

DIGITALIZACE A KONVERZE DAT

V. Voženílek: *Digitising and Data Conversion*. - Sborník ČGS, 98, 4, pp. 236 - 241 (1993). - The paper deals with the substance and use of data digitisation and conversion. Basic digitising methods are explained and described. Advantages and disadvantages of discussed methods are shown as well as the contemporary situation on the domestic market with digitising devices.

KEY WORDS: digitising - data conversion.

1. Úvod

Proces v oblasti přenosu a zpracování informací, který nastal na začátku 80. let ve Spojených státech amerických, dosáhl takových rozměrů, že se začaly vytvářet nové dílčí disciplíny mnoha základních vědních oborů, geografii z toho nevyjímaje. Přestože se matematizace a kvantifikace geografie začala výrazněji projevovat již v 60. letech, nejvíce ovlivnila geografické metody výzkumu právě po roce 1980. S prudkým rozvojem geografických informačních systémů (GIS) vyvstala potřeba efektivního zpracování geografických informací a následovně pak začal souběžný vývoj technického a programového vybavení příslušných pracovišť. V současné době širšího využívání softwarových produktů v geografii se proces sběru, přenosu, uchování a analýzy dat stává jedním ze základních problémů aplikace geografických informačních systémů v geografické praxi.

2. Digitalizace

Způsob získávání dat pro informační systém je různý. Ať již data získáme jakkoli, vždy je musíme vložit do vstupního podsystému geografického informačního systému v předem určeném (standardizovaném) tvaru. Tento proces, nazývaný digitalizace (digitising), je jednou z nejdůležitějších součástí počítačového zpracování geografických informací.

V současnosti je však třeba chápat dvojí digitalizaci. Digitalizace v širším slova smyslu je proces, na jehož počátku je hmotná předloha (mapa, plán, náčrtek, fotografie, tabulka, text aj.) určená k převodu do digitálního tvaru a na konci digitalizace vystupuje datový soubor (nebo sada datových souborů) komplexně zpracovaný, přehledně uspořádaný, "vyčištěný" od šumu a nežádoucích prvků a svým formátem přizpůsobený dalšímu zpracování. Výstupy mají zpravidla vektorový formát a všechny objekty (entity) jsou konkrétně identifikovány. Digitalizace v širším slova smyslu se skládá z vlastní digitalizace (digitalizace v užším slova smyslu) a z konverze digitálních dat z rastrového formátu do vektorového formátu.

Digitalizace v užším slova smyslu je proces transformování analogových prostorových dat do jejich číselného (digitálního) vyjádření (3) nejčastěji rastrového formátu. Digitalizace je nejčastěji prováděna pomocí speciálních zařízení, buď ručně, poloautomaticky nebo automaticky. Pro uchování dat jsou aplikovány nejčastěji rastrové a vektorové struktury, výjimečně struktury jiné, např. smíšené struktury na principu hypergrafů (2). Digitalizace zahrnuje řadu úkolů s cílem zakódovat lokalizační identifikátory prostorových dat. Vlastním úkolem digitalizace je sejmutí obrazu manuálně nebo příslušným zařízením (skener) a vytvoření přijatelného, bezchybného, pro počítač čitelného datového souboru získaného z rozdílných zdrojů nebo zdrojových materiálů.

V oblasti technického vybavení se vývoj zaměřil na výkonné snímače pro hromadné zpracování dat s možností zpracování i ručně pořízených dokladů. Další vývojové

tendence pak směřují ke zlepšení rozlišovací schopnosti snímačů, ke snímání černobílého obrazu s více odstíny šedi, popřípadě pak i ke snímání obrazu barevného. Programové vybavení se vyvíjí ke zlepšení poskytovaných služeb a aplikací, např. zvětšení či zmenšení obrazu, zpracování jen části předlohy, otáčení, slučování, popřípadě překrývání obrazu. Pro uživatele je pak rozhodující ekonomická efektivnost, plynoucí ze vztahu mezi pořizovacími náklady a přínosem pro uživatele, odpovídající výkon zařízení, minimální chybovost, popřípadě minimální pracnost při opravování zjištěných chyb a samozřejmě i technická a programová kompatibilita. Pokud jde o konfiguraci pracoviště pro digitalizaci a konverzi dat, k osobnímu počítači se připojuje zpravidla jen jeden snímač. Existuje však již několik modelů komplexních modelů pro zpracování textu a obrazu, vybavených vlastním mikropočítačem s vysoce kvalitní hlavní pamětí.

Vedle toho je však uchování informací na tradičních materiálech, jakými jsou například papír nebo mikrofilmy, dodnes populární především díky nízkým nákladům a snadnosti jejich manipulace při archivaci. V současné době však vzniká problém, jak informace z těchto tradičních médií převést do formátu kompatibilního s technologiemi GISů a jiných softwarových produktů zabývajících se analýzou dat.

3. Konverze dat

Konverzí dat se rozumí proces převodu digitálních dat z rastrového formátu do formátu vektorového (tzv. vektorizace), převod opačný (z vektorového do rastrového) nebo převod datové struktury do jiné datové struktury.

Rastrová a vektorová data jsou dva základní typy digitálních dat. Rastrová data vystupují ve formě jednotlivých bitů, jež nenesou žádnou objemnou informaci. Aplikací programy zpracovávají jeden nebo skupinu bitů jednoduchými operacemi. Rastrová data jsou především užívána pro informace, které jsou určeny k zobrazení, nikoliv k modifikování či analýze. Rastrová data lze kompresí zhustit, a proto často zabírají méně paměti než data vektorová. Pro rastrová data se nejčastěji užívají formáty RLE (run-length encoding), RCL, TIFF a CCITT.

Vektorová data byla vyvinuta a poprvé použita v CAD systémech a používají se pro informace, jež mají být dále zpracovány a analyzovány. Jak již sám název napovídá, vektorová data jsou posloupností jednotlivých bodů nebo linií, které jsou geometricky nebo analyticky sloučeny. Toto sloučení znamená, že informace může být přiřazena bodu nebo linii k vytvoření inteligentní množiny (vektoru), jež obsahuje objemnější informaci, například obsahovou. Pro vektorová data se nejčastěji používají formáty DXF, DWG (Autodesk format), IGES (Initial Graphics Exchange Specification), DGN (Intergraph format) a HPGL (Hewlett Packard Graphic Language).

Většina GISů dnes zpracovává data rastrová i vektorová, např. ARC/INFO, INTERGRAPH, ERDAS, SPANS aj. Existují však i GISy rastrově založené, např. IDRISI.

Stěžejní atributy procesu konverze dat je rychlost, výběr a slučování. Teprve v posledních letech se tyto atributy začínají společně uplatňovat v softwarových produktech pro konverzi dat. Problém konverze dat není pouze problémem převodu náčrtů. Je to zároveň proces převodu negrafických dat na data grafické povahy.

Z hlediska konverze ve smyslu pořizování digitálních dat lze rozdělit data do tří skupin. Za prvé jsou to grafická data na mapách a výkresech (symboly, linie, areály). Za druhé jsou to atributy grafických dat (barva, struktura symbolů a linií, názvy, měřítko apod.). Tyto jsou nejčastěji specifikovány ještě před začátkem vlastní konverze. Třetí skupinu tvoří negrafická data. Jsou to detailní charakteristiky objektů, nejčastěji texty nebo symboly. Většina negrafických dat se na mapách a výkresech vůbec nevyskytuje. Jsou shromážděny buď v doprovodných materiálech (vysvětlivky, seznamy, popisný text, použité normy atd.) nebo v databázích. Během procesu konverze jsou jednotlivé skupiny dat zpracovávány mnohdy naprosto odlišným způsobem.

4. Metody digitalizace

Existuje několik metod umožňujících převedení dokumentu do digitální formy kompatibilní s programy GIS. Uživatelé si mohou vybrat z velké nabídky programových souborů specializovaných na převod (konverzi) dat. A. Ogata (5) rozděluje konverzní metody na: ruční, stolní, poloautomatickou, automatickou a interaktivní automatickou digitalizaci.

Ruční (manuální) digitalizace (point-to-point digitising) je nejstarší metoda konverze dat do digitálního tvaru. Operátor vykonávající digitalizaci definuje nejprve soustavu rovinných souřadnic na digitalizovaném dokumentu. Potom vybírá jednotlivé body, jejichž souřadnice měří na podkladovém materiálu a manuálně je zapisuje do souboru, který je pomocí grafických programů zobrazen na obrazovce počítače. Výhodou této metody je skutečnost, že nevyžaduje hardwarově ani softwarově specializované vybavení ani zkušenosti s jejich užíváním. Naopak lze využít operátorovy odborné znalosti související s povahou digitalizovaného materiálu. Tato metoda je však časově velmi náročná, protože je prováděna ručně a operátor se při práci často pohybuje mezi digitalizovaným dokumentem a počítačem. Přesnost digitalizace závisí jednak na operátorově fyzické a psychické kondici a únavě, jednak na výběru bodů při digitalizaci křivek a nepravidelných polygonů. Tato metoda se nejvíce využívá pro digitalizaci dokumentů neobvyklých velikostí a tvarů, pro nákresy s velkým množstvím průsečíků nebo pro digitalizaci nákrusů na poškozených nebo velice křehkých materiálech, které nemohou být zpracovány na digitalizačních zařízeních.

Stolní digitalizace (table digitising) na rozdíl od bodové digitalizace odstraňuje přebíhání operátora od počítače k dokumentu a zpět. Podstata stolní digitalizace spočívá ve využití digitizéru, což je tablet s tlačítkovým kurzorem se záměrným křížem. Využívá se k identifikaci bodů na nákresu a vydávání příkazů pro počítač. Digitalizovaný dokument je umístěn na zkalibrovanou desku digitizéru a operátor na něm vybírá body, které jsou určeny k převodu do digitální formy ve vektorovém formátu. Digitalizací jsou určeny x,y souřadnice bodů, které se zapisují na obrazovce počítače. Stolní digitalizace je časově velmi náročná, protože body, linie a objekty musí být manuálně identifikovány. Přesnost závisí na psychických a fyzických schopnostech operátora a jeho únavě. Výhodou stolní digitalizace je zavedená technologie digitizéru, relativně spolehlivý výsledek (je-li správně nastavená soustava souřadnic), snadná obsluha a možnost zpracovat velké a tvarově netradiční formáty. Nejčastěji se používá pro digitalizaci nákrusů s velkým počtem průsečíků, dokumentů nepravidelných formátů a nákrusů na poškozených nebo velice křehkých materiálech, které nemohou být zpracovány na digitalizačních zařízeních. Digitizéry jsou na našem trhu nabízeny širokým okruhem firem v provedení stolních desek. Příkladem mohou být digitizéry firmy Accugrid nabízené na našem trhu za ceny od 99 000 do 259 000 Kč nebo digitizéry Mantissa za 199 710 až 346 180 Kč.

Poloautomatická digitalizace (Heads-up Digitising) je metoda využívající speciálně vytvořeného softwaru ke snadnější konverzi rastrových dat na data vektorová. Zpracovávaným materiálem je tedy již soubor v rastrovém formátu, který je získán nejčastěji využitím skeneru. Operátor prohlíží rastrový soubor na obrazovce počítače jako na pevné desce a vybírá jednotlivé objekty k vektorizaci. Celá vektorizace je řízena a prováděna operátorem, který musí identifikovat každý objekt zvlášť. Objekty jako např. texty a symboly musí být speciálně označeny. Většina programových vybavení pro vektorizaci zahrnuje prostředky k snadší manipulaci a komunikaci při konverzi těchto objektů, např. CAD Overlay ESP od firmy Image Systems Technology, Inc. (USA). Vektorizace je nejpomalejší metoda, přestože již bylo prezentováno zařízení 2krát až 3krát rychlejší než proces stolní digitalizace. I zde zůstává problém obtížného reprodukování přesného průběhu křivek a nepravidelných polygonů. Stejně jako ruční a stolní digitalizace závisí i tato metoda na schopnostech a únavě operátora. Používá se především v případech, kdy uživatel požaduje převedení do vektorového formátu jen

část rastrového souboru (část nákresu), nebo pro zhotovení vektorového souboru nákresů s menším počtem linií a nepravidelných polygonů.

Automatická digitalizace (Automatic Digitising) je plně automatizovanou metodou převodu rastrového formátu do vektorového. Když byla metoda automatické digitalizace poprvé představena, byla přivítána jako všelék na problém konverze dat ve světovém měřítku. Automatická digitalizace vyžaduje minimální nebo vůbec žádné zásahy operátora během vlastní konverze dat. Přestože toto je případ metody s nejvyšším podílem automatizace na současném světovém trhu, vyvíjejí výrobci těchto zařízení systémy, u kterých jsou některé kroky nutně vyžadovány od operátora. Tato metoda zpracovává soubor rastrového formátu, který byl pořízen nejčastěji skenováním dokumentu, a převádí do vektorového formátu vše, co převést lze. Automatická digitalizace se skládá z úvodní editace (pre-process editing), vlastní konverze a závěrečné editace (post-process editing). Úvodní editace obsahuje prohlédnutí rastrového souboru a nastavení parametrů pro vlastní konverzi. Konverze dat pak probíhá automaticky nezávisle na operátorových zásazích pod řízením z dávkového souboru (batch file). Jakmile je konverze vykonána, musí být každý vektor editován (závěrečná editace). V případě, že byly vytvořeny vektory z nežádoucích objektů na dokumentu, např. znečištěné plochy, ohyb papíru, vpisky aj., příslušné vektory se smažou. Závěrečná editace je nezbytná pro upřesnění výsledků konverze, která může chybně převést některá rastrová data. Například linie, kterou protínají jiné linie, je interpretována jako oddělené liniové segmenty, v jiném případě text nemusí být poznán jako symboly, ale jako objekty. Nejčastěji se automatická digitalizace využívá pro konverzi dat velkého počtu podobných dokumentů nebo dokumentů s jednoduchými liniovými prvky, opakovanou symbolikou, to vše ve výborném provedení, tj. z kvalitního rastrového souboru. Řřednostmi této metody digitalizace je velká výkonnost při převodu liniových dat, nevyžaduje žádné operátorovy zásahy během konverze, spolehlivě rozpoznává strukturně pravidelné nebo navzájem podobné objekty, symboly a texty. Nevýhod má tato metoda poměrně dost. Vyžaduje vyškolený personál, rozpozná pouze omezený počet symbolů, konverze ostatních objektů není zcela přesná, úvodní a závěrečná editace mohou zabrat větší množství času, je velice obtížné převést pouze vybranou část rastrového souboru (např. jen vrstevnice), závisí na kvalitě rastrového souboru čili kvalitě skeneru a snímané předlohy. Nezanedbatelná je i cena zařízení pro automatickou digitalizaci. Některé výrobky mají cenu pod 2 000 \$, ale nedisponují prostředky pro závěrečnou editaci a vyžadují další zpracování jiným softwarovým nástrojem, nejčastěji GIS. Nejdražší produkty, např. Scorpions SRV 386 Software 16 000 \$, GTX Professional Conversion Series 38 000 \$ nebo Parallel Reader firmy Caere 10 995 \$, poskytují velice výkonné nástroje pro zpracování velkého množství specifických objektů (čárkované a tečkované linie, přerušované oblouky, dvojité linie, texty apod.). U těchto produktů uvádějí výrobci, že úvodní editace zabírá 8-17 % a závěrečná editace 63-75 % celkové pracovní doby automatické digitalizace, v závislosti na tom, zda se pracuje na PC nebo na workstation.

Interaktivní automatická digitalizace (Intelligent Digitising) kombinuje hlavní rysy vektorizace a automatické digitalizace. Operátor prohlíží na obrazovce počítače soubor v rastrovém formátu, který byl předem pořízen skenováním, a vybírá objekty, které budou vektorizovány. Vektorizace je prováděna interaktivně a může probíhat automaticky nebo manuálně. Operátor může v průběhu konverze umísťovat nebo opravovat již převedené objekty. Původní termín "intelligent digitising" je použit, protože produkty této kategorie dovolují operátorovi rozhodovat v době, kdy probíhá vlastní konverze dat. Většina výrobků této kategorie, např. Pixel Trak od firmy Cadix Research and Development Corp. (Kanada), obsahuje také ruční režim, a tak operátor může některé procesy přizpůsobovat podle jeho potřeb. Poskytují základní vlastnosti vektorové editace, které dovolují opravy konverze v přijatelném čase, linie a areály mohou být položeny až do 200 vrstev a vektory mohou být zobrazeny v x-y-z souřadnicích, tak jak to vyžaduje většina GIS. Jejich cena nedosahuje velkých hodnot (např. Pixel Trak 1 795 \$, CadCore Tracer od Hitachi America LTd. 4 500 \$). S rozšířením služeb těchto softwarů,

zejména o CAD-nástroje, roste i jejich cena. Nejčastěji se interaktivní automatická digitalizace využívá při konverzi řady aplikací a databází rastr-vektorových formátů nebo dokumentů vyžadujících operátorovo odborné zhodnocení. Přestože je výsledek této digitalizace závislý na kvalitě rastrového souboru a zkušenostech a schopnostech operátora (únava a kondice), poskytuje tato metoda řadu výhod. Obsahuje režim pro manuální vektorizaci, zcela odpadá fáze závěrečné editace, snadněji probíhá konverze symbolů, textů a specifických linií a lze interaktivně pracovat v průběhu konverze. Velkou předností je také možnost vektorizovat pouze vybraná data. Produkty této a předcházející skupiny nabízejí nejen čtení textu zaznamenaného strojově na standardních psacích strojích a různých typech počítačových tiskáren, popřípadě ručně pomocí kuličkového pera nebo tužky, ale i přímý záznam obrazu, například z výkresu, plánu, mapy, tabulky, kresby, novinového článku, časopisu, knihy, fotografie nebo výstupu ze souřadnicového zapisovače.

5. Závěr

Česká geografie se v současnosti nachází na prahu intenzivního zavádění GISů do geografické praxe. Digitalizace a konverze dat je prvním úskalím, které je třeba úspěšně překonat pro jejich efektivní uplatnění.

Vývoj v oblasti digitalizace je velmi bouřlivý a směřuje k plně automatizovanému záznamu a převodu dat. Proto je také věnováno značné úsilí k vytvoření takového procesu digitalizace, který v co největší míře omezuje a vylučuje vznik chyb.

Je však třeba konstatovat, že domácí trh nabízí pro fázi zavádění těchto nových přístupů odpovídající zařízení (pro ruční, stolní a poloautomatickou digitalizaci). Digitalizace a konverze dat je velice důležitou a v jistém smyslu i širokou oblastí v teorii GISů. Ovšem nelze ji stavět na přední místa v důležitosti fungování celého systému, protože existuje mnoho závažnějších témat. Postačí proto, bude-li digitalizace a konverze dat bezchybně pochopena a interpretována.

Literatura:

1. BROUGH, P. A.: Principles of Geographic Information Systems for Land Resources Assessment. Oxford, Monographs on Soil and Resources Survey, No. 12, 1986.
2. BOUILLÉ, F.: Structuring Cartographic Data and Spatial Processes with the HBDS. In: Harvard Press on GIS, Harvard, vol. 5., s. 48-86.
3. KONEČNÝ, M.: Úvod do geografických informačních systémů. 1. vyd. Brno, UJEP 1985, 110 s.
4. LEMMENS, M.J.P.M.: GIS: the data problems. EGIS 91, Brusel, 1991, s. 626-637.
5. OGATA, A.: Making the Most of Today's Data Conversion Methods. In: GIS Europe, Lemmer, vol. 1, s. 70-73.
6. PEUGUET, D.J.: Raster Data Handling in Geographic Information Systems. In: Harvard Papers on GIS, Cambridge Mass. Laboratory for Computer Graphics and Spatial Analysis, vol. 2, 112 s.
7. RADO, B.Q. et al.: Raster-Vector integration: case study examples of urban, wetlands and regional studies. EGIS 91, Brusel, 1991, s. 914-920.
8. SHAND, P.J.: Hitachi? CADCore speeds GIS data conversion. In: Mapping Awareness. Oxford, vol. 5, No. 2, s. 49-53.

DIGITISING AND DATA CONVERSION

A number of methods converting hard-copy documents into a data format compatible with GIS programmes exist. As a result, there is a choice from many specialized conversion products. Conversion methods include point-to-point digitising, heads-up digitising, automatic digitising and intelligent digitising.

In the process of point-to-point digitising an operator selects specific points from a hard-copy document and places them into the computer image with help of GIS drafting functions. This method is of best use in case when scanning is impossible (digitisation of odd-sized drawings, drawings with many intersections and damaged or fragile drawings).

Table digitising includes placement of a hard-copy document on a calibrated digitiser board. Then, operator selects points to be converted into a vector form. This method is used in analogous cases as point-to-point digitising.

Heads-up digitising is described as follows. Operator examines a raster file on the computer screen and selects specific points to be vectorized. Each item needs to be vectorized manually. This method is used for simple drawings with a few lines only, curves and irregular polygons or when only part of a drawing is to be vectorized.

In the process of automatic digitising the computer itself examines the raster format file that has been created by a document scanner and converts all items into the vector format. This method is mostly used when digitisation of a number of similar documents with simple data and regular symbology in a good quality is required.

Intelligent digitising may be described as follows. Operator examines a raster file on the computer screen and selects items to be vectorized. Vectorisation may be done automatically or manually. Intelligent method is used for hybrid raster/vector applications and databases and for drawings to be only partially digitised.

(Pracoviště autora: katedra geografie Přírodovědecké fakulty Univerzity Palackého, tř. Svobody 26, 771 46 Olomouc.)

Došlo do redakce 27.3.1992

Lektoroval Dalibor Moravec