United Nations Group of Experts on Geographical Names (UNGEGN)

East Central and South-East Europe Division (ECSEED)
_____

The Nineteenth Session of the
East Central and South-East Europe Division
of the UNGEGN
Zagreb, Croatia, 19 – 21 November 2008
Item 5c, 5d, 5e, 7 and 14 of the agenda

Document Symbol: ECSEED/Session.19/2008/6

Presentation of geographical names using Google Earth

Submitted by Željka Jakir and Željko Hećimović*

_____

* Prepared by Željka Jakir and Željko Hećimović, Croatian Geodetic Institute, Croatia.

Presentation of geographical names using Google Earth

1. Introduction

Google Earth application provides virtual 3D presentation of the Earth surface and space, formed from lot of different satellite pictures subsequently built-up. What is HTML (*HyperText Markup Language*) for Internet, it is KML (*Keyhole Markup Language*) for Google Earth. Google Earth users all around the world have opportunity of presentations and sharing their KML works.

2. A way to KML

To present geographical names in Google Earth we need minimum required data. Required data (object description, point coordinates and object code) for the realization of graphic display are located in XLS file (total 7797 objects). The way of realizing the task we meet is described on the schema below:
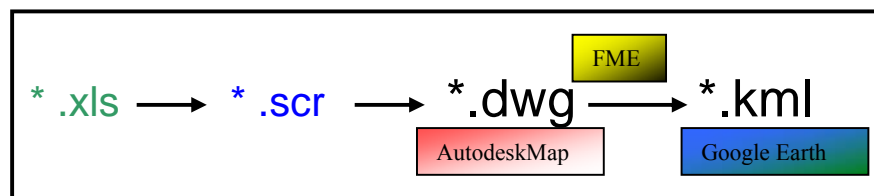

Fig. 1: Step by step realization

The data are transformed from source XLS table to a script SCR file. It is used to input the data in Autodesk Map. Script file are used to:

- define style and font,
- create layers and their colors,
- input data in certain layer.

Feature classification is made after EuroGeoNames (EGN) project feature classification. The names of new-defined layers match to new classification described in *Proposal of feature type catalogue for EuroGeoNames (EGN) project, ver. 3 February 8th of 2008* (s. table 1). EuroGeoNames (EGN) project feature classification is realized in two levels, with different number of classes and subclasses:

- 8 classes,
- 27 subclasses.

Table 1: Proposal of feature type catalogue for EuroGeoNames (EGN) project, ver. 3 February 8th, 2008 (subclasses are excluded from the table)

Code	Feature Type	Short Definition
1	COUNTRY, ADMINISTRATIVE UNITS AND OTHER AREAS	Country, territorial units of a country for administrative purposes and other man-made areas
2	POPULATED PLACES	Buildings for housing of any category like cities, towns, villages, etc.
3	NON-RESIDENTIAL STRUCTURES AND BUILDINGS	Any kind of structures and buildings, except populated places, transport, telecommunication and hydrographic features
4	TRANSPORT AND TELECOMMUNICATION FEATURES	Structures related to transports and telecommunications
5	TERRAIN FEATURES	Land features of natural environment including vegetation. In general they are natural elements but can be modified by man
6	HYDROGRAPHIC FEATURES	Natural or man-made features related to water
7	CONSERVATION AREAS	Terrain or hydrographic conservation areas of natural environment and World Heritage Sites
8	MISCELLANEOUS	Other type of features not included in classes 1 to 7

According to EGN feature classification, Autodesk Map file *topos250K.dwg* with sub layers and their attributes is created (s. Fig. 2).
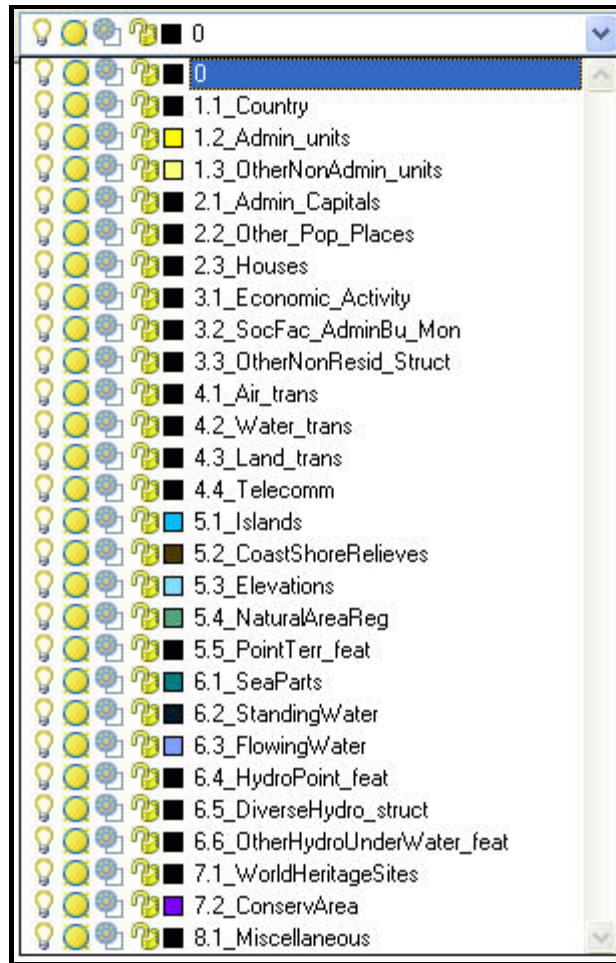
Fig. 2: Layers in Autodesk Map

The file *topos250K.dwg* is transformed in to the file *topos250K.kml* by using Feature Manipulation Engine (FME), Translator application recognizable and legible by Google Earth. Before opening a file in Google Earth, KML code should be edited and defined with different styles and regions. They are joined to more significant geographic names that is the way to simplify presentation in Google Earth according to large number of data.

3. Unicode Standards

It is very important to be able to visualize national characters. Unicode Transformation Format (UTF) is a variable-length set character encoding for Unicode. It allows almost any character in any written human language to be encoded/decoded. The initial encoding of byte codes and character assignments for UTF-8 is backwards compatible with ASCII. For these reasons, it is steadily becoming the preferred encoding for e-mail, web pages and other cases where characters are stored or streamed.

UTF-8 encodes each Unicode character as a variable number of 1 to 4 octets, where the number of octets depends on the integer value assigned to the Unicode character. It is an efficient encoding of Unicode documents.

UTF-8 character encoding advantages:

- UTF-8 is a superset of ASCII. Since a plain ASCII string is also a valid UTF-8 string, no conversion needs to be done for existing ASCII text. Software designed for traditional code-page-specific character sets can generally be used with UTF-8 with few or no changes.
- Sorting of UTF-8 strings using standard byte-oriented sorting routines will produce the same results as sorting them based on Unicode code points.
- UTF-8 is standard encoding for Extensible Markup Language (XML) documents.
- Any byte oriented string search algorithm can be used with UTF-8 data (as long as one ensures that the inputs only consist of complete UTF-8 characters). Care must be taken with regular expressions and other constructs that count characters.

UTF-8 character encoding disadvantages:

- A badly-written (and not compliant with current versions of the standard) UTF-8 parser could accept a number of different pseudo-UTF-8 representations and convert them to the same Unicode output. This provides a way for information to leak past validation routines designed to process data in its eight-bit representation

As it is already known, UTF-8 is the default encoding for XML. XML is used to describe content and it supports Unicode, allowing almost any information in any written human language to be communicated. Referring to the fact that KML is an XML-based language schema. It is used for expressing geographic annotation and visualization on Web-based, two-dimensional maps and three-dimensional Earth browsers.
Google Earth supports UTF-8. It is proved by the following statement written at the beginning of KML code.

<?xml version="1.0" encoding="UTF-8"?>

It enables proper display of every character set used to present toponym (Latin and non-Latin scripts).

4. Loading of the geographical names into the Google Earth

A procedure of opening KML file in Google Earth is very simple. In the main menu, we choose option *FILE – Open* and select name of desired file (in our case *topos250K.kml*).
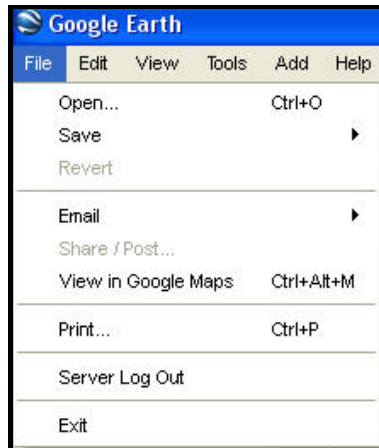
Fig. 3: Window *Menu*

After the loading file in to the Google Earth the names of the layers can be seen (s. fig. 4):
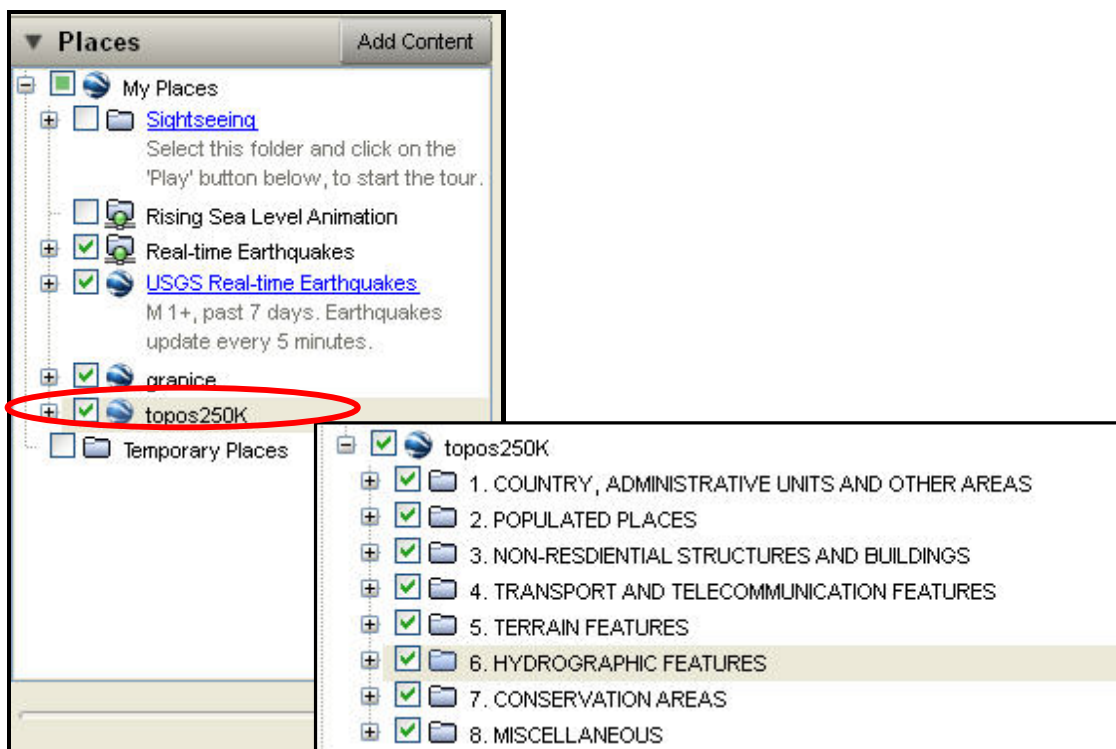

Fig. 4: Google Earth sub window *Places* with opened files and layers

Geographical names feature classes after the EGN feature classification can be seen on the figure 5.
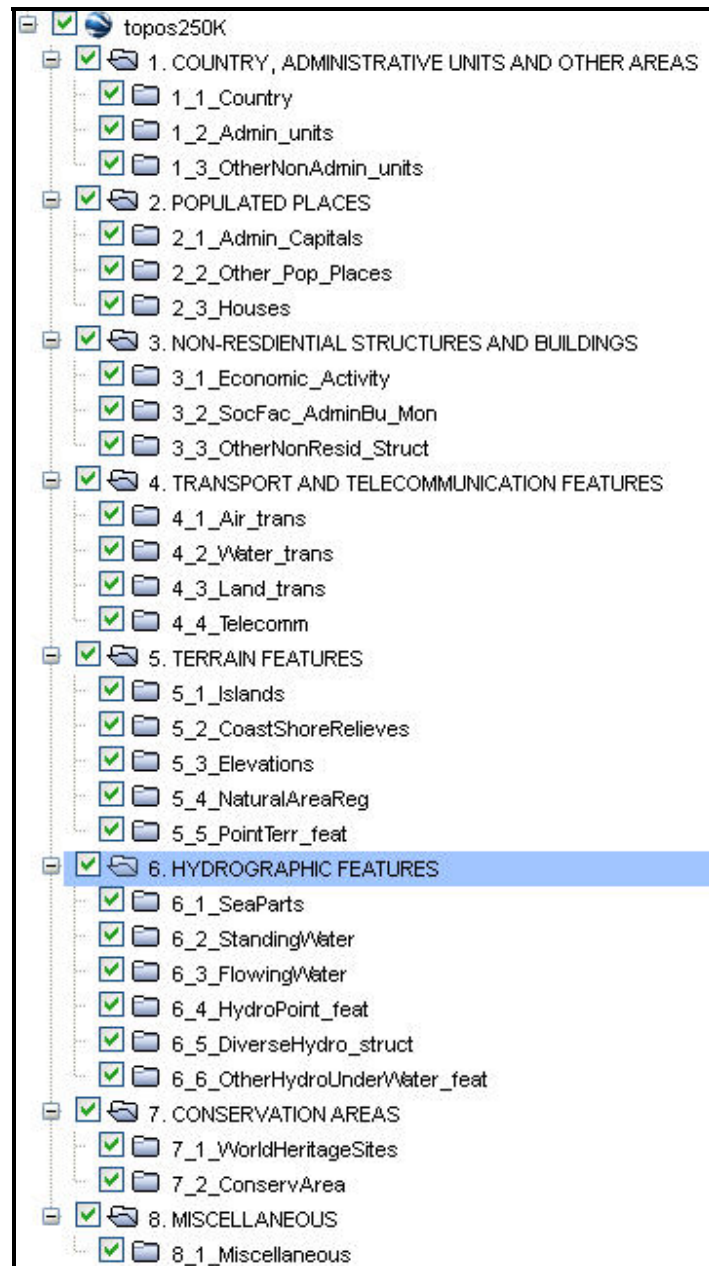
Fig. 5: Feature classes presented in Google Earth

The layers can be selected or deselected depending on user interest. On the figures 6, 7 and 8 are presentations of geographical name in different scales.

Fig. 6: Initial presentation of geographic names

Fig. 7: Small-scale presentation of geographic names

Fig. 7: Small-scale presentation of geographic names

For every feature class, we can see diversities displayed as a variety of colors, font size and other visualization parameters. The visualization depends on zooming level. Level of Detail (LOD) condition has to be met. LOD describes the size of the projected region on the screen that is required in order for the region to be "active". It also specifies the size of the pixel ramp used for fading in and fading out. Each of the geographical names features classes can be selected, activated or deactivated.

5. Conclusion

The presentation of geographical names in Google Earth is practically exposing problems of standardization of geographical names. Implementation of Unicode character encoding is giving possibility to present geographical names using almost all-national character sets.

Google Earth is giving opportunity to present geographical names in effective way allowing implementation of the geographical names standardization criteria.