

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA GEOGRAFIE

František STUPAL

**VYUŽITÍ VOLNĚ DOSTUPNÝCH (GEO)DAT
PRO TVORBU TEMATICKÝCH MAP EVROPY**

Bakalářská práce

Vedoucí práce: Mgr. Zuzana ŠTÁVOVÁ

Olomouc 2009

Vysoká škola: Univerzita Palackého

Fakulta: Přírodovědecká

Katedra: Geografie

Školní rok: 2008/09

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

student

František **STUPAL**

obor

Regionální geografie

Název práce:

Využití volně dostupných (geo)dat pro tvorbu tematických map Evropy

Free (Geo)Data for Thematic Maps of Europe

Zásady pro vypracování:

Cílem bakalářské práce je shromáždit volně dostupná (tj. zdarma dostupná na Internetu nebo jako sample data sets) geodata popř. i atributová data v digitální podobě, která je možné zdarma získat a poté dle potřeby upravit a využít pro tvorbu tematických map Evropy. Bude se jednat o (geo)data v územním rozsahu celé Evropy (přesné prostorové vymezení je třeba provést na začátku práce) a ke každé datové sadě nebo (geo)databázi budou vyplněna metadata ve formě metadatového formuláře.

Dalším krokem bude zpracování získaných (geo)dat ve vybraném softwaru GIS včetně návrhů zpracování různými kartografickými metodami a sestavení ukázek tematických map Evropy pro každou datovou sadu nebo (geo)databázi.

Struktura práce:

1. Úvod
 2. Cíle práce, metody a postup zpracování
 3. Volně dostupná (geo)data
 4. Zpracování volně dostupných (geo)dat v GIS
 5. Tvorba tematických map Evropy
 6. Diskuse
 7. Závěr
 8. Shrnutí – summary, klíčová slova – key words (v českém a anglickém jazyce),
 9. Literatura
- Přílohy

Bakalářská práce bude zpracována v těchto kontrolovaných etapách:

Student v první fázi své práce shromáždí zdarma dostupné datové sady a (geo)databáze obsahující polohové i atributové údaje pro všechny státy Evropy resp. pro největší možný počet těchto států.

Ke každé datové sadě nebo (geo)databázi vytvoří potřebná metadata.

U získaných (geo)dat zjistí možnost jejich zpracování v GIS a navrhne možnosti kartografické vizualizace ve vybraném GIS softwaru.

Z vybraných (geo)dat sestaví ukázkové tematické mapy Evropy.

Rozsah grafických prací: ukázkové tematické mapy Evropy.

Rozsah průvodní zprávy: 10 000–12 000 slov základního textu včetně všech příloh v elektronické podobě

Seznam odborné literatury:

Dublin Core Metadata Initiative. URL: <<http://dublincore.org>> [cit. 2008-6-30].

EuroGeographics:Products & Services. URL:

<http://www.eurogeographics.org/eng/04_products.asp> [cit. 2008-6-30].

Global Map. URL: <http://www1.gsi.go.jp/geowww/globalmap-gsi/iscgm-sec/index.html> [cit.2008-6-30].

INSPIRE Geoportal. URL: <<http://geoportal.jrc.it/geoportal>> [cit. 2008-6-30].

International Society for Digital Earth. URL: <<http://www.digitalearth-isde.org>> [cit. 2008-6-30].

International standard ISO 19115. URL:

<<https://committees.standards.org.au/COMMITTEES/IT004/PRIVATE/I0028/ISO%2019115%20.pdf>> [cit. 2008-6-30].

KAŇOK, Jaromír. *Tematická kartografie*. Ostrava: Ostravská univerzita v Ostravě, 1999. 318 s. ISBN 80-7042-781-7

MIKULÍK, Oldřich, VOŽENÍLEK, Vít, VAISHAR, Antonín. *Studium rozvoje regionu založené na vizualizaci geoinformačních databází*. Olomouc: Vydavatelství Univerzity Palackého, 2008. 181 s. ISBN 978-80-244-1928-2

PRAVDA, Ján, KUSEDOVÁ, Dagmar. *Počítačová tvorba tematických máp*. Bratislava: Vydavateľstvo Univerzity Komenského, 2004. 264 s. ISBN 978-80-893-1700-4

VEVERKA, Bohuslav. *Topografická a tematická kartografie 10*. Praha: Vydavatelství ČVUT, 2004. 220 s. ISBN 80-0102-381-8

VOŽENÍLEK, Vít. *Aplikovaná kartografie I.: Tematické mapy*. Olomouc: Vydavatelství Univerzity Palackého, 2004.187 s. ISBN 80-244-0270-X

Vedoucí bakalářské práce: Zuzana ŠTÁVOVÁ

Datum zadání bakalářské práce: 30. 6. 2008

Termín odevzdání bakalářské práce: duben 2009

Prohlašuji, že jsem zadanou bakalářskou práci vypracoval sám a že jsem uvedl veškerou použitou literaturu a další zdroje.

Olomouc, 18. května 2009

.....

podpis

Děkuji Mgr. Zuzaně Šťákové za odborné vedení, vstřícný přístup a četné připomínky během tvorby bakalářské práce.

SEZNAM POUŽITÝCH ZKRATEK

ANZLIC	Australia and New Zealand's Information Council
CEN	Comité Européen de Normalisation (European Committee for Standardization)
CSV	Comma separated values (hodnoty oddělené čárkou, formát určený pro výměnu tabulkových dat)
ČSN	Česká státní norma
DBF	Database File (tabulkový formát pro atributová data)
DBMS	DataBase Management System (systém řízení báze dat)
EU	Evropská unie
FAO	Food and Agriculture Organization of the United Nations
FGDC	Federal Geographic Data Committee
GIS	geografický informační systém
HTML	Hyper Text Markup Language
INSPIRE	Infrastructure for Spatial Information in Europe
IPCC	Intergovernmental Panel on Climate Change
ISO	International Organization for Standardization (Mezinárodní organizace pro normalizaci)
IUCN/SSC	International Union for Conservation of Nature/Species Survival Commission
JPEG	Joint Photographic Experts Group (metoda ztrátové komprese)
JPG	formát užívaný pro bitmapy s kompresí JPEG
MDB	hlavní formát databázového programu Microsoft Access
OFDA	Office of Foreign Disaster Assistance
OSN	Organizace spojených národů
PDF	Portable Document Format
PNG	Portable Network Graphics (grafický formát pro bezztrátovou kompresi rastrových dat)
SHP	ESRI Shapefile (vektorový formát dat)
SQL	Structured Query Language (strukturovaný dotazovací jazyk používaný v relačních databázích)
TSV	Tab Separated Values (formát určený pro výměnu tabulkových dat)
UN	United Nations (Organizace spojených národů)

UNEP	United Nations Environment Programme (Program OSN pro životní prostředí)
USD	United States dollar (Americký dolar)
WB	World Bank (mezinárodní finanční instituce)
XLS	hlavní tabulkový formát programu Microsoft Excel
XML	Extensible Markup Language (jazyk určený pro výměnu dat mezi aplikacemi)
ZIP	formát pro kompresi a archivaci dat

OBSAH

SEZNAM POUŽITÝCH ZKRATEK	6
1 ÚVOD	10
2 CÍLE PRÁCE	11
3 METODY A POSTUP PRÁCE	12
4 VYMEZENÍ ZÁKLADNÍCH POJMŮ	13
4.1 (GEO)DATA	13
4.1.1 DATA MINING.....	14
4.2 METADATA	14
4.2.1 NORMA ISO 19115	15
4.2.2 INSPIRE	16
4.2.3 DUBLIN CORE.....	17
4.3 GEOGRAFICKÝ INFORMAČNÍ SYSTÉM.....	18
4.4 DATABÁZE	18
4.4.1 DBMS	18
4.4.2 DATOVÝ TYP	19
4.4.3 DATOVÁ SADA.....	19
5 VYMEZENÍ ZÁJMOVÉHO ÚZEMÍ	20
6 ZDROJE DAT	21
6.1 EUROSTAT.....	21
6.2 UNITED NATIONS ENVIRONMENT PROGRAMME – GEO DATA PORTAL.....	22
6.3 GOS – GEOSPATIAL ONE STOP	24
6.4 DALŠÍ MOŽNÉ ZDROJE.....	24
6.4.1 GEOGRAPHY NETWORK.....	24
6.4.2 GEOCOMMUNITY	25
6.4.3 CIA – THE WORLD FACTBOOK	26
6.5 HODNOCENÍ STAŽENÝCH DAT.....	27
7 KVALITA DAT	30
7.1 ZLEPŠOVÁNÍ DATOVÉ KVALITY.....	31
7.1.1 VYLUČOVÁNÍ OKRAJOVÝCH DAT.....	31
7.1.2 DOPLŇOVÁNÍ CHYBĚJÍCÍCH DAT	31
7.1.3 VYLOUČENÍ DUPLICITNÍCH DAT	32
7.2 HODNOCENÍ KVALITY DAT.....	32

7.2.1	OBJEKTIVNÍ METRIKY	33
7.2.2	SUBJEKTIVNÍ METRIKY	33
8	DATABÁZE	34
8.1	PROSTŘEDÍ MICROSOFT ACCESS.....	34
8.2	POPIS DATABÁZE	34
9	PRÁCE SE ZÍSKANÝMI DATY	41
9.1	PŘEDZPRACOVÁNÍ DAT	41
9.2	SYMBOLIKA V PROSTŘEDÍ ARCMAP	44
9.2.1	FEATURES	44
9.2.2	CATEGORIES.....	44
9.2.3	QUANTITIES.....	44
9.2.4	CHARTS	45
9.2.5	MULTIPLE ATRIBUTES	45
9.3	BARVA	45
9.4	KOMPOZICE MAPY.....	46
9.5	KARTOGRAFICKÉ VYJADŘOVACÍ PROSTŘEDKY	47
9.6	TEMATICKÝ OBSAH	48
9.7	TOPOGRAFICKÝ PODKLAD	48
10	KOMENTÁŘ K MAPÁM.....	50
10.1	PODÍL MĚSTSKÉHO OBYVATELSTVA NA CELKOVÉM POČTU OBYVATEL.....	50
10.2	HUSTOTA ZALIDNĚNÍ V EVROPSKÝCH STÁTECH.....	51
10.3	CELKOVÁ PRODUKCE MASA V EVROPSKÝCH ZEMÍCH	51
10.4	ZÁSOBA DŘEVA V EVROPSKÝCH LESÍCH	51
10.5	OHROŽENÁ ZVÍŘATA V EVROPĚ	52
10.6	POROVNÁNÍ CELKOVÉHO OBJEMU MOŘSKÉHO RYBOLOVU.....	52
10.7	HRUBÝ DOMÁCÍ PRODUKT	53
10.8	PŘÍMÉ ZAHRANIČNÍ INVESTICE.....	54
10.9	PODÍL RŮZNÝCH TYPŮ ZEMĚDĚLSKÉ PŮDY	54
10.10	POČET OBYVATEL ŽIJÍCÍCH VE MĚSTECH	54
11	ZÁVĚR	56
12	SUMMARY	57
	LITERATURA.....	58
	SEZNAM PŘÍLOH.....	61

1 ÚVOD

Data jsou nejdůležitější a zpravidla nejdražší součástí každého geoinformačního systému. Pro kvalitní mapy platí, že jejich sběr podléhá nejpřísnějším kritériím a je často záležitostí i několika let. Uživatelé, kteří nemají kvalifikaci nebo časové možnosti pro jejich shánění jsou odkázáni na druhé. Ti za poplatek nabízejí buďto surová data anebo jejich výslednou formu (mapa, síť bodů). Je to jednoduché, je to obchod jako každý jiný. Na druhé straně existují volně stažitelná data, která se nacházejí na rozsáhlé síti jménem Internet. Vznikla totiž myšlenka, že pokud se data sbírají za peníze daňových poplatníků, měla by být i volně přístupná. A ve vyspělých státech se tato myšlenka uchytila. U nás stačí navštívit stránky Českého statistického úřadu, objeví se tolik čísel, až se hlava zatočí a nadšenci můžou vytvářet grafy třeba celý rok. Na mezinárodní úrovni je potřeba dořešit datovou integritu, kvalitu a pravidelný přísun dat.

Nezbytným nástrojem pro tvorbu tematických map je rovněž odpovídající software. Geografický informační systém mnozí charakterizovali jako jednu z nejmocnějších informačních technologií, protože se zaměřuje na integrování znalostí z různých zdrojů a vytváří tak příznivé prostředí pro spolupráci více vědních oborů. GIS je nadto přitažlivý pro většinu uživatelů, kteří se s ním setkali, protože jde o systém zároveň intuitivní a založený na znalostech, který kombinuje efektní vizualizaci životního prostředí se silným analytickým a modelujícím systémem, který je zakotven ve vědecké geografii. Výsledkem této kombinace je důvěryhodná technologie spočívající na vědeckých základech, pomocí níž je možné snadno komunikovat napříč různými kulturami, společenskými třídami, obory a jazyky.

2 CÍLE PRÁCE

Cílem bakalářské práce je analýza volně dostupných (geo)dat, která jsou vhodná pro tvorbu tematických map Evropy. Práce se zabývá hledáním zdrojů nabízejících volně stažitelná (geo)data a hodnocením takto získaných dat. Mezi dílčí cíle náleží tvorba vhodné databáze pro správu metadat a v neposlední řadě také praktické využití dat – tvorba mapových výstupů se všemi kartografickými náležitostmi.

Hlavní myšlenkou práce je podat přehled o současných možnostech praktického využití poskytovaných dat, upozornit na nedostatky a pokusit se odhadnout budoucí vývoj této problematiky.

3 METODY A POSTUP PRÁCE

První část práce se zabývá definováním pojmů, které jsou klíčové pro další kapitoly. Pro samotný výzkum bylo nejprve důležité získat přehled o jednotlivých zdrojích a ty následně popsat, zmínit pozitiva a negativa získaných dat. Pro správu metadat byla navržena databáze v programu Microsoft Access. Závěrečná část práce popisuje práci s daty v prostředí GIS a tvorbu kartografických výstupů.

Pro účely zpracování dat a tvorby map byl použit software ArcView GIS 3.2 a ArcGIS 9.2 od firmy ESRI s licencí pro Univerzitu Palackého v Olomouci. Databáze byla vytvořena v programu Microsoft Access, samotná práce byla sepsána v programu Microsoft Word.

4 VYMEZENÍ ZÁKLADNÍCH POJMŮ

4.1 (GEO)DATA

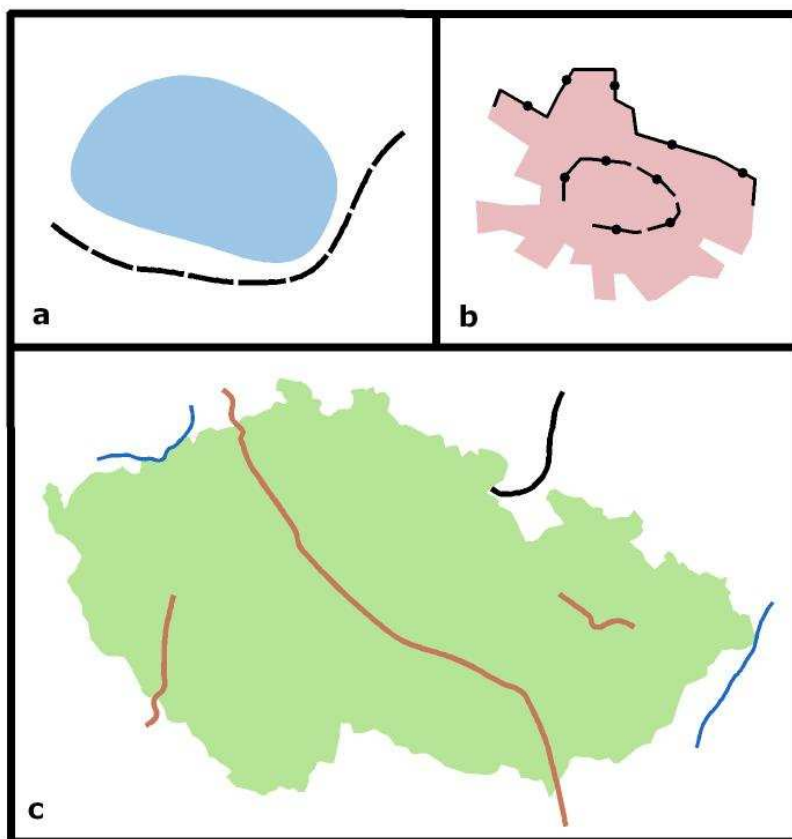
Data jsou obrazy vlastností objektů, vhodně formalizované pro přenos, interpretaci nebo zpracování prostřednictvím osob nebo počítačové techniky (Voženílek, 1998). Data je možné získat měřením, pozorováním nebo pouhým zaznamenáním reality. Rozlišujeme data **analogová** (vytištěná mapa), **alfanumerická** (tabulka), **číselná** (číselná matice). V této bakalářské práci bude řeč výhradně o **digitálních datech**, tj. datech zobrazených číslicemi, popř. zvláštními znaky a znakem mezery, na záznamových médiích.

Geodata neboli **geografická** data jsou polohově lokalizovaná data obsahující tematické informace vázané k údaji o poloze. Jsou určena svým geometrickým tvarem a polohou na zemském povrchu. Mají prostorový, atributový (tematický) a časový aspekt. Rozlišujeme základní geodata (nezbytná pro většinu aplikací), referenční geodata, která používá každý uživatel pro georeferencování svých vlastních geodat a která poskytují obecné prostorové propojení mezi různými aplikacemi. Dále geodata společného základu, což jsou data pro optimální využívání většiny geoinformačních aplikací. Tvoří je: geologická data, data o životním prostředí, registry a databáze. Aplikační geodata jsou specifická pro konkrétní aplikace.

Mezi geodaty se utváří různé **vztahy**. Jedná se o vztahy geometrické, tedy topologické vztahy mezi jednotlivými prvky. Časové vztahy zahrnují nahodilý spoluvýskyt, koexistenci (geoprvek¹ se vyskytuje současně s jinými geoprvky), sukcesí (geoprvek se vyskytuje s časovým odstupem po vzniku jiného geoprvku) či kauzální vztah, kdy geoprvek vzniká jako přímý důsledek vzniku jiného geoprvku. Metrické vztahy měří přímou vzdálenost mezi jednotlivými geoprvky. Syntaktické vztahy zkoumají, které prvky spolu mohou mít vzájemný vztah (geometrický, časový, metrický), a které prvky spolu vzájemný vztah mít nemohou. Vztah „part of“ (je částí) vyjadřuje skladebnost, tedy fakt, že např. říční síť se skládá z řek nebo že území České republiky je, co se administrativní stránky týče, složeno z krajů. Ostatní vztahy jsou všechny ty, jež se netýkají času ani prostoru (např. vlastnické, vztahy nadřazenosti a podřazenosti, apod.).

¹ Geoprvek [angl. Feature] je základní na dané rozlišovací úrovni dále nedělitelná jednotka (entita) na jednotky (entity) stejného druhu.

Existují tři základní typy geoprvků; polygonové geoprvky, liniové geoprvky a bodové geoprvky. Ty je možné ještě dále rozšířit o speciální typ odvozené ze třech základních typů. Každý z geoprvků je možné popsat složkami popisu geoprvku.



Obr. 1 Geometrické vztahy mezi linií a polygonem: a) mimoběžnost; b) vnoření; c) křížení – různé typy (podle Voženílek, 1998, vlastní zpracování)

4.1.1 DATA MINING

Anglický termín „data mining“ (dolování z dat) chápe Petr (2008) jako proces extrakce relevantních, předem neznámých nebo nedefinovaných informací z velmi rozsáhlých databází. Při přípravě dat se obvykle z dat uložených ve složité struktuře, vytváří jedna tabulka obsahující relevantní údaje (hodnoty atributů) o sledovaných objektech. Tato práce se též zabývá „dolováním“ dat z rozsáhlé sítě (Internetu), získáváním informací a jejich praktickým využitím. Informace znamenají smysluplný význam, jenž přisuzujeme datům a vztahům mezi nimi.

4.2 METADATA

Metadata jsou data popisující datové prvky, datové modely a datové struktury. (Voženílek, 1998). Rozdíl mezi metadaty a daty je v jejich použití. Metadata jsou často uložena odděleně od vlastních dat (Dobešová, 2004). Stejně tak je tomu v případě bakalářské práce – metadata jsou uchovávána v autorem navržené databázi v programu Microsoft Access.

Metadata lze chápat jako určité charakteristiky, atributy sloužící k popisu jiných dat, jenž jsou dobrým pomocníkem při práci s daty. Neúplná či dokonce chybějící metadata jsou původcem nižší datové kvality. Obecně jsou metadata také data a mohou mít své vlastní charakteristiky (atributy nebo též deskriptory) a tak vznikají vyšší úrovně metadat. Hlavním úkolem metadat je podrobně popsat data. Slouží pro přesnou a korektní identifikaci a interpretaci dat. Metadata pro geodata by měla být organizována a spravována s využitím metainformačních systémů. Metainformační systém tak podporuje efektivní využívání samotných dat. Důležitou složkou metadat kromě popisu vlastních dat bývají i údaje o kontaktním místě (organizaci, osobě). Nejlepší variantou strukturovaných metadat je jejich zápis v mezinárodně uznávané normě.

4.2.1 NORMA ISO 19115

Norma ISO 19115 standardizuje metapopis prostorových dat. V současné době začíná nahrazovat místní standardy, zejména FGDC (USA), ANZLIC (Austrálie), CEN (EU). Stejně tak i výrobci softwaru implementují tuto normu do svých produktů pro podporu sémantické interoperability. V České republice se standard ISO 19115 stal ČSN (Českou státní normou) pro popis prostorových dat. Postupně nahrazuje národní metadatové standardy a stává se tak hlavním integrujícím činitelem pro standardizaci popisu prostorových dat. Oficiální dokument zahrnující plné znění normy je možné zakoupit na stránkách www.iso.org, český překlad najdete pod názvem ČSN ISO 19115.

Purpose:	Global Forest Resources Assessment update (FRA 2005) focuses on key trends and builds on the thematic elements of sustainable forest management, drawn from regional and ecoregional criteria and indicators processes, as a reporting framework. Thus, the information collated in the assessment is relevant to national monitoring of progress towards sustainable forest management and reporting to various forest-related international organizations and processes. FRA 2005 continues the FAO tradition of reporting on the world's forests. The periodic global assessment reports have tracked and reflected the changes that both the resource and forestry have undergone over the past 50 years. level				
Character set:	utf8				
Language:	English				
Status:	Complete				
Update frequency:	Annually				
Topic Category					
GEO theme:	Forests				
GEO data category:	Vegetation and Land Cover				
Theme Keywords					
GEMET theme:	Forestry				
GEMET keywords:	Nov 2005				
Free keywords:	Land cover, forestry, vegetation, forest extent, deforestation				
Temporal Extent					
Covered time:	1990, 2000, 2005				
Geographic Extent					
Coverage:	World				
West longitude:	-180	East longitude:	180	South latitude:	-90
				North latitude:	90

Obr. 2 Příklad metadat zapsaných v normě ISO 19115 (zdroj: [18])

4.2.2 INSPIRE

Iniciativa INSPIRE (Infrastructure for Spatial Information in Europe) se stala evropskou směrnicí po jejím schválení Evropským parlamentem. Vyšla 25. dubna 2007 a v platnost vstoupila 15. května 2007. Směrnice INSPIRE si klade za cíl vytvořit evropský legislativní rámec potřebný k vybudování evropské infrastruktury prostorových informací. Stanovuje obecná pravidla pro založení evropské infrastruktury prostorových dat zejména k podpoře environmentálních politik a politik, které životní prostředí ovlivňují. Hlavním cílem INSPIRE je poskytnout větší množství kvalitních a standardizovaných prostorových informací pro vytváření a uplatňování politik Společenství na všech úrovních členských států.

INSPIRE je založena na souboru základních principů:

- Data by měla být sbírána pouze jednou a udržována na té úrovni, kde je sbírání dat nejúčinnější.
- Mělo by být možné propojit prostorové informace z různých evropských zdrojů a mezi mnoho uživatelů aplikací.
- Mělo by být možné pro informace sebrané na jedné úrovni jejich sdílení do všech ostatních úrovní.
- Na všech úrovních by mělo být dostatečné množství geodat za podmínek, které umožní jejich rozsáhlé použití.
- Mělo by být snadné najít, která geodata jsou dostupná, která se hodí pro zvláštní použití a za jakých podmínek mohou být získána a používána.
- Geodata by měla být snadno pochopitelná a interpretovatelná.

V současné době je potřeba, aby byla směrnice začleněna do národní legislativy a aby vznikl plán, jak do roku 2013 splnit požadavky, které na každý členský stát EU klade text směrnice. Nově vytvořená evropská prostorová informační struktura by měla uživatelům zpřístupnit integrované prostorové informační služby. Tyto služby by měly uživatelům pomoci při vyhledávání a zpřístupňování geografických informací z široké škály zdrojů. Datové sady musí vyhovovat standardům, které zajistí propojení s daty z jiných zemí a z podkladů různých měřítek. INSPIRE skýtá výhody pro EU (jednotlivé státy), jakož i pro běžného uživatele. Pokud stát zpřístupní a otevře používání prostorových dat, umožní mimo jiné prostřednictvím internetu jejich kontrolu, např. zapsání osobního vlastnictví nebo s použitím mobilního telefonu nalezení nejbližšího bankomatu či autoservisu.

4.2.3 DUBLIN CORE

Dublin Core je soubor metadatových prvků, jehož záměrem je usnadnit vyhledávání elektronických zdrojů. Snaží se nalézt minimum souborů vlastností nezbytných pro podporu hledání v databázi obecně. Pracuje jak s prostorem, tak i s časem, avšak nestanovuje, jak by měly být popsány specifické vlastnosti, jakými jsou prostorová rozlišitelnost, projekce nebo datum. Uplatňuje princip **Light Metadata** (omezený soubor vlastností, jež je možné udržovat za relativně nízkou cenu a přitom jsou užitečné pro podporu výzkumu a objevování). Tato jednoduchá metadata poskytují stručný, ale užitečný popis datových sad, jejichž cena je navíc nižší. Na vytvoření českého standartu Dublin Core pracuje Knihovnicko-informační centrum Masarykovy university v Brně.

15 základních prvků metadatových standardů Dublin Core:

- Název (Title), jméno zdroje zadané tvůrcem nebo vydavatelem;
- Tvůrce (Creator), osoba nebo organizace primárně zodpovědná za vytvoření obsahu zdroje;
- Předmět (Subject), předmět, klíčové slovo nebo fráze popisující předmět nebo obsah zdroje;
- Popis (Description), textový popis obsahu zdroje;
- Vydavatel (Publisher), entita zodpovědná za zpřístupnění zdroje v jeho současné podobě;
- Příspěvatel (Contributor), entita, která přispěla k obsahu zdroje;
- Datum (Date), datum, kdy byl zdroj zpřístupněn v současné podobě;
- Typ (Type), povaha nebo žánr (druh) obsahu zdroje, např. webová stránka, báseň, technická zpráva, slovník;
- Formát (Format), datová reprezentace zdroje, např. HTML, JPG, MDB;
- Identifikátor (Identifier), unikátní číslo nebo textový řetězec používaný k identifikaci zdroje;
- Zdroj (Source), odkaz na zdroj, ze kterého je popisovaný zdroj odvozen;
- Jazyk (Language), jazyk intelektuálního obsahu zdroje;
- Vztah (Relation), vztah k dalším zdrojům;
- Pokrytí (Coverage), rozsah (prostorové umístění) zdroje;
- Práva (Rights), informace o právech vztahujících se k popisovanému zdroji.

4.3 GEOGRAFICKÝ INFORMAČNÍ SYSTÉM

Geografický informační systém (GIS) je na počítači založený systém používaný na získávání, ukládání, upravování, zobrazování a vykreslování geografických dat. Pro uplatnění GIS v geografii uvádí Voženílek (1998) následující definici: „GIS je organizovaný, počítačově založený systém hardwaru, softwaru a geografických informacích vyvinutý ke vstupu, správě, analytickému zpracování a prezentaci prostorových dat s důrazem na jejich prostorové analýzy.“ GIS se většinou skládá ze čtyřech hlavních částí:

- **Hardware** (počítače, vstupní a výstupní periferie);
- **Software** (programové vybavení, zastává všechny systémové operace);
- **Data** (nejdůležitější, zpravidla nejdražší);
- **Lidé** (řídící funkce).

4.4 DATABÁZE

Databáze je určitá uspořádaná množina informací uložená na paměťovém médiu. Strukturovaná data se do ní ukládají v souladu s určitým **datovým modelem** a **datovou strukturou**. Pro reprezentaci dat v databázi používáme vhodný datový model. Donedávna byly používány tři modely: hierarchický, síťový a relační. Většina aplikací dnes využívá relační datový model vycházející z matematického pojetí relace a relační logiky. Jeho autorem je Edgar F. Codd, publikován byl již v roce 1970 (Farana, 2006).

Jednou ze základních částí každého geoinformačního softwaru je **geografická databáze**. Od běžné databáze se liší tím, že navzájem spojuje grafická² a negrafická data³.

4.4.1 DBMS

Přístup k údajům uloženým v databázi obstarává program DBMS (DataBase Management System), jenž je mezičlánkem mezi daty a aplikačním programem. Spravuje data a stará se o fyzické uložení dat. Dále zajišťuje sdílení dat nejefektivnějším způsobem tak, aby nebyla narušena integrita databáze. Řeší víceuživatelský, paralelní přístup k datům s možností nastavení uživatelských práv. Mezi DBMS patří například programy Oracle,

² Grafická data popisují objekty grafického charakteru tak, aby je bylo možno zobrazovat na výstupních zařízeních, která mohou být podle způsobu kódování obrazu do čísel bitmapová nebo vektorová.

³ Negrafická data (atributová data) popisují negeometrickou vlastnost dané entity.

MS SQL Server, Informix či InterBase. Převážná většina dnes používaných DBMS při uspořádání údajů v databázi vychází z relačního modelu dat (Dobešová, 2004).

4.4.2 DATOVÝ TYP

Každý záznam v databázi je určitý datový typ. Databázové systémy se liší počtem a druhy datových typů. Nejběžnějšími základními datovými typy jsou:

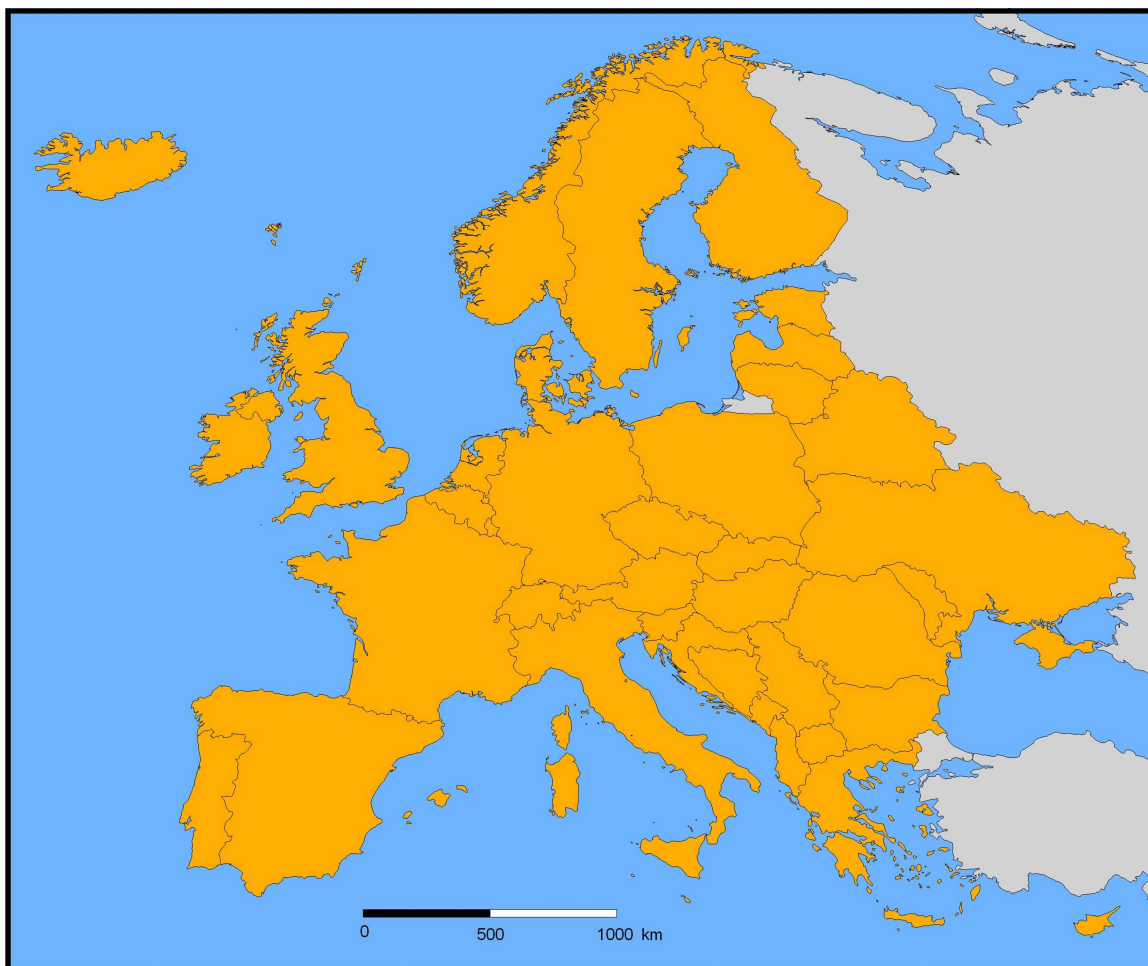
- **Textový** (text o maximální délce 255 znaků);
- **Číselný** (bajt, celé číslo, dlouhé celé číslo, jednoduchá přesnost, dvojitá přesnost, desetinné číslo);
- **Datumový** (date);
- **Logický** (logic, boolean, ano/ne).

4.4.3 DATOVÁ SADA

V praxi se mnohdy setkáváme s výrazem „datová sada“ (angl. dataset). Data se často sdružují do kolekcí, které nazýváme datové sady nebo datové soubory (angl. data files). Pojem „datová sada“ je obecnější oproti pojmu „datový soubor“. Datová sada obvykle představuje data tvořící logický celek v rámci určitého informačního systému či datové báze. Může se tedy jednat o jeden datový soubor či kolekci těchto souborů. Pojem „datová sada“ nebývá omezován jen na digitální data, ale může jím být označena i analogová forma dat (např. atlas). Volně dostupná geodata lze obvykle stáhnout ve formě datasetů. Komprimovaný soubor často obsahuje několik složek, např. tabulková data, mapy, grafy.

5 VYMEZENÍ ZÁJMOVÉHO ÚZEMÍ

Na začátku bakalářské práce bylo třeba definovat zájmové území, protože pojem „Evropa“ by mohl být nepřesně interpretován. Do zájmového území náleží všechny evropské státy s výjimkou Ruska, jelikož jeho větší část se nachází v Asii. Přesná hranice zájmového území je na severu omezena Severním ledovým oceánem, na západě Atlantským oceánem, na jihu Středozemním mořem, na východě západní hranicí Turecka a Ruska. Do zájmového území náleží i členské státy Evropské unie nacházející se ve Středozemním moři, Malta a Kypr. Celkem se jedná o 43 států. Do tohoto počtu není zahrnut církevní stát Vatikán ani nově vzniklý stát Kosovo, jelikož jeho suverenita není jednohlasně uznávána.



Obr. 3 Vymezení zájmového území

6 ZDROJE DAT

Ve snaze získat volně dostupná geodata byl učiněn průzkum Internetu, získaná data uložena na pevný disk a popsána v databázi (příloha bakalářské práce). Nároky na takto získaná data byly následující: a) **volně dostupná a legálně stažitelná**, tomuto požadavku nevyhovují zpoplatněná data, ani data získaná na vyžádání přes email, protože nelze s jistotou tvrdit, že by byla v budoucnu poskytnuta všem zájemcům; b) **prostorový aspekt dat musí odpovídat zájmovému území**, nelze použít data pro část území; c) Nutností je, aby se dala takto získaná data využít pro tvorbu tematických map, tj. aby byla **vhodně uložena** – např. ve formě tabulek či mapových vrstev (např. SHP); d) **důvěryhodný zdroj**, data není přípustné stahovat z neověřených a nedůvěryhodných adres.

6.1 EUROSTAT

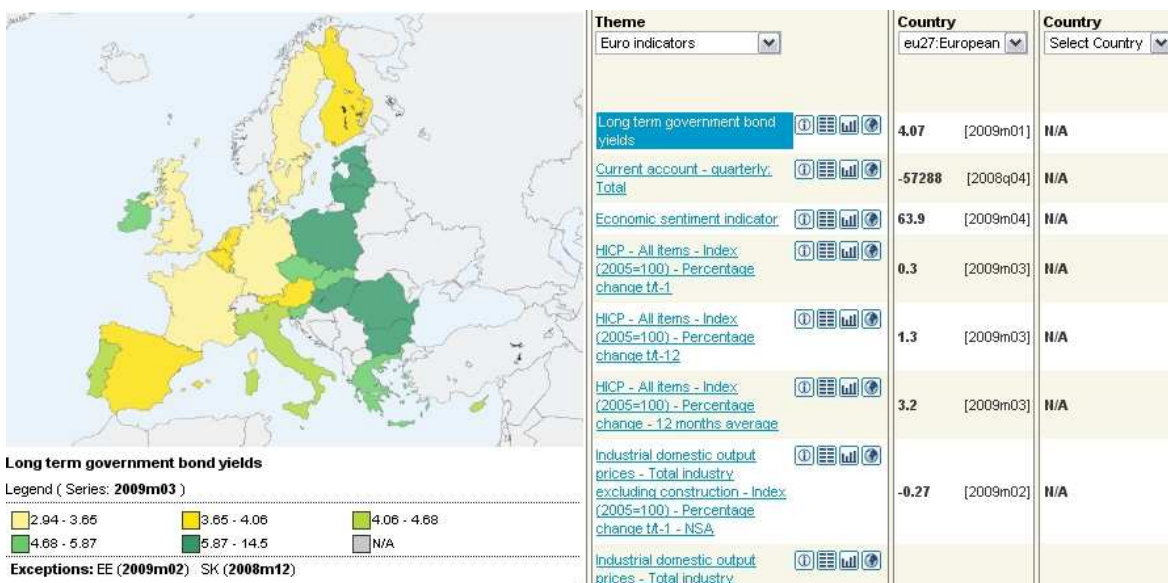
Eurostat (nebo také Evropský statistický úřad) je statistické oddělení Evropské komise produkující data pro Evropskou unii, jehož sídlo se nachází v Lucemburku. Na webové stránce <http://epp.eurostat.ec.europa.eu> je dostupné velké množství aktuálních dat (včetně metadat ve formátu HTML), která jsou dle tématu rozříděna do 9 základních kategorií a 35 podkategorií. K výběru datasetu je možné se dostat klepnutím na záložku „Statistics“ nebo pomocí tlačítka „Statistics Database“ v levé horní části webové stránky.

Data mohou být zobrazena v několika podobách. Ve formě tabulky, jako sloupcový graf nebo jako mapa (kartografickou vyjadřovací metodou je vždy jednoduchý kartogram). Uživatel má možnost jejich volného stažení v následujících formátech: XLS, HTML, XML nebo TSV. Nabízí se rovněž možnost stažení mapy a informačního HTML souboru ve formátu ZIP. Zobrazená mapa je v nízkém rozlišení a slouží především k jednoduchému nahlédnutí na zájmovou oblast, v žádném případě není vhodná pro tisk, nehledě na nemožnost jakýchkoliv úprav (formát PNG). To samé platí i pro graf (také ve formátu PNG). Zobrazením mapy, případně grafu, dostáváme základní informaci o množství dat a o jejich prostorovém rozložení. Jedná se tedy o velmi dobré vodítko pro rozhodování o tom, zda-li data stáhnout, či dát přednost jiné alternativě. K praktickému porovnávání jednotlivých datasetů slouží záložka „Country profiles“.

Účelu bakalářské práce nejlépe vyhovují data ve formátu XLS, která lze po lehké úpravě snadno propojit s mapovou částí a v prostředí GIS efektivně vytvářet mapy dle vlastního uvážení.

The screenshot shows the Eurostat website interface. At the top, there is a navigation bar with 'Home', 'Statistics', 'Publications', 'About Eurostat', and 'Help'. The 'Statistics' menu item is circled in red. On the left sidebar, there is a 'Statistics Database' icon, also circled in red. Below the navigation bar, there are sections for 'Latest news releases', 'Most popular database tables', and 'Selected Statistics'.

Obr. 4 Úvodní stránka Eurostatu (červeně označena cesta k výběru datasetu)



Obr. 5 Zobrazení dat v mapové podobě na Eurostatu

6.2 UNITED NATIONS ENVIRONMENT PROGRAMME – GEO DATA PORTAL

UNEP (United Nations Environment Programme; Program OSN pro životní prostředí) vznikl v roce 1972 pod správou OSN. Jeho posláním je stimulovat a koordinovat akce na ochranu životního prostředí především na mezinárodní úrovni a poskytovat podklady pro rozhodování příznivé životnímu prostředí. Prioritně se zaměřuje na monitoring životního prostředí, hodnocení vlivů, výzkum, shromažďování a rozšiřování informací, včetně systémů včasného varování. Mezi další projekty UNEP patří

mezinárodní informační systém INFOTERRA (www.unep.org/infoterra) a regionální systémy monitorování a vyhodnocování stavu životního prostředí – GEMS (www.ec.gc.ca), EARTHWATCH (<http://earthwatch.unep.net>), GRID (<http://na.unep.net>) a IRPTC (www.chem.unep.ch).

Online databáze geografických dat dostupná na adrese <http://geodata.grid.unep.ch> obsahuje velké množství datasetů ve třech základních formátech: tabulky, grafy a mapy. To vše s podrobnými metadaty dle normy ISO 19115. Vyhledávání funguje na základě klíčových slov, pod kterými jsou datasety uloženy. Mimo specifikace tematického okruhu (klíčového slova) je nutné věnovat pozornost i výběru geografické úrovně. Nejlepší a nejčastější volbou je bezpochyby členění po státech („National“), další možností je dělení regionální („Regional“, po kontinentech) a subregionální („Subregional“, větší územní celky v rámci kontinentů). Užitečná je mapová vizualizace, pro základní představu o datech. Pro stažení datasetů označených jako „Download protected“ je potřeba znát přihlašovací jméno a heslo. Volně stažitelná data jsou nejčastěji uložena ve formátech PDF, CSV, XLS a SHP. Nutno dodat, že u každého datasetu se liší dostupné formáty (např. není dostupný formát SHP) z důvodu nestejného poskytovatele dat. Nejčastějšími poskytovateli jsou UN, WB, FAO, IUCN/SSC, IPCC a OFDA. Data jsou pravidelně aktualizována (ve většině případů každý rok) a pokrývají celý svět. Pro účely bakalářské práce je nejvhodnější stahovat formát SHP (pokud je k dispozici), případně XLS, a ručně odfiltrovat státy, jež neodpovídají zájmovému území.

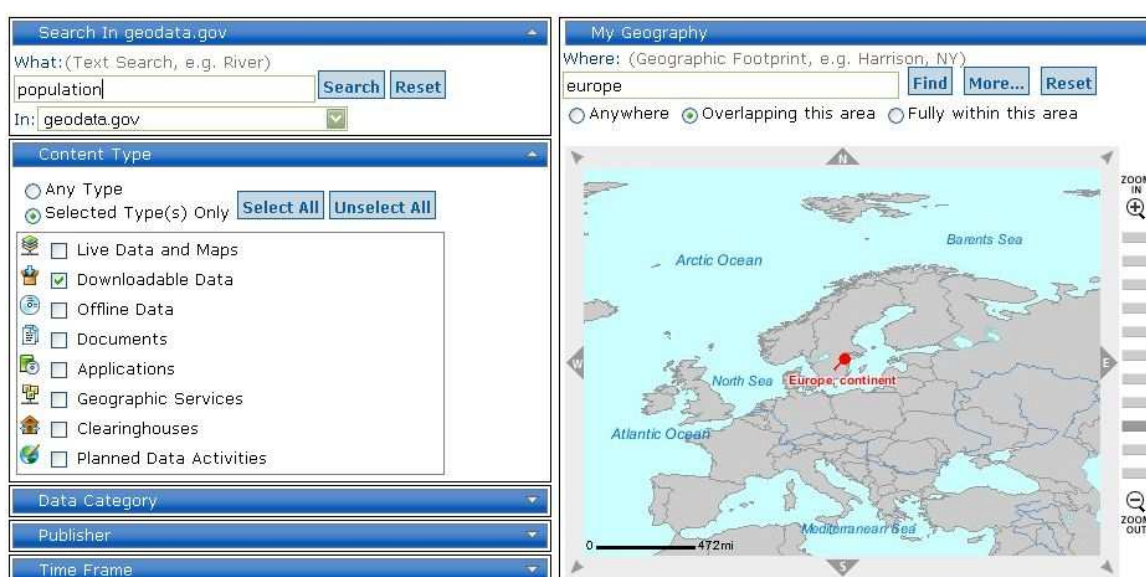
Number of entries found: 74

Variable	Data Set Type	GEO Region	Covered Time	Download protected?	Data Provider	Last updated	Preview
<input type="radio"/> Births	National	World	1950-55, 1955-60, ..., 2045-50	<input type="checkbox"/>	UN Population Division	04/2007	<input type="checkbox"/>
<input type="radio"/> Crude Birth Rate	National	World	1950-55, 1955-60, ..., 2045-50	<input type="checkbox"/>	UN Population Division	05/2007	<input type="checkbox"/>
<input type="radio"/> Crude Death Rate	National	World	1950-55, 1955-60, ..., 2045-50	<input type="checkbox"/>	UN Population Division	05/2007	<input type="checkbox"/>
<input type="radio"/> Diseases of the Respiratory System - Number of Deaths per 100,000 Population	National	World	1979-2003	<input type="checkbox"/>	WHO	05/2005	<input type="checkbox"/>
<input type="radio"/> Economically Active Population	National	World	1980-2020	<input type="checkbox"/>	ILO	11/2007	<input type="checkbox"/>
<input type="radio"/> Enrolment in Tertiary Education	National	World	1999-2005	<input type="checkbox"/>	UNESCO	06/2006	<input type="checkbox"/>
<input type="radio"/> Fertility	National	World	1950-55, 1955-60, ..., 2045-50	<input type="checkbox"/>	UN Population Division	05/2007	<input type="checkbox"/>
<input type="radio"/> Gender Empowerment Measure (GEM)	National	World	2003	<input type="checkbox"/>	UNDP	03/2006	<input type="checkbox"/>
<input type="radio"/> Gender-related Development Index (GDI)	National	World	2003	<input type="checkbox"/>	UNDP	03/2006	<input type="checkbox"/>
<input type="radio"/> Girls to Boys Ratio in Primary Education Enrolment	National	World	1990, 1998-2001	<input type="checkbox"/>	UNSD	03/2006	<input type="checkbox"/>
<input type="radio"/> Gross Domestic Product - per Capita	National	World	1960-2005	<input type="checkbox"/>	World Bank	09/2007	<input type="checkbox"/>
<input type="radio"/> Gross Enrolment Ratio in Primary Education	National	World	1999-2005	<input type="checkbox"/>	UNESCO	05/2006	<input type="checkbox"/>
<input type="radio"/> Gross Enrolment Ratios in Secondary Education	National	World	1999-2005	<input type="checkbox"/>	UNESCO	06/2006	<input type="checkbox"/>
<input type="radio"/> Human Development Index (HDI)	National	World	1975, 1980, ..., 2000-2003	<input type="checkbox"/>	UNDP	03/2006	<input type="checkbox"/>
<input type="radio"/> Human Poverty Index for Developing Countries (HPI-1)	National	World	2003	<input type="checkbox"/>	UNDP	03/2006	<input type="checkbox"/>
<input type="radio"/> Improved Drinking Water Coverage - Rural Population	National	World	1990, 1995, 2000, 2004	<input type="checkbox"/>	WHO	11/2006	<input type="checkbox"/>
<input type="radio"/> Improved Drinking Water Coverage - Rural Population with Household Connection	National	World	1990, 1995, 2000, 2004	<input type="checkbox"/>	WHO	11/2006	<input type="checkbox"/>
<input checked="" type="radio"/> Improved Drinking Water Coverage - Total Population	National	World	1990, 1995, 2000, 2004	<input type="checkbox"/>	WHO	11/2006	<input type="checkbox"/>
<input type="radio"/> Improved Drinking Water Coverage - Total Population with Household Connection	National	World	1990, 1995, 2000, 2004	<input type="checkbox"/>	WHO	11/2006	<input type="checkbox"/>
<input type="radio"/> Improved Drinking Water Coverage - Urban Population	National	World	1990, 1995, 2000, 2004	<input type="checkbox"/>	WHO	11/2006	<input type="checkbox"/>
<input type="radio"/> Improved Drinking Water Coverage - Urban Population with Household Connection	National	World	1990, 1995, 2000, 2004	<input type="checkbox"/>	WHO	11/2006	<input type="checkbox"/>

Obr. 6 Výběr datasetu na geo-portálu UNEP

6.3 GOS – GEOSPATIAL ONE STOP

Geografický portál GOS – Geospatial One Stop provozovaný na internetové adrese <http://gos2.geodata.gov/> je zdrojem rozličných datasetů nabízených v řadě různých formátů. Web funguje jako vyhledávač datových sad s tím, že odkazuje na lokace, kde již data mnohdy nejsou dostupná. Volně lze stáhnout satelitní snímky, mapy, data ve formátu SHP a XLS. Specializace na USA není na závadu, největší potíží je dopracovat se ke kýženým datům, jelikož vyhledávání je pomalé a souboje s nefunkčními odkazy ztrpčují jinak příjemný pobyt na tomto serveru.



Obr. 7 Vyhledávání podle atributů (vlevo) a prostorového určení (vpravo)

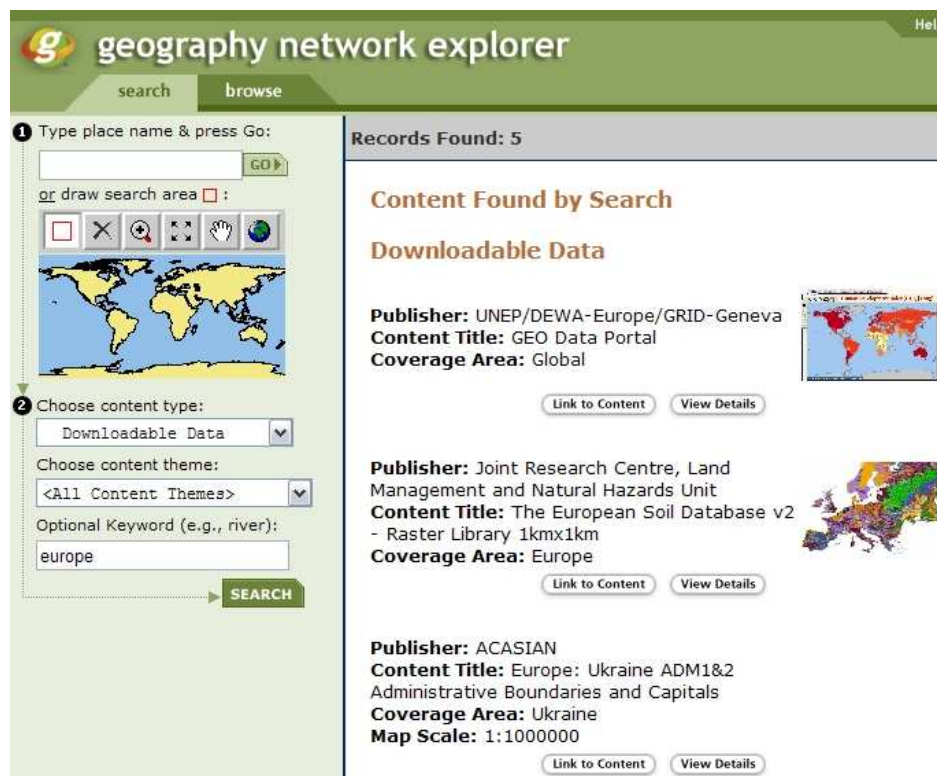
6.4 DALŠÍ MOŽNÉ ZDROJE

Mimo tří výše uvedených zdrojů se nabízela eventuelní možnost stahovat data i z dalších lokací. Nabídka se však střetla s požadavky uvedenými na začátku této kapitoly a tudíž data nebylo možné využít. Pro ilustraci uveďme několik zdrojů s připojením stručného komentáře.

6.4.1 GEOGRAPHY NETWORK

Geography Network reprezentoval slibně vypadající zdroj nacházející se na adrese <http://geographynetwork.com>, který však trpěl hned několika nedostatky. V první řadě fungoval jako rozcestník odkazující na adresy, odkud je možné data stáhnout. Tyto adresy měly značně rozdílný charakter – od propracovaných databází (kapitola 6.2), přes různé geoinformační portály až po nepřehledné až pochybné weby. Data bylo možné často pouze

zobrazit v ArcExploreru, jenž byl součástí prohlížeče. Zájmová oblast datasetů byla různorodá: Svět, Evropa, různě vymezené regiony (především v USA), jednotlivé státy, města, oblasti. Data byla ve většině případů zpoplatněná nebo dostupná pouze e-mailem na vyžádání. Na úvodní stránce je psáno: „Some data may be downloaded for free while other data may be ordered through the publisher.“ První možnost bohužel nesplnila očekávání, druhá nevyhovovala zadání. Bylo třeba poohlédnout se po dalších zdrojích.



Obr. 8 Prostředí vyhledávače Geography Network

6.4.2 GEOCOMMUNITY

Server GeoCommunity (<http://data.geocomm.com>) funguje od roku 1995 a v současné době má více než 41 000 uživatelů, kteří aktivně využívají datasety určené speciálně pro GIS. Ty jsou uspořádány po jednotlivých státech, navíc za poplatek 0,1-3 USD za stát. Vytvoření jedné tematické mapy, řekněme pro třicet států, by bylo velmi pracné a nákladné. Zastoupeno je malé množství tematických okruhů, zdaleka ne všechny evropské státy mají data (dostupné pouze pro následující státy: Belgie, Dánsko, Francie, Itálie, Německo, Nizozemí, Norsko, Rakousko, Španělsko, Švédsko, Švýcarsko a Velká Británie). Web je zaměřen na profesionální uživatele především z USA.



Obr. 9 Logo GeoCommunity

6.4.3 CIA – THE WORLD FACTBOOK

Statistická ročenka CIA – The World Factbook, kterou najdete na adrese (<https://www.cia.gov/library/publications/the-world-factbook/>) zpřístupňuje statistická data geografického charakteru o jednotlivých státech světa. V knižní podobě vychází od roku 1981, počínaje rokem 2008 je dostupná pouze v elektronické formě (zdarma ke stažení). Data jsou tříděna po státech. Výhodou je, že zde lze najít velmi podrobné údaje, které jsou navíc často aktualizovány (aktuální rok). U nezjištěných hodnot je uváděn odhad (angl. est.). Nevýhoda spočívá ve formátu HTML – data nelze stáhnout, nýbrž pouze zobrazit. Pro vytvoření jedné tematické mapy lze tento zdroj doporučit, pro sériovou výrobu už nikoliv, protože přepisování hodnot do tabulky je časově náročné a mohou při něm vznikat chyby.

People	Belgium	Top of Page
Population:	10,414,336 (July 2009 est.)	
Age structure:	0-14 years: 16.1% (male 857,373/female 822,303) 15-64 years: 66.3% (male 3,480,072/female 3,419,721) 65 years and over: 17.6% (male 760,390/female 1,074,477) (2009 est.)	
Median age:	total: 41.7 years male: 40.4 years female: 43 years (2009 est.)	
Population growth rate:	0.094% (2009 est.)	
Birth rate:	10.15 births/1,000 population (2009 est.)	
Death rate:	10.38 deaths/1,000 population (2008 est.)	
Net migration rate:	1.22 migrant(s)/1,000 population (2009 est.)	
Urbanization:	urban population: 97% of total population (2008) rate of urbanization: 0.3% annual rate of change (2005-10 est.)	
Sex ratio:	at birth: 1.04 male(s)/female under 15 years: 1.04 male(s)/female 15-64 years: 1.02 male(s)/female 65 years and over: 0.71 male(s)/female total population: 0.96 male(s)/female (2009 est.)	

Obr. 10 CIA – The World Factbook zpřístupňuje data pouze ve formátu HTML

6.5 HODNOCENÍ STAŽENÝCH DAT

Celkem bylo staženo, prozkoumáno a popsáno 60 datových vrstev. Vybírány byly takové vrstvy, které zachycují vhodné jevy a ze kterých je vhodné tvořit tematické mapy. Následné hodnocení se bude zabývat právě těmi 60 vrstvami, které tak tvoří reprezentativní vzorek volně stažitelných geodat.

Zaměříme se nejprve na problém dostupnosti. Vzhledem k neustálým změnám na webu Eurostatu nebyl umožněn nepřetržitý přístup k datům. Současná podoba stránek existuje teprve od konce dubna 2009 a právě během přechodu na novou grafiku bylo statistické oddělení často nedostupné. Oproti tomu geo-portál UNEP představuje stabilní online databázi, která je bez vážnějších problémů provozována již od roku 2006. Avšak ani tento často aktualizovaný server není bez chyb. V jednom případě (33_plantae_threatened) nebyl k dispozici formát SHP, nýbrž pouze XLS. Na tuto skutečnost nebylo nikde upozorněno. Převažujícím tématem obou datových zdrojů byla socioekonomická sféra (45 vrstev z celkových 60). Velká pozornost byla věnována zejména populaci, znečištění ovzduší a ekonomickým ukazatelům. Z fyzické geografie byly oblíbeným tématem ohrožené druhy rostlin a živočichů, vodní zdroje a především nejrůznější přírodní katastrofy od zemětřesení až po lesní požáry. Datasetsy byly shledány přijatelně přesnými pro tvorbu dostatečně kvalitních tematických map. Vzhledem k tomu, že sběr dat byl v každém případě řízen dostatečně relevantní nadnárodní společností, uživatel může důvěřovat jejich obsahu. Žádné markantní odchylky nebyly zjištěny.



Sorry, but we could not process your request!

Reason:

- Application error occurred

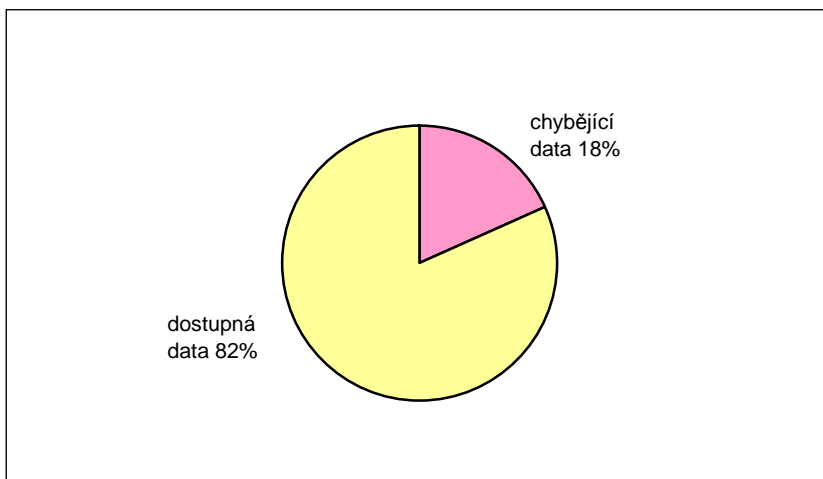
Obr. 11 Nedostupná data na Eurostatu

Nejvíce vadila neúplnost. Pro řadu států data zcela chyběla, a to i opakovaně. Ze zkoumaných 60 vrstev více než čtyřicetkrát chyběla data pro Andorru, Černou Horu, Monako a San Marino. Podíly chybějících dat podle zdroje jsou přehledně znázorněny na obr. 13-14; celkové počty chybějících údajů pro jednotlivé státy vyjádřené grafem obsahuje příloha 11. Časové řady nebyly úplné, občas byl jeden nebo více roků vynechán,

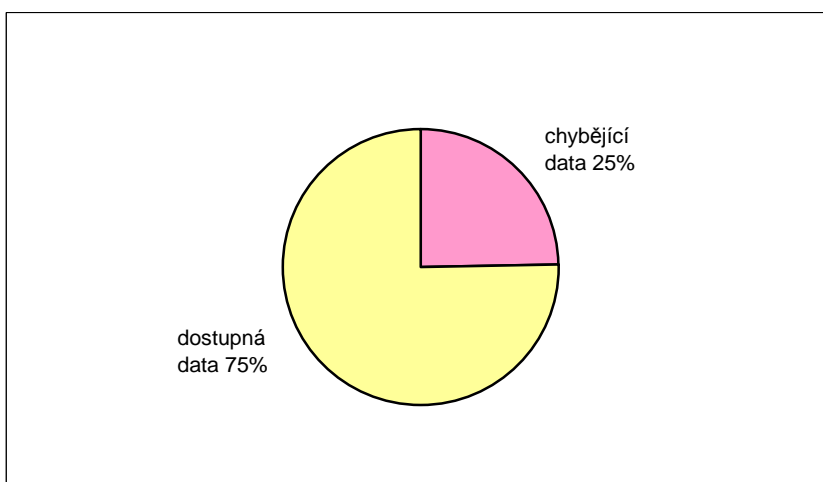
uživatel měl možnost se setkat i s časovým označením „Various“ (tzn. není jasné, ze kterého roku údaj pochází). Data jsou často aktualizována (týká se i metadat) a k existujícím datasetům přidávány aktuální roky. Pokud nebereme v úvahu prognózy, nejnovější data byla za rok 2007, nejstarší za rok 1958. Nutno dodat, že u starších roků zpravidla chybí data pro velké množství států. Tabulková struktura dat ve formátu XLS nebyla ve všech případech zcela korektní. V některých tabulkách (26_tourism arrivals, 30_expend_military, 31_land_wooded_extent, 34_precipitation_mm) se vyskytovaly duplicitní informace, jež bylo potřeba dodatečně odstranit. Datasetsy z UNEP obsahovaly i volné řádky pro závislá území. Jednalo se o tato území: Faerské ostrovy, Gibraltar, Guernsey, Holy See, Isle of Man, Jersey, Špicberky a Jan Mayen. Bohužel ani jednou nebylo možné odečíst žádný údaj – jejich zastoupení v databázi bylo pro účely bakalářské práce zcela zbytečné. Je s podivem, že UNEP řadí do evropského regionu Arménii, Gruzii, Izrael, Rusko a Turecko. O datasetech úspěšně stažených z geo-portálu GOS – Geospatial One Stop se dá hovořit jediné pozitivně. Číselné údaje byly pro většinu států zjištěny, metadata integrovaná v XLS poskytla bezproblémovou orientaci v množství zkratk a čísel.

Name	Sovereign	iso_2_code	iso_3_code	Un_code	Geo_region	Geo_subreg	Region
Belgium	Belgium	BE	BEL	56.0000000000	1003.0000000000	1016.0000000000	Europe
Bosnia and Herzegovina	Bosnia and Herzegovina	BA	BIH	70.0000000000	1003.0000000000	1014.0000000000	Europe
Bulgaria	Bulgaria	BG	BGR	100.0000000000	1003.0000000000	1014.0000000000	Europe
	1319	444	Ita				IT ITA
	-9999	464	d				
	539	476	i				KZ KAZ
312	1287	500					LV LV
437	432	422	576				YT
-9999	-	-9999	60				MN
6167	5589	6					MZ
17060	15342						NA
-9999	-9999						
393	324						
-9999	-9999						
178	161						
-9999	-9999	999	-9999	cipe			
4424	4373	4264	4269				
1452	1424	1371	1345				
116	11	104	100	and Jan Mayen I	lands		
210	21	191	185				
1685	1576	18	nd				
30	31	3	nidad and Tobago				
26727	24557		Wake Isla	nd			erica
Liechtenstein	Liechtenstein	LI	LIE	438.0000000000	1003.0000000000	1016.0000000000	Europe
Lithuania	Lithuania	LT	LTU	440.0000000000	1003.0000000000	1014.0000000000	Europe
Luxembourg	Luxembourg	LU	LUX	442.0000000000	1003.0000000000	1016.0000000000	Europe
Malta	Malta	MT	MLT	470.0000000000	1003.0000000000	1016.0000000000	Europe

Obr. 12 Výskyt duplicitních informací v atributových datech



Obr. 13 Podíl chybějících dat získaných z geo-portálu UNEP [19]



Obr. 14 Podíl chybějících dat získaných z Eurostatu [18]

7 KVALITA DAT

Jedním z nejdůležitějších a zároveň nejproblematičtějších faktorů volně dostupných geodat je jejich kvalita. Voženílek (2008) tvrdí, že rozvoj geoinformatiky v posledních letech postavil především geografii jako vědní disciplínu před velmi složité téma: na jedné straně je k dispozici podstatně více dat než tomu bylo dříve, ale na druhé straně je problém relevance těchto dat.

Kvalita dat a informací je problém, jehož složitost je často neprávem podceňována. Jedná se o vícedimensionální pojem, protože má i takové aspekty, jako je včasnost, dostupnost a důvěryhodnost. Stále více pracuje s daty a informacemi, která nemusí být přesná či věrohodná. Rozvoj internetu tento problém velmi zesílil. Počet připojených uživatelů vzrůstá každým dnem, každý uživatel může nahrávat data rozdílné kvality. A tu poznáme často až při práci s daty.

V aplikacích se mnohdy mlčky chybně předpokládá, že potřebná data jsou vždy k dispozici v dostatečné kvalitě. V poslední dekádě význam kvality dat značně zesílil. Důvodem je potřeba podpory managementu a rozvoj manažerských informačních systémů. Ty musí využívat data a informace různé kvality dostupné např. na internetu. Problémy kvality dat jsou tématem řady konferencí, pokusů o standardizaci (ISO) a dokonce i zákonů (americký zákon Data Quality Act). Kvalita dat a informací není u nás dostatečně legislativně ošetřena. To je zásadní nedostatek, který do značné míry vylučuje možnost vyžadování, sledování, a kontrolu kvality dat a informací. To neumožňuje obranu proti takovým opatřením, které kvalitu dat a informací, především dostupnost, zbytečně zhoršují (Kráal, 2005).

Kvalita dat je poměrně nový problém, jehož formulace a cesty řešení se rychle vyvíjejí. Není ustálen názor, do jaké míry mají být metriky kvality závislé na potřebách jednotlivých aplikací. Jedním z problémů volby atributů kvality dat je volba optimálního kompromisu mezi obecností a použitelností. Kvalita informací se odvozuje od kvality dat, jedná se ale o specifický problém, který je předmětem intenzivního výzkumu. Typickou úlohou je řešení problému, jaká je kvalita informací využívajících data různé kvality.

Kvalitní data by měla splňovat především následující atributy:

- Přesnost (Accuracy)
- Úplnost (Completeness)
- Včasnost (Timeliness)

- Objektivnost (Objectivity)
- Důvěryhodnost (Believability)

Uživatel by neměl opomíjet ani další faktory, jakými jsou:

- Reputace (Reputation)
- Dostupnost (Accessibility)
- Bezpečnost přístupu (Access security)
- Kontextuální (Contextual)
- Přínos (Value added)
- Rozsah (Amount of data)
- Reprezentační (Representational)
- Interoperabilita (Interoperability)
- Srozumitelnost (Easy of Understanding)

7.1 ZLEPŠOVÁNÍ DATOVÉ KVALITY

„Zlepšování datové kvality (Data Quality Improvement) je proces zvyšování kvality dat na úroveň požadovanou pro podporu informačních potřeb organizace“ (Pirkl, 2004). Jádrem zlepšování kvality dat je založeno na poměrně sofistikovaných metodách vyvinutých v rámci matematické statistiky. Nejčastěji používané postupy jsou:

7.1.1 VYLUČOVÁNÍ OKRAJOVÝCH DAT

Jde o postup, kdy se ze souboru vylučují data, která jsou zjevně nesprávná: úmyslně změněná, chybně změněná (překlepy), chybně změněná. Jestliže například v dostatečně velkém souboru existuje číslo x , které splňuje podmínku, že $x \gg M_{n-1} + 3\sigma_{n-1}$, kde M_{n-1} je průměr a σ_{n-1} směrodatná odchylka souboru, ze kterého bylo vyloučeno x , pak je lépe x ze souboru vyloučit. Existuje na to velmi rozvinutá teorie a postupy, které se používají především při dolování dat a také při zpracování výsledků měření v přírodních a technických vědách.

7.1.2 DOPLŇOVÁNÍ CHYBĚJÍCÍCH DAT

V tomto případě se do souboru doplní chybějící data, aby bylo možno soubor rozumně zobrazovat (například časové řady) a přitom nedošlo k chybným výsledkům (k

významným změnám charakteristik daného souboru). K dispozici jsou dávno existující algoritmy.

7.1.3 VYLOUČENÍ DUPLICITNÍCH DAT

Tato operace zahrnuje sjednocení formátů, vylučování dat s nevhodným formátem a parciální replikaci. Pokud se data používají pouze pro statistické analýzy, lze soubory dat replikovat pouze částečně (aniž dojde k závažnější chybě). To, jak velké procento dat je třeba replikovat, závisí především na rozptylu dat a kvalitě algoritmu, který provádí výběr dat k replikaci. Úspory mohou být dramatické.



Obr. 15 Jednotlivé kroky vedoucí ke zvyšování datové kvality (podle Pirkl, 2004)

7.2 HODNOCENÍ KVALITY DAT

Proces vyhodnocování kvality dat do značné míry závisí na potřebách konkrétního uživatele a dokonce na potřebách konkrétní aplikace. Různí uživatelé kladou rozdílné nároky na kvalitu dat. Existují však obecně stanovené metriky sloužící k hodnocení datové kvality; je vhodné rozlišovat metriky dvojího typu:

7.2.1 OBJEKTIVNÍ METRIKY

Jedná se o metriky, které lze vždy znovu vypočítat z dat, kterých se týkají. Objektivní metriky jsou obvykle číselné. Mezi objektivní metriky patří takové vlastnosti, jako existence okrajových či nevalidních dat, chybějící data, atd.

7.2.2 SUBJEKTIVNÍ METRIKY

Subjektivní metriky jsou metriky hodnotící způsob, jakým data vznikla, případně kvalitu zdroje dat. Subjektivní jsou např. metriky hodnotící důvěryhodnost dat, stupeň jejich utajení, dostupnost, atd. Subjektivní metriky odpovídají metrikám interním (angl. in process metrics, např. doba řešení, pracnost)

Proces zjišťování subjektivních metrik je nutno standardizovat. To je většinou zajišťováno předpisy, které specifikují atributy kvality dat, a postupy, které je nutno při sběru dat a při jejich „čištění“ dodržovat. Přívlastek „subjektivní“ má v případě metrik kvality dat jisté oprávnění, poněvadž tyto metriky většinou nevznikají měřením nějakého procesu, ale de facto subjektivním hodnocením vlastností dat experty, založeném na zkušenostech a nikoliv na měření v běžném slova smyslu. Z definice tohoto typu metrik plyne, že při vymezení typů subjektivních metrik závisí na potřebách konkrétního uživatele a někdy dokonce jediné konkrétní aplikace.

Hranice mezi subjektivními a objektivními metrikami není striktní. Pokud máme dostatečně rozsáhlý soubor, můžeme jeho střední hodnotu a směrodatnou odchylku vypočítat. V opačném případě (např. při problémech s dostupností dat) můžeme použít i kvalifikovaný odhad, tj. postupovat jako v případě subjektivních metrik. Fakt, že se takto postupovalo, by měl být zaznamenán.

8 DATABÁZE

8.1 PROSTŘEDÍ MICROSOFT ACCESS

Přílohou bakalářské práce je databáze pro správu metadat navržená v programu Microsoft Access popisující data, která byla stažena z výše uvedených zdrojů. Microsoft Access byl zvolen proto, že má všechny vlastnosti klasického systému pro správu relačních databází. Zajišťuje veškeré funkce pro definici dat, manipulaci s daty a řízení dat, které jsou potřeba pro správu velkých objemů dat. Access je nejen výkonný, flexibilní a snadno použitelný DBMS, ale také kompletní prostředek pro vývoj databázových aplikací. V operačním systému Microsoft Windows lze vytvářet tabulky, dotazy a formuláře pro prohlížení a změny dat. Samozřejmostí je podpora SQL.

8.2 POPIS DATABÁZE

Databáze nacházející se na CD-ROM pod názvem **Free_geodata.mdb** se skládá ze tří tabulek. Hlavní tabulkou je **t_GEODATA** obsahující veškeré informace o všech 60 datových vrstvách. Vedlejší tabulky **t_Metadata_Author** a **t_Point_of_Contact** rozšiřují informace, které by se jinak v hlavní tabulce duplikovaly. S hlavní tabulkou jsou propojeny relací typu 1:N, která mimo jiné zajišťuje, že pokud se změní jakékoliv pole v hlavní tabulce, změna se projeví i v tabulce vedlejší. V záložce „Dotazy“ se skrývá šest výběrových dotazů, které se dají využít pro vyhledávání v databázi. Jedná se o vyhledávání podle tematické kategorie, podkategorie, názvu, počtu chybějících dat, poskytovatele či data publikování. Pro jednoduché vyhledávání lze samozřejmě využít i funkci **Najít**.

Členění databáze vychází z normy ISO 19115, některé sloupce však byly záměrně vypuštěny z důvodu nadbytečnosti. Jazykem databáze je angličtina. Důvodem pro toto rozhodnutí byla vhodnost existence vzájemné spojitosti mezi databázemi a datovými servery, které jsou rovněž v angličtině. Překlad jednotlivých záznamů by byl časově náročný a nepřinesl by žádné nové informace. Autor práce navíc u uživatele pracujícího s GIS předpokládá alespoň elementární znalosti nejdůležitějšího světového jazyka.

ID	Title	Year	Units	Filename	Definitions	Abstract	Purpose	Character set
1	Agricultural Area - Percent of Land Area	2007	Percent	agri_land_area	Agricultural Area	FAO statistical c	Policy formulato	utf8
2	Air transport of passengers	2009	Number of passengers	air_transport	---	---	---	utf8
3	Animal Species - Threatened	2004	Number of Species	animalia_threatened	"Threatened" inclu	IUCN-The World	The IUCN Red Lis	utf8
4	Area under organic farming	2008	Percent	organic_farming	---	---	---	utf8
5	Built-up areas	2003	Square kilometres	built-up_areas	---	---	---	utf8
6	Civil Aviation Traffic - Total Passengers Carried	2007	Thousand People	aviation_passengers_total	Number of internat	Civil Aviation Sta	Transport infrastru	utf8
7	Disasters of Natural Origin (Excluding: Earthquakes, I	2004	Number of People	disasters_killed_excl_cap	Killed: Persons co	The growing nurr	Since 1988 the V	utf8
8	Diseases of the Respiratory System - Number of Deat	2005	Number of People	mortality_resp_infections	---	---	---	utf8
9	Divorces	2008	Number of divorces per	divorces	---	---	---	utf8
10	Drylands - Percent of Total Area	2000	Percent	dry_percent	---	---	---	utf8
11	Economically Active Population	2007	Thousand People	active_pop	The economically	The ILO program	The main objectiv	utf8
12	Emissions of SO2 - Total (National Reports, UNFCCC	2005	Gigagrams	emissions_so2_tot_unfccc	Sulfur dioxide is a	This site (http://g	In accordance wit	utf8
13	Energy Supply per \$1000 Gross Domestic Product (P	2008	Kilogram Oil Equivalent	energy_mdg	Indicator of energy	Energy is one of	One of our great	utf8
14	Farmland bird index	2008	Index (1990 = 100)	bird_index	---	---	---	utf8
15	Fatal accidents at work	2009	Incidence rate per 100	fatal_accidents_at_work	---	---	---	utf8
16	Fish Catch - Total	2008	Metric Tons	fish_catch	Fish catch relate t	Fish, crustacear	The FAO Fisherie	utf8
17	Floods - Killed People	2008	Number of People	floods_killed	Killed: Persons co	The growing nurr	Since 1988 the V	utf8
18	Foreign Direct Investment - Net Inflows	2005	Million Current US\$	invest_foreign	Foreign direct inve	The annual publi	Recording statist	utf8
19	Forest Fire Extent	2005	Thousand Hectares	forest_fire_extent	Forest Fire Extent	As a part of its fu	Fire has been a r	utf8
20	Forest Plantation Annual Change	2005	Hectares Per Year	forest_plantation_change	Annual change of	Plantation refers	FAOSTAT Forest	utf8
21	Greenhouse gas emissions	2008	Index (Actual base yea	greenhouse_gas_emission	---	---	---	utf8
22	Gross Domestic Product - Annual Percentage Growth	2007	Percent	gdp_growth_rate	Annual percentage	The annual publi	The World Bank	utf8

Obr. 16 Náhled databáze pro správu metadat v prostředí Microsoft Access

Na následujících řádcích se budeme věnovat stručnému popisu jednotlivých sloupců, aby nedošlo informačnímu šumu mezi uživatelem a softwarem. Jednotlivé sloupce jsou rovněž česky popsány v programu a slouží jako jednoduchá nápověda.

- **ID**

Číslo sloužící k identifikaci zdroje, v databázi má funkci primárního klíče. Pod stejným číslem jsou uložena data na CD-ROM (příloha bakalářské práce).

- **Title (název)**

Úplný název zdroje podle metadat.

- **Year (datum publikování)**

Datum publikování zdroje s přesností na roky. Název dle normy ISO 19115 je „Publication year“, v databázi zkráceno na „Year“.

- **Units (jednotky)**

Jednotky, ve kterých jsou hodnoty uváděny. Mohou být absolutní (milióny lidí) nebo relativní (procenta, různé indexy). Údaj „Various“ značí různé jednotky v případě zpracování více témat do jedné tabulky.

- **Filename (název souboru)**

Pracovní název zdroje bez přípony pod kterým je uložen na webu. Stejný název mají i soubory uložené na CD-ROM. Nepsaným pravidlem je, že mezery se nahrazují podtržítkem (např. tourism_arrivals).

- **Definitions (definice)**

Definování metody sběru dat a pravidla pro jednotlivé údaje. Vzhledem k tomu, že toto pole je omezeno počtem 255 znaků, neobsahuje veškeré informace. Podrobnější údaje je možné dohledat na webových stránkách s metadaty.

- **Abstract (úvod)**

Stručné představení zdroje. Vzhledem k tomu, že toto pole je omezeno počtem 255 znaků, neobsahuje veškeré informace. Podrobnější údaje je možné dohledat na webových stránkách s metadaty.

- **Purpose (účel)**

Účel sběru dat. Výhody existence dat, jejich praktické využití. Vzhledem k tomu, že toto pole je omezeno počtem 255 znaků, neobsahuje veškeré informace. Podrobnější údaje je možné dohledat na webových stránkách s metadaty.

- **Character set (kódování)**

Kódování zdroje. Vždy utf8, avšak některá metadata typ kódování neuvádějí.

- **Language (jazyk)**

Jazyk zdroje. Ačkoliv jazykem je vždy angličtina, autor cítí potřebu tento fakt uvádět.

- **Update frequency (frekvence obměny)**

Frekvence aktualizace dat. Většinou se uvádí termín „Annually“, což znamená „jedenkrát do roka“.

- **GEO theme (tematická kategorie)**

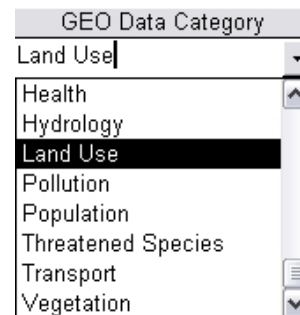
Tematická kategorie zdroje. Pro účely databáze se nevycházelo z kategorií daných normou ISO 19115, kterých je velké množství, nýbrž byly navrženy vlastní kategorie „Socio-Economic“ (Socio-ekonomická sféra) a „Physical Geography“ (Fyzická geografie). Do první jmenované náleží data týkající se činnosti přírody, tj. klima, vegetace, hydrologie a

ohrožené druhy rostlin a zvířat. Oproti tomu mapy socioekonomické zobrazují objekty a jevy, které jsou výsledkem lidské činnosti. Do této kategorie náleží např. zemědělství, využití země, doprava nebo obyvatelstvo.

• **GEO data category (tematická podkategorie)**

Tematická podkategorie zdroje. Stejně jako v předchozím případě navržený vlastní kategorie. Pro „Socio-Economic“ to jsou:

- „Agriculture“ (zemědělství),
- „Economy“ (ekonomické ukazatele),
- „Health“ (zdraví),
- „Land Use“ (využití země),
- „Pollution“ (znečištění),
- „Population“ (obyvatelstvo),
- „Transport“ (doprava).



Obr. 17 Výběr podkategorie

V oblasti fyzické geografie „Physical geography“ jsou zastoupeny následující oblasti:

- „Environmental Hazards“ (přírodní katastrofy),
- „Hydrology“ (hydrologie),
- „Threatened Species (ohrožené druhy rostlin nebo živočichů),
- „Vegetation“ (vegetace).

• **GEMET keywords (klíčová slova-čas)**

Klíčová slova týkající se časového vymezení.

• **Free Keywords (klíčová slova-téma)**

Klíčová slova pro bližší určení tematického obsahu zdroje.

• **Covered time (časové pokrytí)**

Časové období za které jsou data poskytována. Výjimečně se může objevit termín „Various“ značící neznámý rok.

- **Coverage (rozsah)**

Rozsah (prostorové umístění) zdroje. V databázi se vyskytují pouze dvě možnosti: „European Union“ a „World“.

- **Point of Contact (kontaktní místo)**

Odkaz na poskytovatele zdroje. Tabulka **t_Point_of_Contact** podrobně informuje o adrese a kontaktních údajích.

t_Point_of_Contact : Tabulka					
	Point of Contact	Role	Person	Organization	Address
*	CRED	Originator	Prof. D. Guha-Sapir	Centre for Research on the Epidemiology of Disasters	Universite Catholique de Louvain 30.94 Clos Cha
*	Eurostat	Originator	Statistical Office of the Euro	Eurostat	---
*	FAO	Originator	FAO	FAO	Viale delle Terme di Caracalla
*	International Labour Office	Originator	Bureau of Statistics	International Labour Office	4 route des Morillons
*	IPCC	Originator	Dr. Mike Hulme	IPCC - Intergovernmental Panel on Climate Change	Climatic Research Unit University of East Anglia
*	IUCN/SSC	Originator	SSC Red List Programme	IUCN/SSC	219c Huntingdon Road
*	OECD	Originator	OECD	Organization for Economic Co-operation and Developm	2, rue André Pascal External Relations and Publ
*	SEDAC	Originator	SEDAC User Services	SEDAC	61 Route 9W, P.O. Box 1000
*	The World Bank	Originator	Development Data Group	The World Bank	1818 H Street
*	UNFCCC	Originator	UNFCCC	United Nations Framework on Climate Change	Martin-Luther-King-Strasse 8
*	United Nations Population Division	Originator	Mr. Joseph Chamie	United Nations Population Division	2 United Nations Plaza
*	WHO	Originator	James L. Duppenhaler	WHO	20 Avenue Appia

Obr. 18 Údaje o tvůrci zdroje obsahuje samostatná tabulka

- **Use constraints (omezení)**

Informace o právech (omezeních) vztahujících se k popisovanému zdroji.

- **Copyright (autorské právo)**

Držitel autorského práva na poskytovaná data.

- **Online resource (online zdroj)**

Odkaz na zdroj, ze kterého je možné popisovaný zdroj stáhnout (hypertextový odkaz).

- **ULR Original resource (původní online zdroj)**

Uvedení původního zdroje dat (hypertextový odkaz).

- **File format (formát)**

Datová reprezentace zdroje. Nejčastěji uváděno „Excel spreadsheet“, avšak ve většině případů je možné použít více datových formátů.

- **Elipsoid (elipsoid)**

Referenční elipsoid zdroje. Ve většině případů WGS 84.

- **Projection (kartografické zobrazení)**

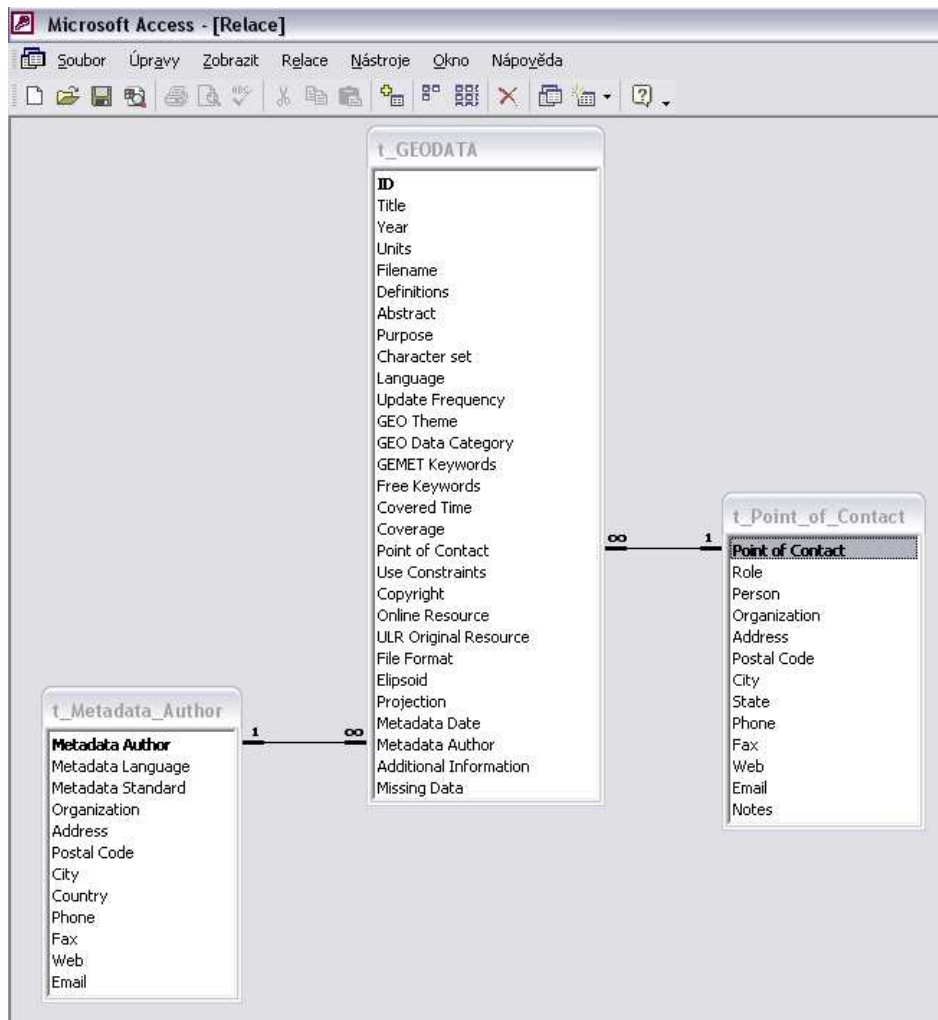
Kartografické zobrazení zdroje. Poskytovatel zpravidla neuvádí, po načtení vrstvy do GISu je třeba kartografické zobrazení nastavit ručně.

- **Metadata Date (poslední aktualizace metadat)**

Datum poslední aktualizace metadat. Uváděno ve formátu RRRR-MM-DD.

- **Metadata author (autor metadat)**

Odkaz na autora metadat. V samostatné tabulce **t_Metadata_Author** je možné se dočíst o jazyku, ve kterém jsou psána, metadatovém formátu, jakož i získat kontakty na autora metadat.



Obr. 19 Ukázka dvojnásobné relace typu 1:N

- **Additional information (doplňkové informace)**

Poznámky týkající se zdroje.

• **Missing Data (chybějící data)**

Číselný údaj uvádějící počet států, pro které v dané vrstvě nejsou dostupná data. V tabulce jsou většinou reprezentována prázdným záznamem nebo hodnotou -9999.

Missing Data	Title	ID	Year	Units	Filename	Update Frequency
0	Urban Population - Total	44	2008	Thousand People	pop_urban_total	Every 2 years
0	Floods - Killed People	17	2008	Number of People	floods_killed	Continually
0	Wind Storms - Killed People	50	2008	Number of People	storms_killed	Continually
0	Disasters of Natural Origin (Excluding: Earthquakes,	7	2004	Number of People	disasters_killed_excl_cap	Continually
0	Urban Population - Percent of Total Population	43	2008	Percent	pop_urban_percent	Every 2 years
1	Rural Population - Total (Projection)	38	2008	Thousand People	pop_rural_project_total	Every 2 years
1	Gridded Population of the World, Version 2 (GPWv2)	56	2000	Various	gridded_population	As needed
1	National Aggregates of Geospatial Data Collection: F	58	2007	Various	population_collection	As needed
1	Fish Catch - Total	16	2008	Metric Tons	fish_catch	Unknown
2	Population Density	60	2007	People per Square Kilome	pop_density	Every 2 years
2	Precipitation - Millimeters	34	2001	Millimeters per Year	precipitation_mm	Unknown
3	Arable Land	53	2007	Square Kilometers	land_arable	Annually
3	Agricultural Area - Percent of Land Area	1	2007	Percent	agri_land_area	Irregular
3	Animal Species - Threatened	3	2004	Number of Species	animalia_threatened	Unknown

Obr. 20 Dotaz na chybějící data: 14 nekompletnějších datasetů

GEO Data Category	GEO Theme	ID	Title	Year	Units	Filename
Agriculture	Socio-Economic	49	Water Use - Agricultural Water Withdrawal	2006	Billion Cubic Meters	water_use_agri
Agriculture	Socio-Economic	41	Total sales of pesticides	2008	Tons of active ingredient	pesticides_sales
Agriculture	Socio-Economic	4	Area under organic farming	2008	Percent	organic_farming
Agriculture	Socio-Economic	29	Meat Production - Total	2006	Metric Tons	meat_production_total
Agriculture	Socio-Economic	37	Rice - Yield	2007	Hectograms per Hectare	rice_yield
Agriculture	Socio-Economic	32	Pesticides Consumption - Herbicides	2004	Metric Tons	pesticide_consump_herbicides
Agriculture	Socio-Economic	16	Fish Catch - Total	2008	Metric Tons	fish_catch
Economy	Socio-Economic	54	Gross Domestic Product - per Capita	2007	Constant 2000 US\$ per P	gdp_capita
Economy	Socio-Economic	18	Foreign Direct Investment - Net Inflows	2005	Million Current US\$	invest_foreign
Economy	Socio-Economic	55	Environmental Sustainability Index (ESI)	2001	Various	environmental_sustainability_index
Economy	Socio-Economic	30	Military Expenditures - Percent of Central Government	2002	Percent	expend_military
Economy	Socio-Economic	13	Energy Supply per \$1000 Gross Domestic Product (P	2008	Kilogram Oil Equivalent	energy_mdg
Economy	Socio-Economic	35	Private expenditure on education	2008	Percent	private_expenditure_on_education
Economy	Socio-Economic	25	Industry Value Added - Percent of GDP	2007	Percent	ind_add_gdp
Economy	Socio-Economic	42	Total Unemployment - Percent of Total Labor Force	2007	Percent	unemployment
Economy	Socio-Economic	22	Gross Domestic Product - Annual Percentage Growth	2007	Percent	gdp_growth_rate
Economy	Socio-Economic	11	Economically Active Population	2007	Thousand People	active_pop
Economy	Socio-Economic	23	Gross National Income	2007	Million Current US\$	gni
Environmental Hazards	Physical Geography	17	Floods - Killed People	2008	Number of People	floods_killed
Environmental Hazards	Physical Geography	19	Forest Fire Extent	2005	Thousand Hectares	forest_fire_extent
Environmental Hazards	Physical Geography	50	Wind Storms - Killed People	2008	Number of People	storms_killed
Environmental Hazards	Physical Geography	7	Disasters of Natural Origin (Excluding: Earthquakes, I	2004	Number of People	disasters_killed_excl_cap
Health	Socio-Economic	39	Total death due to suicide	2009	Standardised death rate b	death_suicide

Obr. 21 Databáze uspořádaná podle tematické podkategorie (dotaz q_GEO_Data_Category)

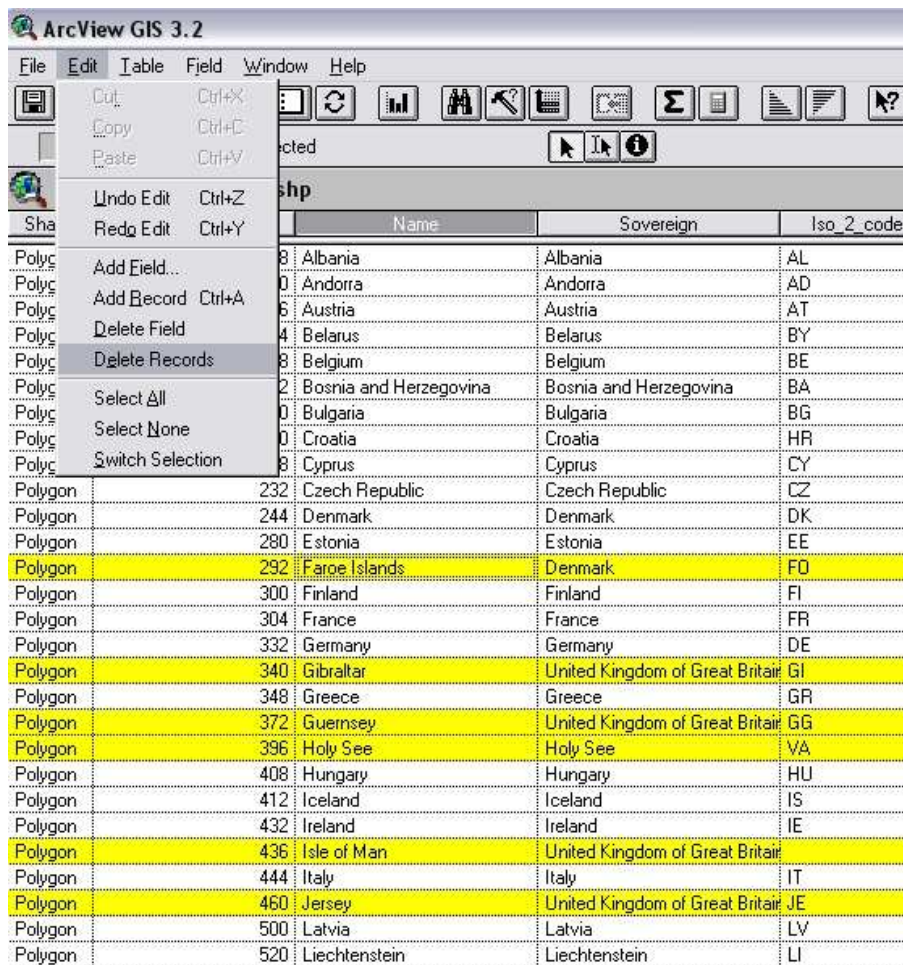
9 PRÁCE SE ZÍSKANÝMI DATY

9.1 PŘEDZPRACOVÁNÍ DAT

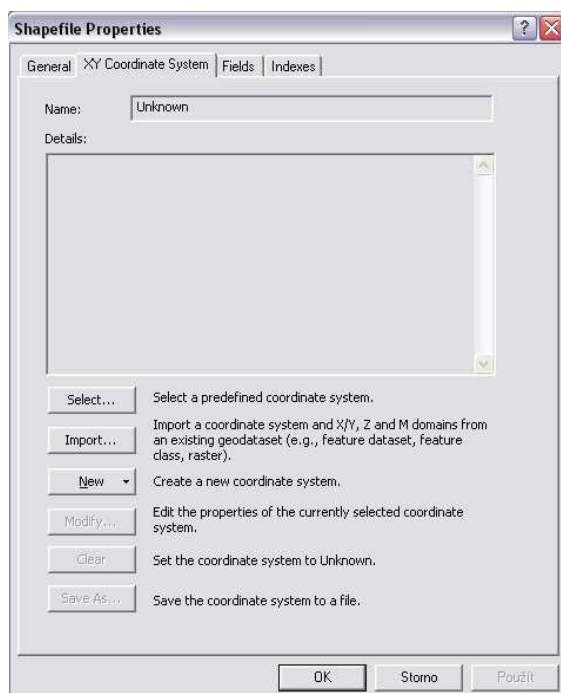
Tento krok zahrnuje základní úpravy, jelikož data v surové formě lze použít jen velmi omezeně. Jedná se především o úpravy tabulkových souborů takovým způsobem, aby bylo možné je asociovat s GISy. Např. tabulky z Eurostatu obsahují i v okleštěné formě spoustu nepotřebných informací, kterých je potřeba se zbavit. Je vhodné smazat základní metadata a ponechat jen názvy států, hodnoty a názvy sloupců. Mimo to je nutné převést datový typ „textový“ na „číselný“ (pokud tomu tak není) a soubor ve formátu XLS uložit jako DBF, což je standardní formát podporovaný ve většině GISů.

Pro základní předzpracování dat byl použit software ArcView GIS 3.2, jenž vyniká svou jednoduchostí a praktičností. Nejprve bylo vhodné vrstvy ve formátu SHP prostorově omezit na zájmové území, tedy zvolených 43 států. Po načtení vrstvy do ArcView a zobrazení tabulky stačí jednoduše smazat nepotřebné řádky (státy) příkazem **Table-Start Editing** a následně **Edit-Delete Records**. Pokud se chystáme propojovat atributové tabulky, je třeba dohlédnout na identické záznamy v klíčovém sloupci (např. Name) u obou tabulek. V tabulce můžeme provádět další změny, např. nezjištěnou hodnotu zastoupenou číslem -9999 změnit na 0 apod. Dva neevropské státy Turecko a Rusko je dobré ponechat, ale protože jejich hodnoty nebudeme používat, je praktické je změnit třeba na shodnou hodnotu a této hodnotě přisoudit šedou barvu, která se hodí k označení států zastoupených na mapě, ale nespádajících do zájmového území.

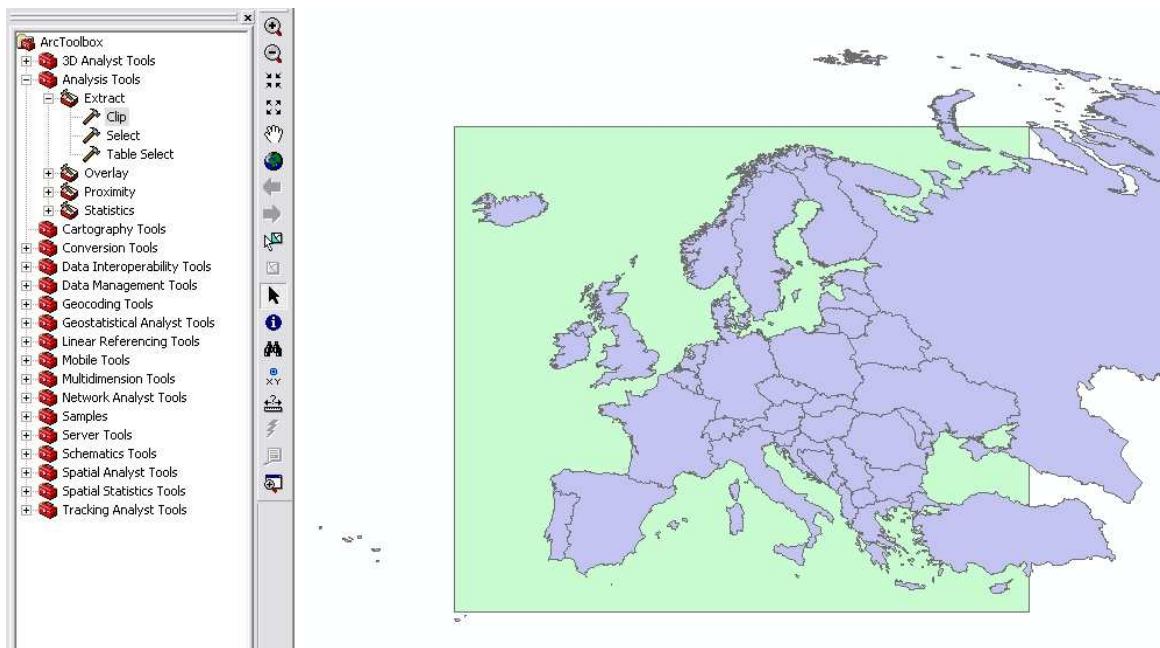
Předzpracování dat zahrnuje i práce v programu ArcCatalog, který zajišťuje správu souborů pro ArcMap. Především je třeba důsledně dbát na definování souřadnicového systému, jinak se vrstva nezobrazí. V metadatach je uvedeno, že soubory SHP mají souřadnicový systém WGS 84, nicméně pro práci v ArcMapu je potřeba systém ručně nastavit v ArcCatalogu příkazem **Properties-XY Coordinate System-Select**. Pokud potřebujete založit nový Shapefile (např. pro operaci Clip na oříznutí území), nezapomínejte, že to lze učinit pouze v ArcCatalogu. Jednoduše vyberete složku, ve které chcete novou vrstvu založit, a pravým tlačítkem myši dáte příkaz **New-Shapefile**. Pro finální oříznutí použijete příkaz **Clip**, který najdete v ArcToolBoxu (Analysis Tools-Extract).



Obr. 22 Smazání nepotřebných řádků v prostředí ArcView GIS 3.2

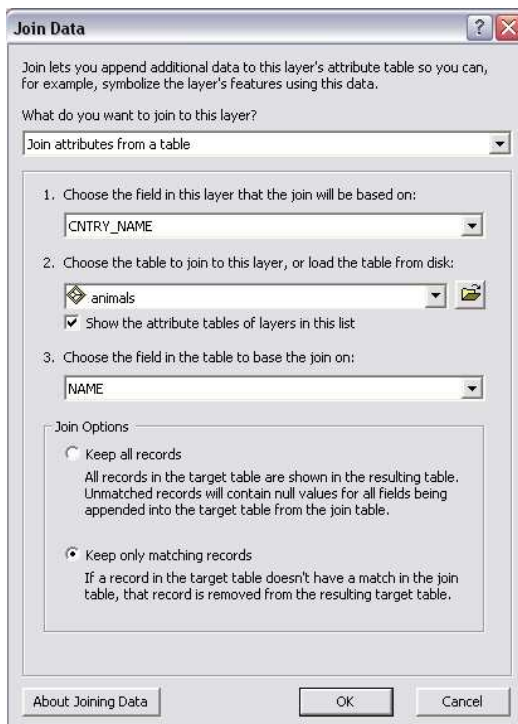


Obr. 23 Nastavení souřadnicového systému pomocí programu ArcCatalog



Obr. 24 Operace Clip se dvěma polygonovými vrstvami

Pokud máte data pouze v tabelární formě, je potřeba použít jiný SHP a tabulku s ním propojit. Lze tak učinit příkazem **Joins and Relates-Join**, přičemž je třeba mít v SHP i v tabulce sloupec shodného názvu, podle kterého budeme propojovat. Pro tyto účely obvykle postačí jakékoliv ID nebo název státu (Name). Po této operaci je dobré zkontrolovat, jestli se úspěšně propojily všechny řádky.



Obr. 25 Propojování atributových tabulek podle klíčového sloupce

9.2 SYMBOLIKA V PROSTŘEDÍ ARCMAP

Když je vrstva načtená a prostor odpovídá zájmovému území, je třeba zvolit vhodnou grafickou metodu pro znázornění údajů. Na vybranou vrstvu stačí poklepat pravým tlačítkem myši, z nabídky vybrat **Properties** (vlastnosti) a zvolit záložku **Symbology** (symbolika).

9.2.1 FEATURES

Volba **Features** (Prvky) – Single symbol (Jediný symbol) znázorní všechny prvky v jedné vrstvě stejnou barvou. Hodí se např. pro mapu kontinentu, na které chceme vyjádřit jiný, dominantní jev.



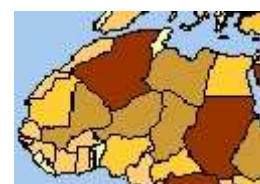
9.2.2 CATEGORIES

Volba **Categories** (Kategorie) slouží pro znázornění kvalitativních dat. Zde je nabídka tří dalších možností: Unique values (Unikátní hodnoty) zobrazí každý prvek unikátní barvou podle vybraného atributu. Používá se velmi často pro politické mapy. Unique values, many fields (Unikátní hodnoty s více atributy) – každý prvek lze zobrazit jinou barvou, u každého prvku je možné zadat až tři rozlišující atributy. Match to symbols in a style (Propojení se symbolem stylu) – každému prvku přiřadí symbol, který je definovaný v externím souboru (*.style). Propojení funguje pomocí stejného atributu.



9.2.3 QUANTITIES

Volba **Quantities** slouží pro znázornění kvantitativních dat. Volba Graduated colors (Odstupňované barvy) se používá k tvorbě kartogramu. Hodnoty je třeba rozdělit do intervalů, přičemž každému intervalu náleží jeden barevný odstín.



Graduated symbols (Odstupňované symboly) lze použít k zobrazení kvantitativních údajů pomocí symbolů. Hodnoty jsou opět rozděleny do intervalů a každému intervalu odpovídá různá velikost předem zvoleného symbolu (nejčastěji kolečko, čtverec, trojúhelník).



Proportional symbols (Proporcionální symboly) umožňují znázornit kvantitativní údaje pomocí symbolu, jehož velikost závisí na hodnotě sledovaného jevu. Je potřeba nastavit minimální a maximální velikost, které odpovídají minimální a maximální hodnotě atributu prvku. Všem ostatním hodnotám je přiřazena velikost znaku podle přímé úměry.

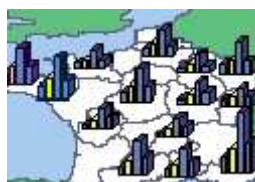


Dot density (Tečková metoda) zobrazuje údaje pomocí teček, přičemž každá tečka má určitou váhu neboli hodnotu sledovaného jevu (např. 500 tun). Tečky neukazují přesné rozmístění jevu, na ploše území se vyskytují náhodně.



9.2.4 CHARTS

Volba **Charts** (Diagramy) umožňuje zobrazovat kvantitativní data v absolutních hodnotách. Lze použít tři typy diagramů: Pie (Kruhový strukturní diagram), Bar, column (Sloupcový diagram) a Stacked (Součtový sloupcový diagram).



9.2.5 MULTIPLE ATRIBUTES

Poslední volba **Multiple Attributes** (Vícenásobné atributy) umožňuje zobrazit více atributů dat najednou. V rámci jedné vrstvy lze kombinovat kvalitativní a kvantitativní data. Jedná se o kombinaci předchozích možností.

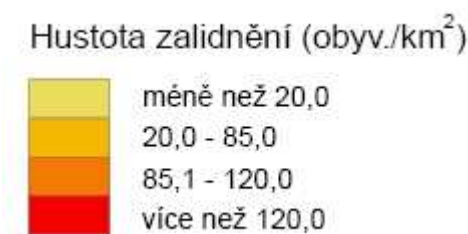


Obr. 26-35 Mapová symbolika v prostředí programu ArcMap

9.3 BARVA

Barva je samostatným vyjadřovacím prostředkem a zároveň je i součástí všech prvků mapy. Barevné provedení tematického obsahu musí být smysluplné, co nejpřirozenější a nejlogičtější tak, aby byla barva součástí mapového jazyka. Pro správné užívání barev je potřeba ctít několik základních pravidel. Dominantní prvky se znázorňují v sytých

odstínech, druhořadé a doplňující prvky v bledých odstínech. Na plošně rozsáhlá území se volí světlejší odstíny a naopak. Respektují se zažité barvy – pro lesy zelená, horstvo hnědá, vodstvo modrá, ekonomické ukazatele zelená nebo modrá. Pro prvky degradující životní prostředí se používají odstíny fialové barvy, pro antropogenní transformace reliéfu se vyčleňují odstíny hnědé barvy. Pro kvantitativní rozlišení jevů se sestavují barevné stupnice, přičemž je nutné dodržet pravidlo: čím vyšší intenzita jevu, tím vyšší intenzita barvy. Nejjednodušší volbou je užití odstínů jedné barvy, pokud se využívá více barevných tónů, je lépe využít barvy ve stejné části barevného spektra. Černou a bílou barvu se doporučuje nepoužívat. Pro znázornění výškových stupňů pomocí barev se používá barevná hypsometrie, zřejmě nejčastěji používanou stupnicí je Sydowova-Wagnerova stupnice (modrozelená – zelená – žlutozelená – žlutá – žlutohnědá – oranžovohnědá – hnědá – hnědočervená).



Obr. 36 Návrh dvoutónové barevné stupnice

9.4 KOMPOZICE MAPY

„Kompozicí mapy se rozumí rozmístění základních náležitostí mapového díla na mapovém listu. Závisí především na účelu a měřítku mapy, kartografickém zobrazení, tvaru a velikosti znázorňovaného území a na formátu mapového listu“ (Voženílek, 2001). V programu ArcMap se mapová kompozice realizuje v okně **Layout View**. Při umísťování základních a nadstavbových kompozičních prvků do mapového listu je třeba dodržovat řadu kartografických zásad. Mezi základní kompoziční prvky, které musí obsahovat každá mapa, náleží název, legenda, měřítko, tiráž a mapové pole. **Název** musí obsahovat věcné prostorové a časové vymezení tematické mapy a nejčastěji je rozdělený na titul (dominantní) a podtitul (doplnění).

HUSTOTA ZALIDNĚNÍ V EVROPSKÝCH STÁTECH v roce 2007

Obr. 37 Věcné, prostorové a časové vymezení tematické mapy

Legenda slouží k výkladu použitých mapových znaků, chybějící legenda činí tematickou mapu nepoužitelnou. **Měřítko** je hlavním ukazatelem stupně podrobnosti vyjádření prvků a jevů. Na přiložených tematických mapách je použito jak měřítko grafické, tak i číselné. Důležité je, aby měřítko mělo dekadické dělení (po 10, 100, 1000, apod.).



Obr. 38 Číselné a grafické měřítko

Tiráž zobrazuje informace o autorovi mapy, místě a roku vydání mapy. Může obsahovat i řadu dalších informací např. o použitých podkladech, kartografickém zobrazení, pořadí vydání, použitém software atd. **Mapové pole** je vlastní mapa, nejdůležitější kompoziční prvek, kterým by se měla řídit kompozice celého listu. Mezi **nadstavbové kompoziční prvky** patří např. směrovka, obrázky, grafy, tabulky a vedlejší mapy. Na vytvořených mapách se vyskytuje nadstavbový kompoziční prvek v podobě textového pole, jenž uvádí zdroj dat. Internetový odkaz spolu s datem přístupu jsou ve většině případů umístěny přímo pod legendou.

Tab. 1 Státy s objemem rybolovu méně než 16 000 t (hodnoty jsou uváděny v tunách)

rok / stát	2002	2004	2006
Albánie	3 655	4 514	5 706
Bosna a Hercegovina	2 005	2 005	2 005
Bulharsko	15 008	8 252	7 532
Černá Hora	nez.	nez.	900
Kypr	1 968	1 567	2 155
Malta	1 084	1 138	2 411
Rumunsko	6 960	5 095	6 661
Slovinsko	1 686	1 022	1 131

Vysvětlivky: nez. = nezjištěno

Zdroj: [19]

9.5 KARTOGRAFICKÉ VYJADŘOVACÍ PROSTŘEDKY

Na přiložených mapách jsou nejčastějšími kartografickými vyjadřovacími prostředky kartogram a kartodiagram. Pro znázornění kvantitativních údajů do mapy se používají kartodiagramy. Kaňok (1999) definuje **kartodiagram** jako mapu s dílčími územními celky, do kterých jsou diagramy znázorněna statistická data (absolutní hodnoty), většinou geografického charakteru. Mimo diagramů se v mapě mohou vyskytovat i další grafické

objekty a jevy (hranice, vodstvo, komunikace, sídla, atd.). Tyto kartografické znaky jsou obvykle potlačeny, generalizovány, aby nebyl zastíněn původní záměr autora kartodiagramu - informovat čtenáře o kvantitě jevu. Kartodiagramy se dělí se na plošné, liniové a patří sem i tečková metoda.

Pro vyjádření kvantity se používá **kartogram**. Je to mapa s dílčími územními celky, do kterých jsou plošným způsobem znázorněna statistická data (relativní hodnoty), většinou geografického charakteru (Kaňok, 1999). Při tvorbě kartogramu se kvantitativní data přepočítávají na jednotku plochy dílčího územního celku. Velmi častou tematickou mapou je hustota zalidnění obyvatel, tedy počet obyvatel/1km². Pokud se data nepřepočítávají na plochu, jedná se o kartogramy nepravé neboli **pseudokartogramy**, protože nemohou plně vystihovat srovnatelnou intenzitu jevu v celém území.

9.6 TEMATICKÝ OBSAH

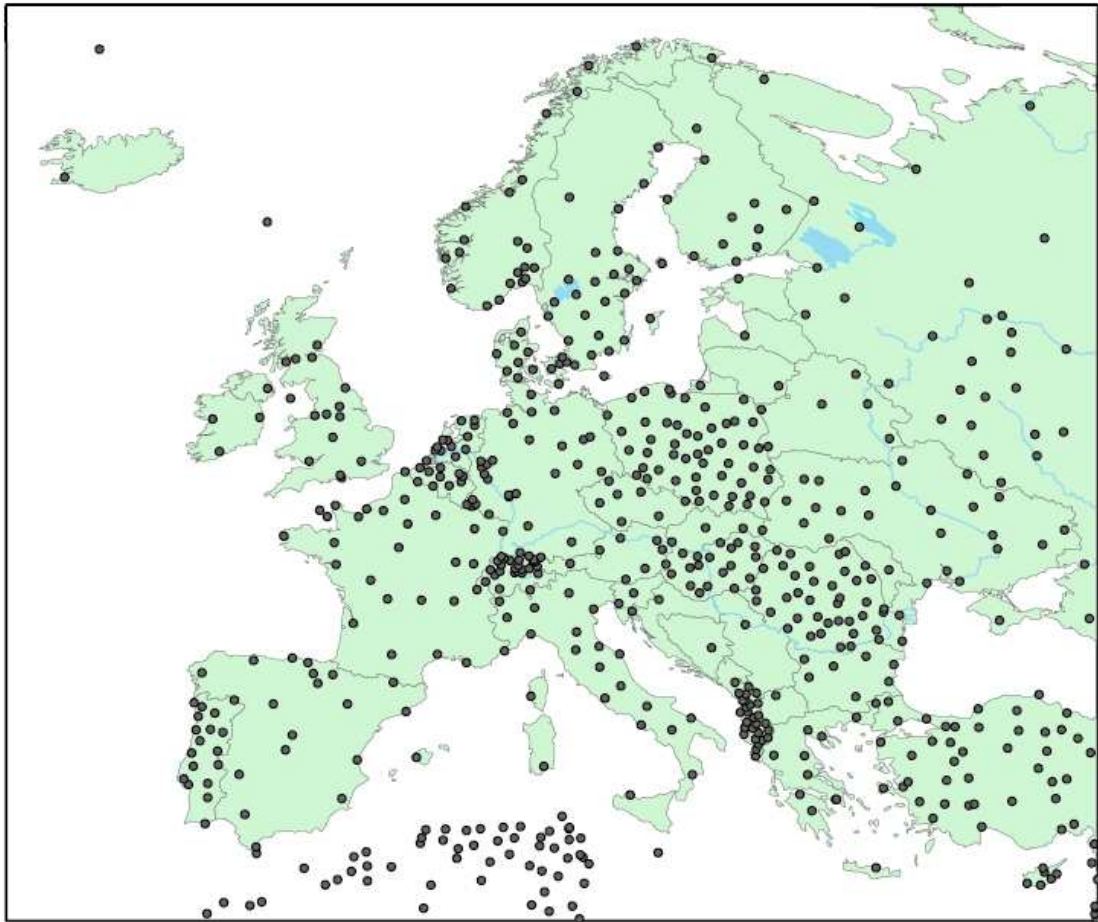
Mapovým produktem bakalářské práce jsou **tematické mapy** – mapy, které na topografickém podkladě přebíraném z vhodné výchozí (podkladové, základní) mapy podrobně zobrazují zájmové přírodní, socioekonomické a technické objekty a jevy a jejich základní vztahy (Veverka, 1995). Hlavní část obsahu tematických map tvoří tematický obsah. Je to souhrn prvků obsahu mapy tvořící mapovanou tematiku nebo s ní úzce související (Voženílek, 1998). Obsahem může být jeden nebo více socioekonomických či fyzickogeografických prvků, který vyjadřuje výsledky vědeckého výzkumu a statistických šetření různých vědních oborů. Tematický obsah tvoří volně dostupná geodata a jsou tak nejdůležitější částí každé mapy.

9.7 TOPOGRAFICKÝ PODKLAD

Topografický podklad prostorově lokalizuje prvky tematického obsahu mapy a slouží tak k základní orientaci na mapě. Topografický podklad se liší v závislosti na typu tematické mapy. Pro socioekonomické jevy se používá i odpovídající topografický podklad (sídla, hranice administrativního členění), pro jevy z oblasti fyzické geografie se znázorňují další fyzickogeografické prvky, jakými jsou např. vodstvo, vegetace nebo vrstevnice. Pro tematické mapy vytvořené v rámci bakalářské práce byly jako topografické podklady použity mapové vrstvy dodávané společně s programem ArcGIS 9.2 (ArcGIS Media Kit - ESRI Data & Maps), které ochotně poskytla vedoucí práce. Balík dat obsahoval souřadnicové sítě, bodové, liniové a polygonové vrstvy. Využita byla pouze

vrstva sídel (cities), řek (rivers), jezer (lakes) a států (cntry08) pro kontrolu souřadnicového systému.

Hranice mezi tematickým obsahem a topografickým podkladem není stálá ani ostrá. Tematický obsah může tvořit některý z prvků topografického obsahu, který je znázorněn výrazněji a podrobněji než ostatní. Výrazně tematický charakter může mít i popis mapy.



Obr. 39 Použitelné podkladové vrstvy z ArcGIS Media Kit

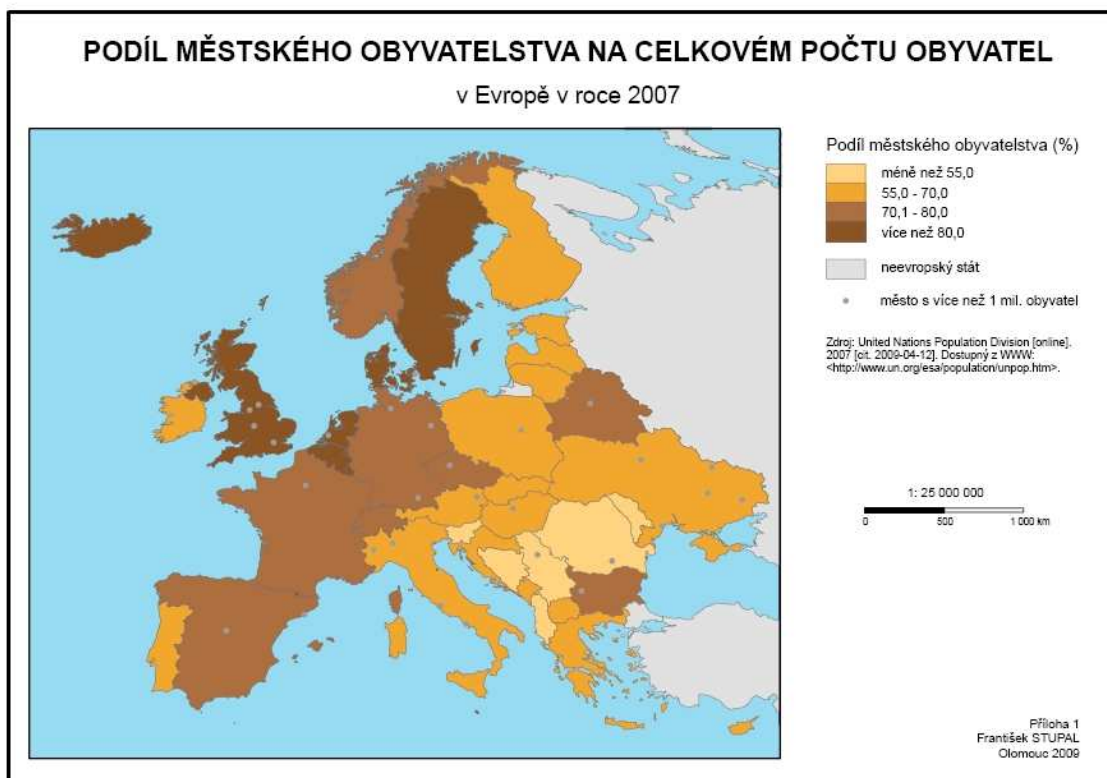
10 KOMENTÁŘ K MAPÁM

Mapové výstupy jsou důležitou součástí bakalářské práce a vhodně ilustrují praktické využití získaných dat. Obsahem této kapitoly jsou komentáře k jednotlivým mapám – použité kartografické vyjadřovací metody, použité barevné stupnice, popř. symboly, specifika mapy.

10.1 PODÍL MĚSTSKÉHO OBYVATELSTVA NA CELKOVÉM POČTU OBYVATEL

Kartografická vyjadřovací metoda: nepravý kartogram

Komentář: První mapa vyjadřuje podíl městského obyvatelstva na celkovém počtu obyvatel v jednotlivých státech Evropy. Údaje jsou vyjádřeny v procentech, proto se jedná o kartogram nepravý. Jako topografický podklad byla zvolena vrstva měst s více než 1 mil. obyvatel (vyjádřena metodou figurálních znaků pomocí teček o velikosti 4 body). Barevná stupnice je laděná do hnědé barvy. Data jsou za rok 2007, s výjimkou dvou států byla dostupná data za všechny země.



Obr. 40 Ukázka nepravého kartogramu

10.2 HUSTOTA ZALIDNĚNÍ V EVROPSKÝCH STÁTECH

Kartografická vyjadřovací metoda: pravý kartogram

Komentář: Druhá mapa reprezentuje typický socioekonomický jev, se kterým se v praxi setkáváme velmi často. Hustota zalidnění vyjadřující podíl počtu obyvatel na 1 km² je hojně používaná a oblíbená tematická mapa. Jako topografický podklad byla opět použita vrstva měst s více než 1 mil. obyvatel, protože i v tomto případě je těsně spjata s mapovanou tematikou. Barevná stupnice je dvoutónová, přechází od světle žluté do tmavě červené. Stupnice je rozdělena do čtyřech intervalů, které jsou voleny tak, aby byl v každém zastoupen přibližně stejný počet států.

10.3 CELKOVÁ PRODUKCE MASA V EVROPSKÝCH ZEMÍCH

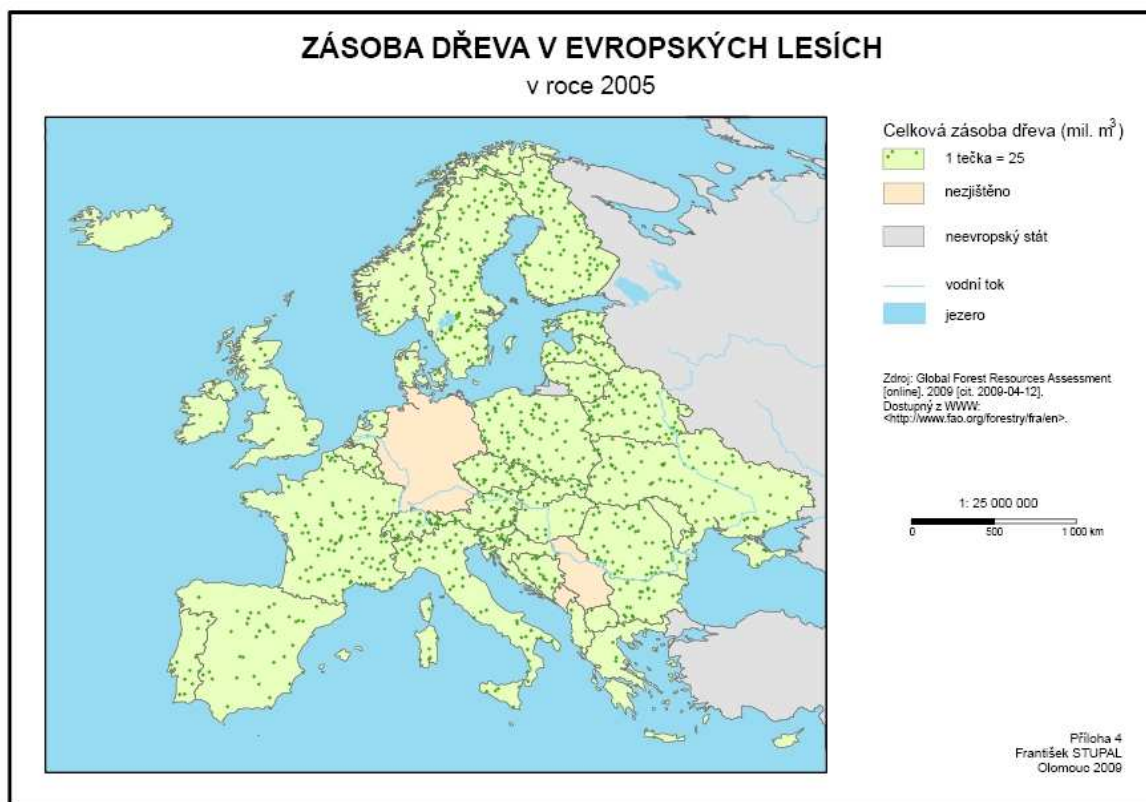
Kartografická vyjadřovací metoda: bodový kartodiagram – jednoduchý

Komentář: Do celkové produkce masa je zahrnut rybolov, produkce hovězího, drůbežího, vepřového a koňského masa. V tomto případě jsou kvantitativní hodnoty za jednotlivé státy vyjádřeny jednoduchým bodovým kartodiagramem – kolečkem – červené barvy (asociuje barvu masa). Jedná se zároveň o kartodiagram plošný, jelikož hodnoty sledovaného jevu se vztahují k celkové ploše státu, respektive produkci masa na území státu. Pro některé státy nebyla dostupná data, v mapě jsou pro snazší rozlišení vyznačena malým čtverečkem fialové barvy. Topografickým podkladem je administrativní členění států. Světlá, pastelová barva každého státu je unikátní.

10.4 ZÁSoba DŘEVA V EVROPSKÝCH LESÍCH

Kartografická vyjadřovací metoda: tečková metoda

Komentář: Čtvrtá mapa vyjadřuje pomocí tečkové metody celkovou zásobu dřeva v Evropě. Jedná se o tečkovou metodu k vyjádření kvantitativního jevu, jedna tečka představuje 25 mil m³ dřeva, jenž se nachází na území státu. V tomto případě bylo nejdůležitější určit správnou váhu tečky a její velikost tak, aby se tečky nepřekrývaly, ale zároveň nebyly ani řídce rozmístěny. Autor mapy dal přednost menší velikosti tečky, aby jedna tečka vyjadřovala co nejmenší hodnotu a mapa tak byla podrobnější. Nevýhodou je, že nebyla dostupná data za Andorru, Černou Horu, Monako, Německo, San Marino a Srbsko. Na území těchto států se nevyskytuje ani jedna tečka, a proto jsou od ostatních států barevně odlišeny. Jako topografický podklad byla použita vrstva vodních toků a jezer.



Obr. 41 Zásoba dřeva vyjádřená tečkovou metodou

10.5 OHROŽENÁ ZVÍŘATA V EVROPĚ

Kartografická vyjadřovací metoda: jednoduchý kartodiagram s plynulou velikostní stupnicí

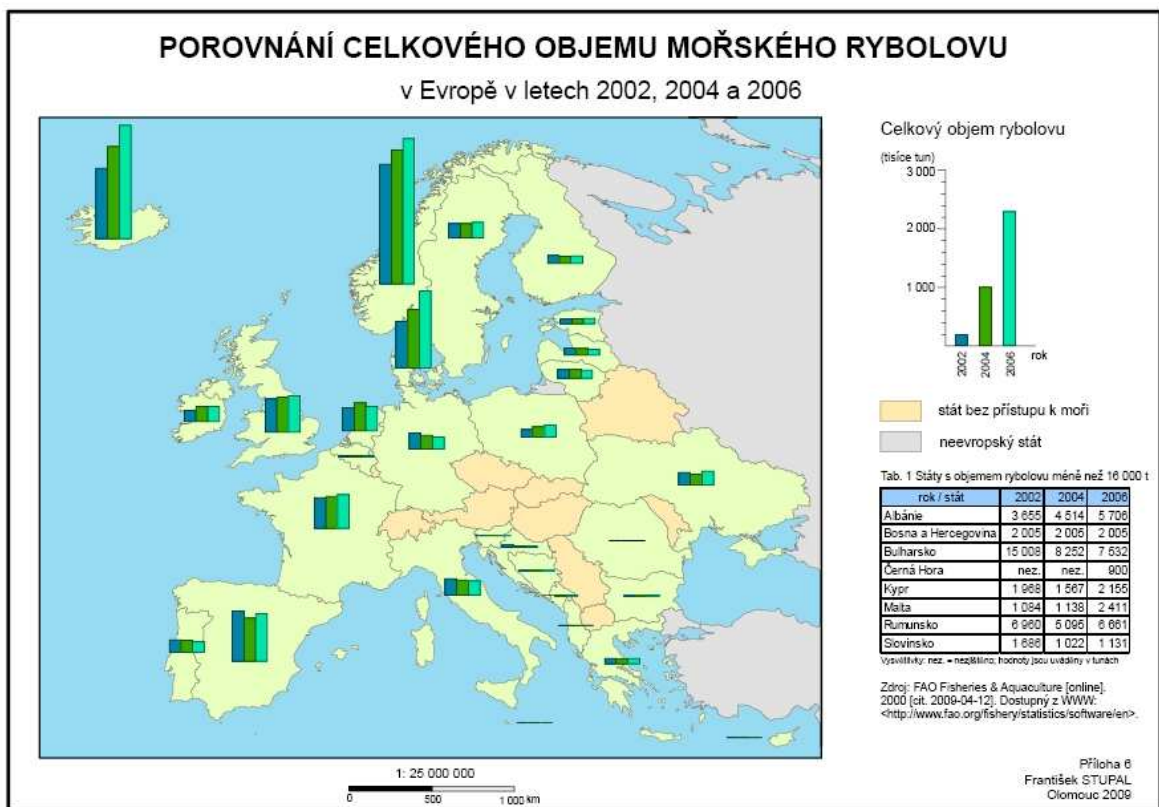
Komentář: Tematický obsah této mapy spadá do oblasti fyzické geografie – znázorňuje počty ohrožených druhů zvířat v jednotlivých státech. Kartografickou vyjadřovací metodou je jednoduchý kartodiagram s plynulou velikostní stupnicí, která nabývá hodnot od 6 do 142. U každého symbolu se nachází i číselný údaj, jenž informuje o počtu zvířat zapsaných na Červené listině (Red List). Státy, pro které nebyla k dispozici data, jsou barevně odlišeny.

10.6 POROVNÁNÍ CELKOVÉHO OBJEMU MOŘSKÉHO RYBOLOVU

Kartografická vyjadřovací metoda: kartodiagram jednoduchý sloupcový

Komentář: Jak je již zřejmé z názvu, tematickým obsahem šesté tematické mapy je mořský rybolov a porovnání jeho objemu v třech různých letech (2002, 2004 a 2006). Jednotlivé hodnoty vyjadřují sloupcové grafy, každému roku přísluší jeden sloupec. V tomto případě bylo nutné se vypořádat s velkým rozpětím hodnot. Zatímco rybářské velmoci jako Norsko nebo Island dosahovaly hodnot až 2,2 mil. tun vylovených ryb, mnoho přímořských států

jen 200 tisíc tun a méně. U osmi států byly hodnoty dokonce tak nízké, že je mezi sebou nebylo možné porovnávat. Tento nedostatek byl vyřešen tabulkou, která informuje o přesných hodnotách. V ideálním případě by tabulka obsahovala všech 43 států a porovnávání by tak bylo kompletní, avšak nedostatek místa na mapovém listu nutil omezit tabulku pouze na osm států s nejnižším výlovem (hranice byla stanovena na 16 tisíc tun). Barevně je mapa laděna převážně do modro zelené, státy bez přístupu k moři jsou reprezentovány písčitou, světle žlutou barvou.



Obr. 42 Porovnání hodnot za jednotlivé roky pomocí sloupcových diagramů

10.7 HRUBÝ DOMÁCÍ PRODUKT

Kartografická vyjadřovací metoda: nepravý kartogram

Komentář: Hrubý domácí produkt je důležitým ekonomickým ukazatelem a proto je zastoupen mezi deseti vytvořenými mapami. Hrubý domácí produkt je definován jako celková peněžní hodnota statků a služeb vytvořená za dané období na území jednoho státu. Nepravý kartogram je přepočítán na jednoho obyvatele, hodnoty jsou rozděleny do čtyřech intervalů laděné do odstínů zelené. Státy, pro které nebyly dostupná data jsou zastoupeny světle fialovou barvou.

10.8 PŘÍMÉ ZAHRANIČNÍ INVESTICE

Kartografická vyjadřovací metoda: bodový kartodiagram – jednoduchý

Komentář: Osmá mapa vyjadřuje množství přímých zahraničních investic, jenž přitekly do evropských států v roce 2003. Diagram je vyjádřen bodovým (písmenkovým) symbolem – znakem amerického dolaru, jelikož údaje jsou v milionech USD. Nejvyšších hodnot dosahuje s náskokem Lucembursko, dále Francie, Belgie a Irsko. Údaje poskytla World Bank, nejčerstvější data jsou za rok 2003.

10.9 PODÍL RŮZNÝCH TYPŮ ZEMĚDĚLSKÉ PŮDY

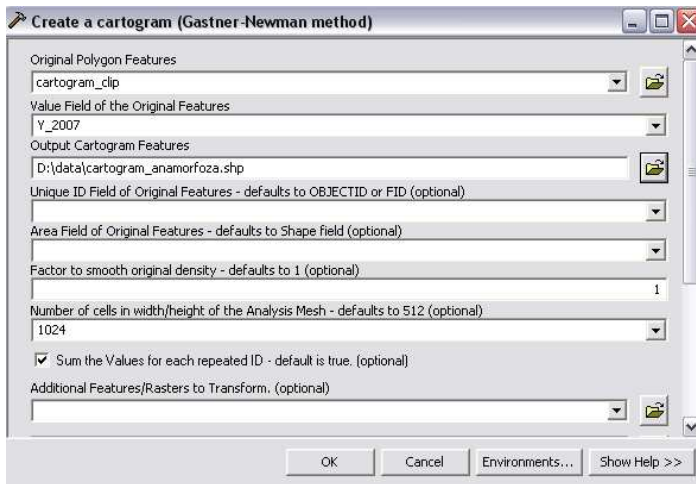
Kartografická vyjadřovací metoda: strukturní kartodiagram a jednoduchý kartogram

Komentář: Devátá mapa kombinuje kvantitativní a kvalitativní data – jednoduchý kartogram dělí státy do třech kategorií podle podílu zemědělské půdy na celkové ploše území. Kartodiagramy je vyjádřeno zastoupení jednotlivých typů půd – pastvin, orné půdy a třetí kategorii tvoří plodiny, jejichž plody se každý rok sklízí (vinná réva, chmel, ovocné stromy) – v legendě označeno jako „ostatní“.

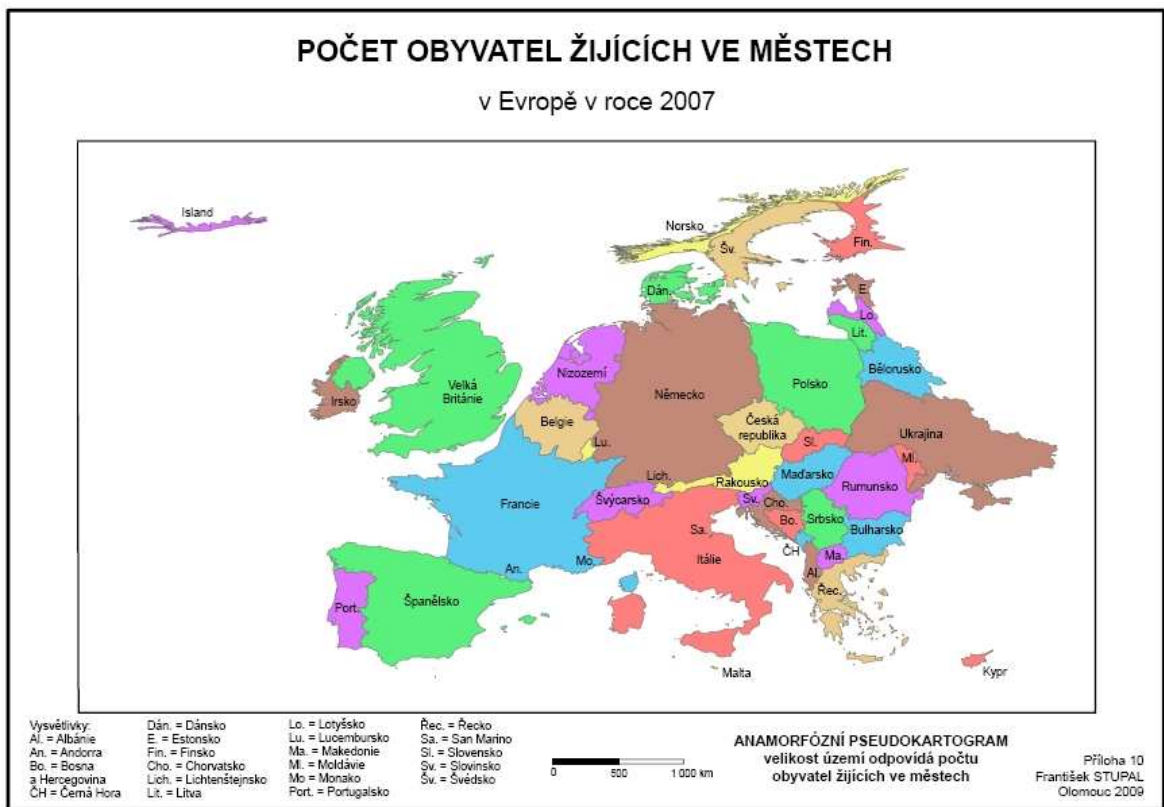
10.10 POČET OBYVATEL ŽIJÍCÍCH VE MĚSTECH

Kartografická vyjadřovací metoda: anamorfózní pseudokartogram

Komentář: Nejzajímavější mapou je bezesporu Počet obyvatel žijících ve městech. Tematická mapa je zpracovaná metodou anamorfózy – rozloha státu odpovídá počtu obyvatel trvale žijících ve městech. V tomto případě se jedná o pseudokartogram, protože zvětšením hodnoty intenzity jevu se mění velikost plochy území, avšak tvar území zůstává přibližně stejný. Podmínkou k vytvoření této mapy byla aplikace skriptu staženého ze zdroje [15]. Státy jsou barevně vyjádřeny metodou opakujících se areálů, na mapě se vyskytuje i popis jednotlivých států, protože jejich tvar je v několika případech drasticky deformován. Vzhledem ke zvětšení území je mapa jako jediná v měřítku 1: 32 000 000 (uvedeno pouze grafické měřítko).



Obr. 43 Použití skriptu k vytvoření anamorfózního pseudokartogramu



Obr. 44 Velikost území na této tematické mapě odpovídá počtu obyvatel žijících ve městech

11 ZÁVĚR

Volně dostupná geodata skýtají pro tematickou kartografii široké pole využití. V současné době existuje velké množství zdrojů nabízejících volně stažitelná data v rozličných datových formátech. Uživatel si může vybírat z velké nabídky tematických okruhů, z nichž jednoznačně převažuje socio-ekonomická sféra nad fyzickou geografii. Datasets je dobré stahovat z relevantních zdrojů, v opačném případě se je třeba k získaným údajům přistupovat kriticky, ověřovat jejich pravdivost, upravovat systém uložení, doplňovat chybějící údaje a tím celkově zlepšovat datovou kvalitu. Vážným nedostatkem je množství chybějících údajů za jednotlivé státy, což uživatele omezuje při tvorbě tematických map. Možným řešením je uvádět odhady (jako na CIA – The World Factbook) a tuto skutečnost uvést do poznámek. Z praktického hlediska se nejlépe pracuje s formáty typu Shapefile nebo CSV.

Klíčovým aspektem jsou i kvalitní metadata, jenž do značné míry vypovídají i o kvalitě dat. Geograf pracující s volně dostupnými daty by měl preferovat standardizovaná a pravidelně aktualizovaná metadata. V tomto ohledu je velkým příslibem do budoucna směrnice INSPIRE a „jednoduchý“ princip popisování metadatových prvků, jaký uplatňuje Dublin Core.

12 SUMMARY

The aim of this bachelor thesis is the analysis of free geodata, which can be used to make thematic maps of Europe. Firstly, the exploration of the Internet was undertaken with the aim of finding relevant sources that offer free geodata, which were subsequently collected and described. The next part focuses on reviewing these data, after which follows practical use of gathered data with the aim of creating thematic maps of Europe while using various cartographic methods. The author also developed a database of metadata administration, which is attached as an appendix on a CD-ROM. Hence, the main target of the thesis is to explore free geodata – their quality, structure, availability and practical use.

KEY WORDS: Data quality – Database – Europe – Free data – Geodata – GIS – Maps – Metadata – Thematic cartography

LITERATURA

Knihy a sborníky:

- [1] BERKA, Petr. *Dobývání znalostí z databází*. 1. vyd. Praha : Academia, 2003. 366 s., CD-ROM. ISBN 80-200-1062-9.
- [2] DOBEŠOVÁ, Zdena. *Databázové systémy v GIS*. 1. vyd. Olomouc : Univerzita Palackého v Olomouci, 2004. 76 s. ISBN 80-244-0891-0.
- [3] FARANA, Radim. *Databáze : speciální postupy*. 1. vyd. Praha : Český svaz vědeckotechnických společností, 2006. 169 s. ISBN 80-02-01876-1.
- [4] CHARVÁT, Karel, et al. *Geografická data v informační společnosti*. Zdíby : Výzkumný ústav geodetický, topografický a kartografický, 2007. 269 s. ISBN 978-80-85881-28-8.
- [5] KAŇOK, Jaromír. *Tematická kartografie*. Ostrava : Ostravská univerzita v Ostravě, 1999. 318 s. ISBN 80-7042-781-7.
- [6] KRÁL, Jaroslav, ŽEMLIČKA, Michal. Kvalita dat. In *Moderní databáze 2005 : 20. ročník*. [s.l.] : [s.n.], 2005. s. 5-14. ISBN 80-239-4844-X.
- [7] MIKULÍK, Oldřich, VOŽENÍLEK, Vít, VAISHAR, Antonín. *Studium rozvoje regionu založené na vizualizaci geoinformačních databází*. 1. vyd. Olomouc : Univerzita Palackého v Olomouci, 2008. 181 s. ISBN 978-80-244-1928-2.
- [8] PETR, Pavel. *Data Mining : Díl I.* Pardubice : Univerzita Pardubice, 2008. 139 s. ISBN 978-80-7395-098-9.
- [9] ŘEZNÍK, Tomáš. Standard ISO 19115 pro krizové řízení. In *Sborník přednášek 1. národní kongresu geoinformatiky v Česku : Geoinformatika pro každého*. Mikulov : MSD, 2007. s. 1-6. ISBN 978-80-86633-79-4.

[10] VEVERKA, Bohuslav. *Topografická a tematická kartografie*. 2. přeprac. vyd. Praha : ČVUT, 1995. 202 s. ISBN 80-01-01245-X.

[11] VIASCAS, John. *Mistrovství v Microsoft Access 2000 : Kompletní průvodce efektivního uživatele i tvůrce databází*. Praha : Computer Press, 2000. 819 s., CD-ROM. ISBN 80-7226-2742.

[12] VOŽENÍLEK, Vít. *Geografické informační systémy I. : Pojetí, historie, základní komponenty*. 1. vyd. Olomouc : Univerzita Palackého v Olomouci, 1998. 173 s. ISBN 80-7067-802-X.

[13] VOŽENÍLEK, Vít. *Aplikovaná kartografie I. : tematické mapy*. 2. vyd. Olomouc : Univerzita Palackého v Olomouci, 2001. 187 s. ISBN 80-244-0270-X.

Internetové zdroje:

[14] *ArcGIS Desktop 9.2 Help* [online]. c1999 , last updated 2007-03-15 [cit. 2009-04-30].

Dostupný z WWW:

<<http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=welcome>>.

[15] *ArcScripts Details : ESRI Support Center* [online]. [2001] [cit. 2009-05-14]. Dostupný z WWW: <<http://arcscripts.esri.com/details.asp?dbid=15638>>.

[16] *CIA - The World Factbook* [online]. 1994 [cit. 2009-04-27]. Dostupný z WWW: <<https://www.cia.gov/library/publications/the-world-factbook/index.html>>.

[17] *Dublin Core : Czech homepage* [online]. 2006 , poslední změna 20.11.2006 [cit. 2009-04-12]. Dostupný z WWW: <http://www.ics.muni.cz/dublin_core/elems.html>.

[18] *Eurostat* [online]. 2002 [cit. 2009-05-02]. Dostupný z WWW: <<http://epp.eurostat.ec.europa.eu/>>.

[19] *GEO Data Portal : The Environmental Database* [online]. c2006 [cit. 2009-04-20]. Dostupný z WWW: <<http://geodata.grid.unep.ch/>>.

- [20] *GeoCommunity : Free GIS Data* [online]. c1995 [cit. 2009-04-27]. Dostupný z WWW: <<http://data.geocomm.com/>>.
- [21] *Geography Network* [online]. 2006 [cit. 2009-04-27]. Dostupný z WWW: <<http://geographynetwork.com/>>.
- [22] *GOS - Geospatial One Stop* [online]. 2005 [cit. 2009-04-27]. Dostupný z WWW: <<http://gos2.geodata.gov/wps/portal/gos>>.
- [23] *INSPIRE : Infrastructure for Spatial Information in Europe* [online]. 2007 [cit. 2009-05-05]. Dostupný z WWW: <<http://www.czinspire.cz/o-inspire>>.
- [24] PIRKL, David. *Kvalita dat : Využití asociačních pravidel pro zvyšování kvality dat* [online]. 2004 [cit. 2009-04-29]. Dostupný z WWW: <http://keg.vse.cz/_slides/pirkl.ppt>.
- [25] ROUDNÁ, Milena. *Český národní komitét UNEP : program OSN pro životní prostředí* [online]. 2002 [cit. 2009-04-29]. Dostupný z WWW: <<http://www.unep.cz/unep.php>>.
- [26] VOJTEK, David. *Cizí slova a pojmy : Studijní opory pro cvičení* [online]. VŠB-TU Ostrava : Institut geoinformatiky, c2008 , Poslední úprava: 23. 10. 2008 [cit. 2009-04-29]. Dostupný z WWW: <<http://gis.vsb.cz/vojtek/index.php?page=dict/index>>.

SEZNAM PŘÍLOH

- **Příloha 1 (volná)**

Podíl městského obyvatelstva na celkovém počtu obyvatel

Tištěná mapa

- **Příloha 2 (volná)**

Hustota zalidnění v evropských státech

Tištěná mapa

- **Příloha 3 (volná)**

Celková produkce masa v evropských zemích

Tištěná mapa

- **Příloha 4 (volná)**

Zásoba dřeva v evropských lesích

Tištěná mapa

- **Příloha 5 (volná)**

Ohrožená zvířata v Evropě

Tištěná mapa

- **Příloha 6 (volná)**

Porovnání celkového objemu mořského rybolovu

Tištěná mapa

- **Příloha 7 (volná)**

Hrubý domácí produkt

Tištěná mapa

- **Příloha 8 (volná)**

Přímé zahraniční investice

Tištěná mapa

• **Příloha 9 (volná)**

Podíl různých typů zemědělské půdy

Tištěná mapa

• **Příloha 10 (volná)**

Počet obyvatel žijících ve městech

Tištěná mapa

• **Příloha 11 (volná)**

Chybějící data za jednotlivé evropské státy

Pruhový graf

• **Příloha 12 (volná)**

CD-ROM s elektronickou formou dat