

Metadata and Website Design for Statistical Data Dissemination

UDC: 004.738.12:519.22

Miroslav Hudec, Pavol Büchler

INFOSTAT, Institute of Informatics and Statistics, Bratislava, Slovakia

Dissemination of statistical data targeting to web-based audience is one of inevitable requirements for statistical organisations. This poses a significant challenge to the statistical organisation to provide the suitable web page format, accuracy, timeliness, and reliability of data and especially metadata. Descriptive metadata are strongly content oriented and their aim is to ensure a correct interpretation of presented data on the website and to avoid the misuse. Metadata assisting the search and navigation give general view of the statistical website and help users to understand the general website structure. The form of metadata and data presentation facilitates legibility and apprehensibility of the disseminated data meaning. The aim of this paper is to present experience in the design of websites for statistical purposes with emphasis on metadata, data and the appropriate form of data dissemination related to territorial units.

These important factors for website design are analysed, and discussed on examples of following websites: MOJ/MIS (Urban and Municipal Statistical Project and Information System of the Slovak Republic), Urban Audit for selected Slovak towns and Censuses (Population and housing censuses in Slovakia). Each of these projects has its own requirements, goals and user groups as well as common features which are highlighted. The common features of these websites are researched to obtain as unified presentation of information and navigation structure as possible. The purposes and goals of each website are preserved in unchanged form.

1. Introduction

Data dissemination is broad area of research and application. Dissemination of statistical data targeting web-based audience is one of essential task of statistical organisations. The statistical websites should present vast variety of collected and analyzed data. New developing tools and new knowledge about dissemination offers new possibilities for statistical organisations to create more sophisticated websites with large data sets. This poses a significant challenge to statistical organisations to provide suitable format to the users and also ensure accuracy, timeliness, and reliability of data and especially metadata.

As the way in which websites are designed can either improve or impede users to look up for information on Internet, great attention has to be paid to the website design to facilitate the access to data. A website with an easy to use and comprehensible navigation is required. This paper points out metadata on the one hand in relation with giving context and increasing meaning of presented data and on the other hand in relation with navigation and search on statistical websites. The purpose of tables and graphs is to show important information effectively. This paper briefly discusses the appropriate design of websites as well as design of tables and graphs to support better legibility of information presented on these websites.

The websites have to satisfy the requirements of users and to attract them to repeatedly visit websites and to

apply the content of the website for their needs. The purpose of websites analysed in this paper is to enhance knowledge of home and foreign general and professional public about some statistical information concerning the Slovak Republic.

The theoretical aspects and practical solutions of statistical websites are discussed in this paper. First, common aspects for the website design are mentioned and then these aspects and experiences with the web design are discussed on examples of the MOŠ/MIS (Urban and Municipal Statistical Project and Information System of the Slovak Republic), Urban Audit and Population and Housing Censuses in Slovakia (Censuses) websites. Each of these projects has its own requirements, goals and user groups as well as common design elements which are highlighted.

2. Metadata

Data are raw representation of facts; the metadata are necessary to supply the context for the data to be usable. IAIDQ's definition of data is as follows: Data: (1) Symbols, numbers, or other representation of facts; (2) the raw material from which information is produced when it is put in a context that gives it meaning [13]. Information is data in context. IAIDQ's definition of information: Information: (1) Data in context, i.e., the meaning given to data or the interpretation of data based on its context; (2) the finished product as a result of processing, presentation and interpretation of data [13]. Thereby, information is data

plus description given by metadata. They enable better understanding meaning of data (in statistics data are mostly figures). Metadata describe who, when, how, where, in which unit... collected data and these metadata are very useful for end users. According to these definitions the following equation is correct: $\text{Information} = \text{Data} + \text{Metadata}$.

Data are raw facts. Information takes into account the context of data. Without any context the figure is just a number, and a number by itself has no meaning associated with it. If there is an error in the data value the error can be found and corrected only by means of an adequate context. The essence of metadata is reducing or eliminating the communication barrier between the human and the presentation unit (in this web pages case), so that the data obtained from web pages are in general clear and can facilitate actions based on these data.

The metadata quality is more important than the data quality. It is absolutely necessary that the description of the data is up-to-date, accurate, accessible, complete, and consistent. Shortly, the metadata have to be of such high quality and so appropriately presented that users can rely on them. If any metadata component does not meet user's quality expectations, then the user will deem the metadata and data to be unusable. If an error occurs in data, it can be relatively easily found with the help of reliable metadata and corrected as long as the error in data is not essential.

Metadata mentioned above are descriptive metadata and they are strongly content oriented. Their aim is to ensure a correct interpretation of presented data on the website and to avoid misuse. Metadata assisting search and navigation give general view of the statistical website and help users to understand the general website structure. Metadata for searching and navigation are also interconnected with descriptive metadata to enable navigation across three-dimensions in this case: indicator, territorial unit and period. Such navigation helps users to obtain adequate data more efficiently. The good navigation structure can attract occasional users or users with limited statistical knowledge to use the websites and allow professional users to save labour and time.

3. Website design

Besides the sufficient amount of the relevant metadata and data, the form of metadata and especially data presentation is very important too. The intent is to point out the most important data. The user's preat-

tentive process is an extremely fast process. If the designer wants a particular set of objects to be seen as belonging together, then he/she should design them differently from the surrounding information. This visual information is moved into the user's short-term-memory. In this context the most important fact concerning the short-term-memory is that readers of graphs or tables can only hold a few chunks of information at any time. It means that in tables and graphs the relevant legend information should be kept as close as possible to the data and to have adequate supporting-information. Only well designed tables and graphs are usually stored in the user's long-term-memory. Some tables and graphs from the websites analysed in this paper are shown in this chapter. It is more useful if data are in several smaller tables than in big one. When such separation is not possible, is better to put data into logical parts in the same table to point out differences and common chunks of data. "The numbers have an important story to tell and it is up to us to help them tell it." [2].

In this case study three above mentioned websites are considered. At the beginning of deeper analyses goals and purposes of each website are mentioned. The environment for data is the same for all websites. Data are situated in a cube of three dimensions (axes): territorial units, indicators and periods. These dimensions are basis for navigation on web sites to retrieve an appropriate data. Common design elements are highlighted bellow.

Statistics about users' behaviour shows that: 34% of users selects the search engine, 23% chooses the menu list (the toolbar, tree structure), 10% goes to or looks at key figures and 4% goes to or looks at marketing part of the page (relevant statistical publications for example) and none looks at the institution or of project logo. In above mentioned survey on Statistics Denmark websites log files and Eye tracking method were used. Eye tracking method uses special designed cameras to catch movement of user eyes during browsing web pages. Log files were analysed to trace how users attain statistical web pages. This data mining shows that users attain a web page by different ways. Users mostly come "out of the blue". The second way is searching by the search engine, then follow direct links and links from other statistical pages. This survey also has shown that Google is the most important searching engine. All of examined parts except the marketing one can be find on MO[MIS webpage shown in Figure 1. Detailed description of this web page is in the next section.

3.1 MOŠ/MIS

The purpose of this presentation is to enhance awareness of home and foreign community about basic indicators concerning all Slovak municipalities and whole territorial structure of the Slovak Republic. The goals are following: to simplify the access to basic data concerning municipalities, to enable the presentation of whole structure of monitored indicators for all periods of their monitoring and to make available the contact

for ordering further data. This website presents data only for the last year approved by the Statistical Office.

The website fully meets the World Wide Web Consortium (W3C) standards (HTML 4.01 Transitional and CSS) and the regulation of the Ministry of Finances of the Slovak Republic Nr. MF/013261/2008-132 on the way of design of information systems and websites for public administration (national regulation).

SO SR - Urban and Municipal Statistics SR - Mozilla Firefox

http://www.statistics.sk/mosmis/eng/run.html

Mestská a obecná štatistika MOŠ

Home page | Slovenská verzia | Textual version | Contact | Site map

Find municipality by name | Select municipalities by region-district

Slovak Republic

- Region Bratislava
- Region Trnava
- Region Trenčín
- Region Nitra
- Central Slovakia
- Region Žilina
 - Bytča
 - Čadca
 - Dolný Kubín
 - Kysucké Nové Mesto
 - Liptovský Mikuláš
 - Martin
 - Námestovo
 - Ružomberok
 - Turčianske Teplice
 - Tvrdošín
 - Žilina
- Region Banská Bystrica
- East Slovakia
- Region Prešov
- Region Košice

Territorial units (NUTS) navigation tree

District Žilina
list of municipalities

Map of NUTS4 structure

Selected data: 31.12.2007

Indicator	Value
Number of Municipalities [LAU 2]	53
of which with town statute	3
Population density [popul. per km2]	194
Population	157 787
of which females [%]	51,4
Surface area [km2]	815,1

Some figures for selected district

List of municipalities in the: District Žilina

Belá	Bitarová	Brezanov	Divina	Divinka
Dlhé Pole	Dolná Tlážina	Dolný Hričov	Fačkov	Gbeľanov
Horný Hričov	Hričovské Podh...	Hôrky	Jasenové	Kamenná Poruba
Konská	Kotrčiná Lúčka	Krasňany	Kunerad	Kľačče
Lietava	Lietavská Lúčka	Lietavská Svín...	Lutiše	Lysica
Malá Čierna	Moiš	Nedetzka	Nezbudská Lúčka	Ovčiarsko
Paština Závada	Podhorie	Porúbka	Rajec	Rajecká Lesná
Rajecké Teplice	Rosina	Strečno	Stránske	Stráža
Stráňav	Svederník	Teplička nad V...	Terchová	Turie
Varín	Veľká Čierna	Višňové	Zbýňov	Šuja
Žilina	Čičmany	Ďurčiná		

List of municipalities

Optimized for: 800x600. Last modification on April 20, 2009
 Design - Copyright © 2007 INFSTAT BRATISLAVA, Institute of Informatics and Statistics Bratislava, Slovak Republic
 Content - Copyright © 2007 SO SR, Statistical Office of the Slovak Republic
 Questions, remarks or comments: MOŠ/MIS - Administrator

Figure 1: Key figures and navigation to municipalities – MOŠ/MIS

The navigation tree of territorial units is in fact metadata describing the whole territorial structure of country and assisting in navigation. The territorial units navigation tree has been done according to the Nomenclature des Unités Territoriales Statistiques (NUTS) structure. The map presents the navigation through the fourth NUTS level. Six key figures (data) and list of municipalities are presented for the selected district.

There are three ways for reaching municipalities. First two ways are done by metadata. The first one is fast selecting the district (NUTS 4) by the aid of the map

shown in Figure 1 and choosing the particular municipality from the list of municipalities belonging to the selected district. The second method consists in navigation across the hierarchical tree from NUTS 1 to NUTS 4 shown on the left side in Figure 1 and 2. These two navigation methods enable to preview the municipality position in the map and get some statistical key figures for the selected district, region, group of regions and country. Finally, the user can attain the requested municipality directly through the search engine. In a final step indicators for selected municipality divided into domains are presented in Figure 2.

The screenshot shows a web browser window with the URL <http://www.statistics.sk/mosmis/eng/run.html>. The page title is "SO SR - Urban and Municipal Statistics SR - Mozilla Firefox". The main content area displays "Mestska a obecna statistika MOŠ" for the municipality of Terchová. A navigation tree on the left shows the hierarchy from Slovakia to Terchová. The main content area has tabs for "Basic Characteristics and Gravity", "Technical equipment", "Demography", "Sport and Culture", "Health service and Environment", "Selected services", and "Selected results of census 1991 and 2001". The "Basic characteristics (31.12.2007)" table is shown below.

Indicator	Value
Code of municipality	518042
Name of district	Žilina
Name of region	Žilinský
Municipality statute	municipality
Postcode	013 06
Phone code	041
The earliest recorded mention	1598
Altitude of municipality - town	560
Surface area of municipality [m ²]	84 541 848
Population density [popul.]	48

A tooltip for "Gravitational indicators" is shown, containing the following text: "Municipality is created by one or more cadaster areas (territorial technical units). Cadaster area is understood as a territorial and technical unit, which forms topographically closed and in cadaster registered file of parcels. Geometric assessment of cadaster area is understood as definition of location and parameters of cadaster area by its borders. Borders of the cadaster area are marked by refructure point (border stone, metallic pipe, hall, etc.). The arable land, hopgarden, vineyard, gerdens, orchard, permanent growth of grass, forest area, water area, build-up area and curtyard and other area are registered in cadaster of municipality/town."

Figure 2: Selected domain of indicators and municipality – MOŠ/MIS

Figure 2 contains also metadata but the main stress is on data. Two tables in the middle of page contain rows and fill colours are used here for better information legibility. This is an example of how fill colours in alternating rows can be used to aid scanning across rows and point out that the importance is on the row not on the single value. Properly designed fill colours are for scanning across rows less distracting to the eye than grids. If user wants to know more about an indicator, the tooltip

with full methodical notes concerning an indicator appear by pointing on an indicator or on his value.

In the MOŠ/MIS website the descriptive metadata are used as a navigation tool too. Users can very easy obtain indicators from all domains by one click. The user also can very easy obtain the desired district, region and group of region for the analysed municipality. It is very useful that user can by one click on the tree of territorial units

obtain a totally different district with its municipalities. The territorial unit tree always shows user's position in the whole NUTS structure of the Slovak Republic.

3.2 Urban Audit

The purpose of the Urban Audit Website is to present all indicators concerning eight selected towns and their subcities in the Slovak Republic in years 2001 and 2004. The website was realised in frame of the European Urban Audit project. All indicators are same for both years which allow compact presentation of surveys.

The eight selected towns are shown in tree on left side of screen in Figure 3. All indicators for selected towns (C - city), one level above or districts (L - large) and parts of town (S - sub-city) are presented. The navigation structure is the same as on the MO[MIS] website but only for selected districts, cities and sub-cities participated in this

European project. The user can obtain data for all supported territories by its selection from the tree. Here are significantly more indicators (338) in comparison with the MOŠ/MIS, where are presented 140 indicators. Indicators are divided into domains and sub-domains, which are accessible via the drop-down menu. After clicking on a sub-domain, user can see all indicators and their values. The table is of the same type as in Figure 2 but with a larger number of rows and also fill colours are here used for better legibility. In this website fill colours in alternating rows are more accentuated than the previous one because there are significantly more rows in each table.

The UrbanAudit website is similar to the MO[MIS] website and also meets the W3C standards and the mentioned national regulation. When users become familiar with one of these websites they will be able to use easily the other one.

The screenshot shows the 'Urban Audit - Slovak Republic' website interface. On the left is a navigation tree under 'URBAN AUDIT - Information' listing various cities and districts, with Bratislava (SK001C) selected. The main content area displays 'SK001C BRATISLAVA mesto' and a table of indicators under the sub-domain 'TT1 Travel and Transport - Travel Patterns'. The table lists indicators with their codes and values for 2001 and 2004. Annotations with arrows point to the navigation tree ('Territorial units for UA navigation tree'), the domain menu ('Domains and sub-domains of indicators'), and the indicator table ('Indicators and their values').

Code	Name of indicator	2001	2004
TT1003V	Percentage of journeys to work by car	24	26
TT1010V	Percentage of journeys to work by public transport (rail, metro, bus, tram)	72	70
TT1006V	Percentage of journeys to work by motor cycle	.	.
TT1007V	Percentage of journeys to work by bicycle	.	.
TT1008V	Percentage of journeys to work by foot	3	4
TT1012V	Percentage of journeys to work by car or motor cycle	24	26
TT1019V	Average time of journey to work (minutes)	43	38
TT1020V	Average length of journey to work by private car (km)	12	12
TT1064V	People commuting into the city	89 424	100 000
TT1065V	People commuting out of the city	7 528	7 000
TT1069V	Number of stops of public transport	1 362	1 235
TT1083V	Number of buses (or bus equivalents) operating in the public transport	1 011	975
TT1084V	Average age of the bus (only buses) fleet	8	10
TT1085V	Proportion of buses running on alternative fuels	3	19
TT1066V	Length of public transport network (km)	614	620
TT1077V	Length of public transport network on fixed infrastructure	185	185
TT1078V	Length of public transport network on flexible routes	429	435
TT1082V	Length of restricted bus lanes	.	.
TT1079V	Length of bicycle network (dedicated cycle paths and lanes)	73	85
TT1080V	Cost of a combined monthly ticket (all modes) for 5-10 km in the central zone	.	.
TT1081V	Cost of a taxi ride of 5 km to the centre at day time	2	2
TT1057V	Number of private cars registered	182 002	200 424
TT1013V	Number of motor cycles registered	2 879	4 276
TT1070V	Number of park and ride parking spaces	41 902	43 538
TT1075V	Maximum charge of on-street parking in the city centre per hour	.	1
TT1060V	Number of deaths in road accidents	32	32
TT1061V	Number of persons seriously injured in road accidents	163	152
TT1071V	Accessibility by air (EU27=100)	.	.
TT1072V	Accessibility by rail (EU27=100)	.	.
TT1073V	Accessibility by road (EU27=100)	.	.
TT1074V	Multimodal accessibility (EU27=100)	.	.

Figure 3: Selected sub-domain of indicators for selected city – Urban Audit

3.3. Censuses

The purpose of this project is to enhance the knowledge of domestic and foreign general audience and professionals about population and housing censuses realised in the area of Slovakia during almost one century. The population and housing censuses, give a complex picture of demography and social and economic characteristics of a country. The census data are very specific and it is impossible to replace them by data from other types of surveys.

The objectives of this project partially differ from the objectives of other two projects mentioned above. This presentation offers detailed results and other related basic information concerning population and housing censuses realised in Slovakia in years 1921, 1930, 1950, 1961, 1970, 1980, 1991 and 2001. The user obtains the results in one place, in the unified form and in various representation modes like tables, diagrams, maps and their combinations. Detailed results of each census are in separated web pages because territorial unit structure was changed in the time between every two censuses and the structure of indicators was designed for each census according to actual requirements. This presentation also provides the possibility to compare common selected indicators in their historical context. The de-

sign allows an easy and trouble-free future expansion of this website. More about this project is in [4].

The data from the last three censuses (1980, 1991 and 2001) were computer processed already in the census time. The creation of the three - dimensional data structure (time, territory and indicator), the conversion to the unified data format and the storage of data into the common database were accomplished in first project realisation steps. In the case of older censuses (1921-1971) the selected paper documents were used as data sources. They were digitised and transformed into HTML format.

The territorial navigation for the three latest censuses is realised via the hierarchical tree of territorial units actual for each census. Figure 4 depicts the hierarchical navigation tree of territories, list of available tables and the selected table for the 1991 census. Territorial unit navigation tree is the administration division of the Slovak Republic effective during the 1991 Census. The difference to the MOJ/MIS website is in the territorial unit navigation tree as the administration division of the Slovak Republic changed between years 1991 and 2001. The focus in this table is on individual values in cells, so the grids are used. This table is bidirectionally structured with two sets of subdivisions of categories.

Figure 4: The territory navigation tree, list of tables and the selected table of results – 1991 census

This table presents quantitative values in intersection of one regular and one hierarchical subdivision of categories and shows their relation to both subdivisions of categories. When tables are long or total values are more important than the detailed ones, total values are placed at the top of a table so the user can read total values without moving the vertical bar on the screen.

Besides the data from individual censuses, the presentation offers time series of selected indicators in numer-

ical and diagram form. Some series start as early as in 1848. An example of time series is shown the Figure 5 depicting the population development in years 1848 - 2001 in the area of Slovakia.

The population is presented in the bar graph. Grid lines are used to make easier to perceive differences in the lengths of bars. The Y-axis starts at zero value, what means that bar length expresses nominal values.

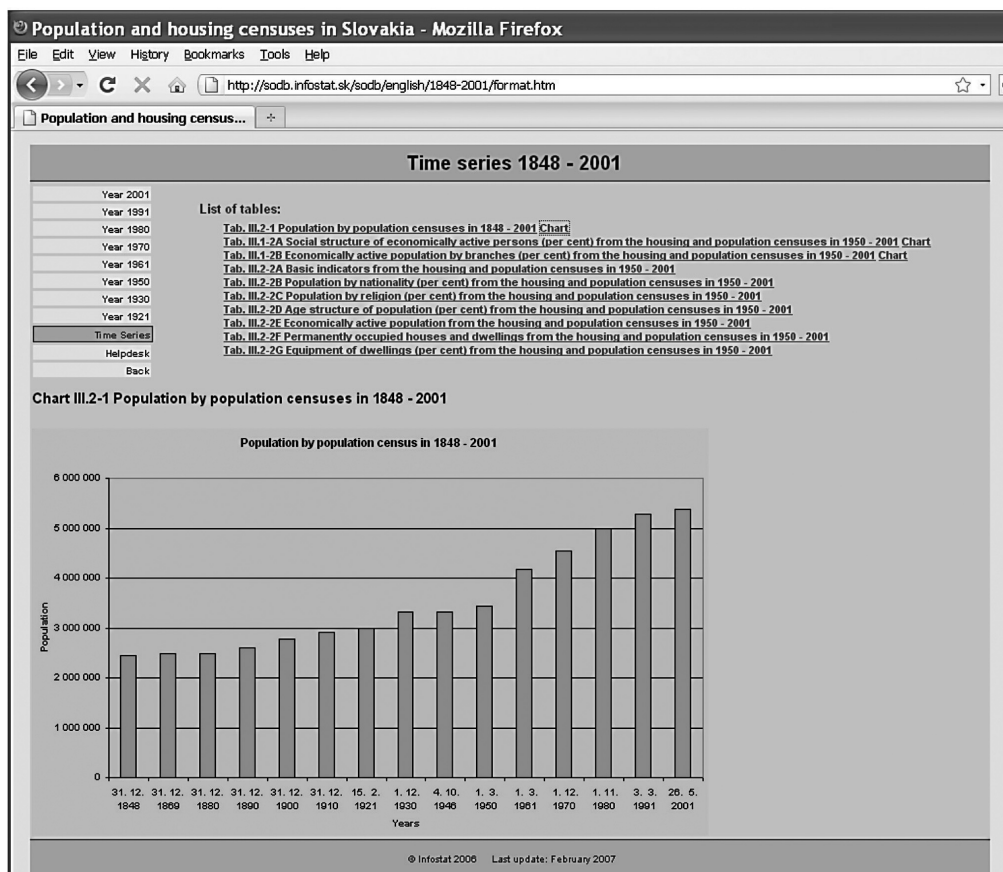


Figure 5: Population development in years 1848 – 2001

4. Conclusion

The MO[MIS] and the Urban Audit websites were already finished and only maintenance of web pages is assumed. Both websites are placed on the website of the Statistical Office of the Slovak Republic. The MO[MIS] website is one of the most visited websites on the Statistical Office web, what proves that the MO[MIS] website is a useful source of the statistical information concerning territorial units of the Slovak Republic. As for the Censuses website, its first version is finished and placed on the INFOSTAT website. In response to the requests, changes in the Censuses website could be realised. For example the enlargement of the amount of presented data can be done by expanding the territorial

units to the municipality level, adding detailed bibliographical part for individual censuses, expanding the length and number of time series, transforming static presentation system from older censuses (censuses till 1970) into dynamic form, adding the search engine working across the website, retrieving and presenting data from censuses older than 1921 census. The presentation should also fully meet the W3C standards and the above mentioned national regulation.

These three projects have specific and common points. The common points are analysed to obtain as unified presentation form and navigation as possible. The common points are analysed with regards to fuse descrip-

tive metadata with metadata for search and navigation into appropriate website design. These websites are running at different web servers (Apache and MS IIS), web pages (jsp, aspx and html) and database management systems (Oracle, SQL Server). The particular solution depends on the hardware and software infrastructure of the respective web server manager - the Statistical Office of the Slovak Republic or INFOS-TAT, where the websites are operated. For end users it is not important because the layout and navigation system of these websites is as unified as possible and the purpose and goals of the each project remain in unchanged form.

Projects mentioned above result in better information of home and foreign community. Relevant websites enable easier accessing the wide variety of statistic indicators describing different territorial levels and time periods. The MO/MIS presents on his website only current data and information. The Urban Audit website presents data for both surveys. The Censuses website presents data and information from almost one century period starting with the 1921 and ending with the 2001 census.

REFERENCES

- [1] Doucek P., "Metainformační systémy a jejich úloha v rozhodovacím procesu", *Statistika* N₀. 4/2005 (344-352), ČSU, Praha, 2005.
- [2] Few S, "Show me the numbers – Design tables and graphs to enlighten", Analytic Press, Oakland, 2004.
- [3] Hudec M., "Practical experience towards designing websites for presentation of statistical data", Joint UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS), Sofia, 2006.
- [4] Hudec M., Büchler P., "Presentation of population and housing censuses in Slovakia on the website", Joint UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS), Geneva, 2007.
- [5] Hudec M., "Web aplikacije za prezentovanje statističkih podataka", Symorg, Zlatibor, 2006.
- [6] Ivory M., Megraw R., "Evolution of Web Site Design Patterns", *ACM, Transactions on Information systems*, V. 23, N₀. 4/2005 (463-497), New York, 2005.
- [7] Inmon W., O'Neil B., Fryman L., „Business Metadata - Capturing Enterprise Knowledge“, Morgan Kaufmann Publishers, 2007.
- [8] Nordbotten S., "Metadata about editing and accuracy for end-users", *Statistical data editing - impact on data quality*, Volume 3 (186-194), United Nations Statistical Commission and Economic Commission for Europe, United Nations, New York and Geneva, 2006.
- [9] Schrey E., "Web-based data dissemination services of the statistical information system Genesis", Joint UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS), Geneva, 2004.
- [10] United Nations Statistical Commission and Economic Commission for Europe, "Guidelines for statistical metadata on the internet", Conference of European statisticians, Geneva, 2000.
- [11] United Nations Statistical Commission and Economic Commission for Europe, "Best practices in designing websites for dissemination of statistics", Conference of European statisticians, Geneva, 2001.
- [12] Ward D., "Draft metadata content standards for statistical metadata on the internet", Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS), Geneva, 2004.
- [13] <http://www.iaidq.org/main/glossary.shtml>, accessed on February, 2008.