruktúry v rámci realizácie konkrétnej aplikácie nie je iba problémom dotýkajúcim sa čisto digitalizácie, resp. vstupu údajov, ale je problémom každej funkčnej úrovne GIS-u. S určitým zovšeobecňujúcim zjednodušením možno povedať, že v prípade archivácie údajov v aplikáciách dominuje čo do rozsahu menej náročný a polohovo presnejší vektorový formát, v prípade analýz v aplikáciách prevláda rastrový formát.

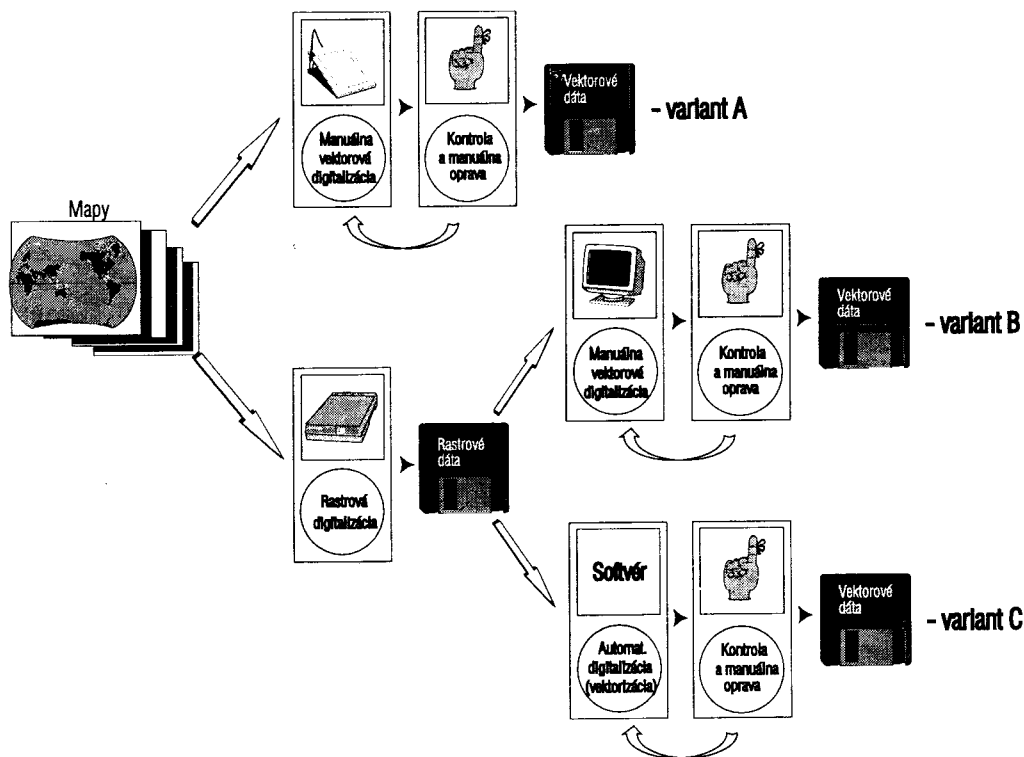
V prípade vstupu i výstupu nemožno jednoznačnejšie hovoriť o dominancii ani jedného z uvedených formátov. Výber formátu je v týchto prípadoch (na rozdiel od archivácie a analýzy) v oveľa väčšej miere závislý od konkrétnych disponibilných hardvérových a softvérových prostriedkov toho-ktorého pracoviska. Ak sme na úrovni vstupu údajov a ak sa na problém *vektor versus raster* pozrieme z historického hľadiska, nemožno nezaznamenať výrazný posun od vektora k rastru (mriežke). Treba poznamenať, že tento vývoj nebol primárne determinovaný vývojom v geografických disciplínach, hoci ani vplyv geografie nemožno úplne marginalizovať. Dominantným faktorom pri určovaní vývoja v oblasti modelov údajov a ich štruktúr sa stali počítačové vedy, priamo alebo nepriamo inšpirované prístupmi zaužívanými na poli geovedných disciplín. Tomuto "boom-u" v oblasti počítačových vied, ktorý mal svoj konkrétny odraz v stále lepšej dostupnosti hardvéru a softvéru (a ktorý napokon trvá dodnes), sa nemohla vyhnúť ani geografia a geografické pracoviská.

Pôvodná preferencia vektorového formátu (variant A na obr. 1) v 70. a čiastočne v 80. rokoch bola primárne daná vtedajšou výlučnou existenciou vektorových zariadení - digitalizátorov a absenciou zariadení rastrového typu (skenerov). V druhej polovici 80. a v priebehu 90. rokov je evidentná snaha potlačiť alebo aspoň úplne vylúčiť ľudský faktor z procesu manuálnej, resp. poloautomatickej vektorovej digitalizácie, a to jednak za účelom zefektívnenia procesu digitalizácie a jednak kvôli eliminácii chýb ľudského subjektu v procese digitalizácie. Výsledkom týchto snáh je skutočnosť, že dnes viaceré špičkové pracoviská používajú na digitalizáciu rastrové zariadenia s vysokou rozlišovacou schopnosťou - skenery a využívanie vektorových zariadení je na ústupe. Následná transformácia týchto rastrových údajov na vektorové (vektorizácia) je buď predmetom manuálnej digitalizácie na displeji (variant B na obr. 1) alebo záležitosťou sofistikovaného softvéru (variant C na obr. 1). Treba spomenúť, že napriek nesporným prednostiam digitalizácie pomocou variantu C, cena softvéru na automatickú vektorizáciu rastrových údajov je t.č. určitou prekážkou v prospech masívnejšieho uplatnenia variantu C v aplikačnej a vedeckej praxi. Zrejme však ide iba o dočasný problém.

Určitým medzistupňom medzi uvedenými variantmi A a C je vektorová digitalizácia rastrového mapového podkladu na displeji počítača (variant B na obr. 1), pričom pôvodná vektorová mapa je podobne, ako v prípade variantu C, najskôr predmetom rastrovej digitalizácie. Aj napriek tomu, že v prípade variantu B ide tiež o manuálnu digitalizáciu, tento variant možno považovať za efektívnejší, a to najmä z dvoch dôvodov:

- a) z hľadiska lepších manipulačných a editačných možností rastrového podkladu v počítačovom prostredí oproti originálnemu mapovému podkladu,
- b) pri digitalizácii pomocou variantu B (na rozdiel od variantu A) vlastná digitalizácia a jej vizuálna kontrola sa uskutočňuje na jedinom a tom istom médiu - obrazovke počítača, dôsledkom čoho sa výrazne znižuje miera chybovosti.

Uviedli sme už, že trend vývoja pri spôsobe digitalizácie smeruje od vektorového módu (variant A) k rastrovému (variant C). Hoci ešte dnes v súvislosti s digitalizáciou nemožno



Obr. 1 Ilustračná schéma troch možných technologických variantov digitalizácie priestorových dajov.

hovorí o výraznejšej dominancii rastra nad vektorom, je reálne predpokladať, že budúce desaťročie bude významne poznačené dominanciou rastra (pritom nemáme na mysli výlučnosť). V súčasnosti určitým, zrejme prechodným problémom masívnejšieho rozšírenia rastrovej technológie digitalizácie je dostupnosť vhodného skenovacieho zariadenia (variant B a C), najmä však dostupnosť vhodného softvéru pre potreby automatizovanej vektorizácie (variant C).

Digitalizácia na Geografickom ústave SAV

Z hardvérových zariadení podporujúcich digitalizáciu GÚ SAV t.č. reálne prevádzkuje digitalizátor TEKTRONIX 4958, (formát A0, rozlišovacia schopnosť 0,025 mm) a skener HEWLETT-PACKARD Scan Jet IIc (216 x 356 mm, 400 dpi). Obe zariadenia sú prepojené s počítačom radu IBM Pentium a PC/486. Zo softvérových produktov ústav vlastní viacero systémov podporujúcich digitalizáciu, napr. IDRISI (Clark University), Design CAD 2-D (American Small Business Computers), TYDIG (Intera TYDAC), Arc View (ESRI) TOPOL (Help Service Mapping) a dva vlastné produkty APTAB a DIGEDIT. Program APTAB je určený na topologickú vektorovú digitalizáciu a DIGEDIT na prvotné spracovanie vektorových údajov a ich geometrické analýzy. Ak vychádzame z požiadavky topologickej digitalizácie, potom systémy IDRISI a Design CAD 2-D nie sú určené na takýto druh digitalizácie, neumožňujú zadávanie topologických parametrov. Tento nedostatok je možné riešiť neštandardnými zreťazeniami údajov v úvodných vetách každého hraničného segmentu.

Pre potreby topologickej vektorovej digitalizácie na GÚ SAV sa t.č. využívajú programy TYDIG, APTAB a TOPOL. Z nich v poslednom čase je preferovaná digitalizácia pomocou programu TOPOL a APTAB (variant A), a najmä program TOPOL (variant B). Nedostatkom uvedených produktov je ich menšia stabilita a tiež aj existencia určitých náhodných chýb pri použití niektorých procedúr. K úplnosti technologickej linky automatickej digitalizácie GÚ SAV chýba softvér na automatickú vektorizáciu rastrových údajov.

Záver

Značná časť primárnych údajových zdrojov, s ktorými prichádza geograf do kontaktu je v analógovej forme. Pri ich hodnotení sa môže uspokojiť s určitým subjektívno-intuitívnym pohľadom, ktorý je závislý od jeho expertnej erudície. Na druhej strane, exaktnejší prístup nevyhnutne predpokladá aj použitie kvantitatívnych metód založených na údajoch v digitálnej forme a rovnako i počítačové prostredie. Predpokladom a cestou k tomuto exaktnejšiemu prístupu je použitie metód digitalizácie. Aj napriek určitým ťažkostiam spojeným s vlastným procesom digitalizácie (náročnosť na čas a náchylnosť k chybám) ide o štandardnú procedúru vstupu priestorových údajov do GIS-u, ktorej sa pri existencii analógových údajov nemožno vyhnúť a ktorá poskytuje bázu na uplatnenie nových technologických možností pri spracovaní priestorovo lokalizovaných údajov.

Príspevok vznikol ako súčasť riešenia vedeckého projektu č. 2/5043/98 "Hodnotenie súčasnej krajiny aplikáciou údajov z databáz CORINE land cover podľa environmentálnych princípov", financovaný grantovou agentúrou VEGA.

Literatúra

- FERANEC, J., OŤAHEL, J., PRAVDA, J., HUSÁR, K. (1994). Formy krajinného krytu identifikované v rámci projektu CORINE Land Cover, *Geografický časopis*, 46, 35-48.
- GOODCHILD, M.F. (1987), A Spatial Perspective on Geographical Information Systems. *International Journal of Geographical Information Systems*. Vol.1,
- HUSÁR, K. (1994). *Vektorová digitalizácia a morfolometrická analýza areálov na príklade foriem krajinného krytu JZ Slovenska*. Kandidátska dizertačná práca. Bratislava (Geografický ústav SAV).

S u m m a r y

Variants of spatial data digitization

The contribution deals with technological variants of digitization of spatial data as linked to vector and raster data structure. Special attention is given to the trends in the field of digitization in relation to the preferences of particular data structure. At present it may seem that vector and raster data structures are two mutually complementary data structures while each of them has its pros and cons. In spite of it the trend suggests that the raster is ever more preferred at the cost of vector. As the Fig. 1 shows out of three possible technological variants of spatial data digitization from the point of view of development the raster digitization (variants B and C, see Fig. 1) is preferred to the detriment of the vector digitization (variant A). We prefer and consider the digitization using variant B and C more efficient for the following reasons: a) it is less time consuming, b) better manipulation and editing of raster source material in computer environment as compared to the original map material, and d) using the variant B (in difference from variant A) the proper digitization and its visual control is carried out using the one and only medium - the computer screen, which results in a considerable lower rate of error.

Fig. 1 Scheme of three possible technological variants of spatial data digitization.

Lektoroval:
Doc. RNDr. J. Feranec, CSc.,
Geografický ústav SAV,
Bratislava