

Internet Presentation of CZSO's Metadata

Miroslava Brchaňová and Petr Eliáš, CZSO

Summary

The Czech Statistical Office (CZSO) is on its way towards a new complex statistical metadata system. One of the advantages of the system is that it will enable the CZSO to present its metadata on one place in a uniform format. At the beginning there are three main metadata domains (application subsystems) included in the presentation. These are classifications, nomenclatures and variables. Other domains are to follow in future (statistical tasks etc.). The paper will focus on the three available at the moment.

The presentations of classifications and nomenclatures are built on the same grounds (Classifications and Nomenclatures Documentation System (KLAS)) and are therefore very similar, yet not the same because they reflect the specific technological differences, mainly the hierarchical structure of classifications on the one hand and the flat structure of nomenclatures on the other. Anyway both presentations are a big step ahead in comparison with their predecessors, both as regards comfort for users and the range of information available.

The presentation of indicators / variables is based on the Variables Documentation System (UKAZ). The system is not built in isolation and is linked to other relevant metadata subsystems like the above mentioned Classifications and Nomenclatures Documentation System (KLAS) and Tasks Documentation System (ULOHY) and plays important role in storage data in data warehouse (DWH). The presentation will play its role as a harmonizing tool for names of variables and their definitions.

The three presented subsystems of the whole metadata internet presentation are to be accessed by all kinds of internal or external users directly from the CZSO's home page.

1. Introduction

Acknowledging the importance of metadata the Czech Statistical Office (CZSO) develops a new database (to store and share metadata) to be used throughout the whole office. The database named Statistical Metadata System (SMS) is planned to have around 10 modules which are to cover all the possible statistical needs. Currently there are 3 modules available. The first one serves to maintain classifications and nomenclatures (called Classifications and Nomenclatures Documentation System (KLAS)). The second one is designed to methodically describe indicator / variables (named Variables Documentation System (UKAZ)). The third module being at hand is the one to keep the evidence of statistical tasks (referred to as Tasks Documentation System (ULOHY)). Other modules for keeping track of respondents, administrative data sources, time series, data users, methods of dissemination and so on are to follow.

The access to the SMS database will be possible not only by internal applications but, to some extent, also via internet. The policy chosen is to provide external users with easily accessible metainformation of quite a high level of detail. Regarding the accessibility, it means that there should be a single portal for all the metadata directly accessible from the CZSO's home page. But, of course, if need be, the specific information from SMS's internet presentation (iSMS) can be referred to from any other page of the CZSO's web presentation. All the data presented here must be available in the uniform format. Moreover, the data should be detailed enough to satisfy needs of external users like ministries or banks who cooperate with the CZSO in the data collection.

Currently the old system of metadata internet presentation is being replaced by the iSMS. What is at hand now is a β -version of iSMS that does not have all the functions desired yet, however it enables CZSO to provide users with sufficient information about following metadata domains:

- classifications,
- nomenclatures,
- indicators / variables.

The full version of iSMS is to follow soon. The presentation will be completed with the rest of its functions such as tree-type structure for classifications (like that in Windows Explorer), exports of classifications, nomenclatures and their correspondences, advanced search for indicators / variables, English version of the presentation etc.

Later on, in accordance with the development of the metadata database also the iSMS is expected to be enriched with some of the new modules from the SMS. For the time being it is counted with statistical tasks to be the next.

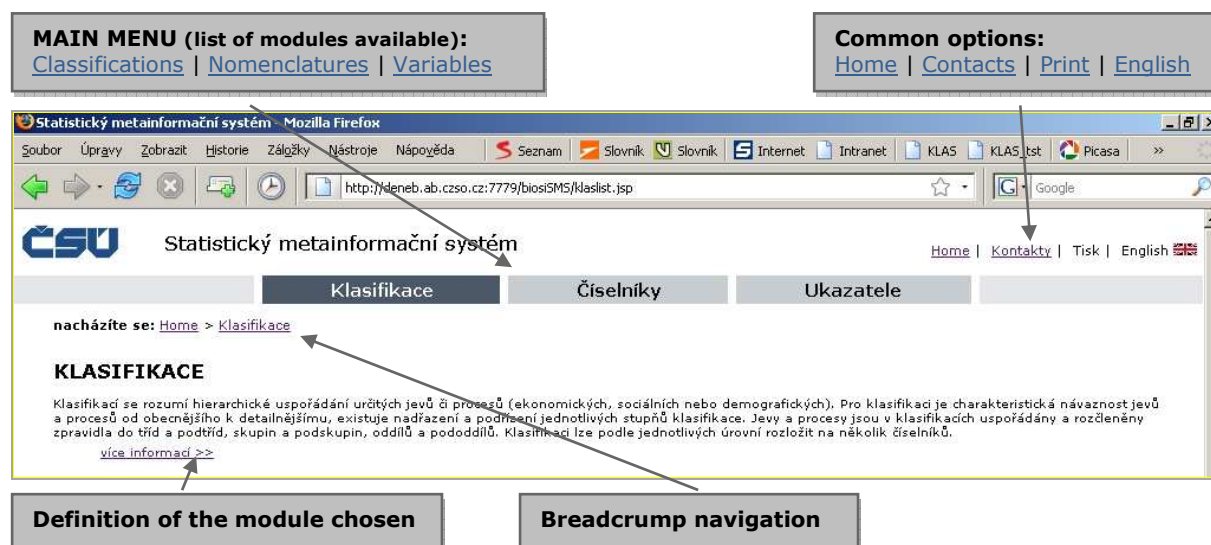
2. Description of the presentation

With iSMS the CZSO's approach towards the internet metadata presentation has been changed. There will be no separate presentations of classifications, nomenclatures or variables any more. On the new web portal they are provided all under one umbrella, all structured in the same manner and all taken directly from the CZSO's metainformation database SMS.

Every single page has the same headline (see *Image 1* below) and footnote. In the headline there is a menu for selection among the metadata domains (classifications, nomenclatures and indicators). The menu together with so-called breadcrumb navigation give users the possibility to check where exactly in the presentation they are and also to browse through the it quickly and easily at any moment. Above the main menu there are common options. Apart from "Home" and "Contacts" links, it will be possible to print the content of the current site and to opt between Czech and English language of the interface.

Below the common uniform headline there is the room for the very metadata. On opening a domain chosen, the domain is introduced by a brief definition rendering it easier for users to orient themselves and understand what data exactly are to be expected here.

Image 1 – Uniform headline of the iSMS



On the bottom of each page there is a common footing providing users with some technical information about the presentation and the address of the CZSO together with the telephone exchange number.

2.1. Classifications

In general, there are tree basic levels of the classifications' presentation. On the first one, there is a list of all available classifications. Here it is possible to choose one and get to the lower level of the presentation where there is all the information about the classification (its structure, definition, links to other classifications or nomenclatures etc.) and where it is possible to download it. The lowest level of the presentation offers information on one item / heading of the classification.

2.1.1. List of classifications

On opening the presentation of classifications users are given the explanation of what exactly is meant by the term “classification” so that they can have an idea of what kind of data to expect here. Right below the definition there is a list of all available classifications – the first level of the presentation. In fact, these are not really all there are in the SMS database. There is the possibility to distinguish between the classifications for public use and those intended just for internal purposes that do not go to the internet presentation.

After opening the list of classifications the presentation gives several possibilities to influence the list. First it is possible to decide to modify the default selection of the currently valid classifications and choose classifications valid on some other date or see all the classifications regardless the date of validity. There is also possibility to make selection according to the classification family (all | activities | products | education | etc. – families respect Eurostat’s distinction). Furthermore, search in abbreviations or names by inserting a string of letters is possible as well.

Results are displayed in pages and users can decide how many lines it is desired to see per page whether 10, 20 or 30.

Image 2 – List of classifications

Table heading:
[Abbreviation](#) | [Name](#) | [Version](#) | [Valid from](#) | [Valid to](#) | [Family](#) | [Additional information](#) | [Download](#)

There are search fields for abbreviations and names, pull-down menu for families.

Statistický metainformační systém - Mozilla Firefox

http://deneb.ab.czso.cz:7779/biosSMS/klaslst.jsp

CSU Statistický metainformační systém

Home | Kontakty | Tisk | English

Klasifikace Číselníky Ukazatele

nacházíte se: [Home](#) > [Klasifikace](#)

KLASIFIKACE

Klasifikací se rozumí hierarchické uspořádání určitých jevů či procesů (ekonomických, sociálních nebo demografických). Pro klasifikaci je charakteristická návaznost jevů a procesů od obecnějšího k detailnějšímu, existuje nadřazení a podřazení jednotlivých stupňů klasifikace. Jevy a procesy jsou v klasifikacích uspořádány a rozčleněny zpravidla do tříd a podtříd, skupin a podskupin, oddílů a pododdílů. Klasifikaci lze podle jednotlivých úrovní rozložit na několik číselníků.

[více informací >>](#)

Výběr klasifikace:

Akronym	Název	Verze	Platnost od	Platnost do	Rodina	Doplňující informace	Ke stažení
CZ-CC_KL	Klasifikace stavebních děl (CZ-CC)	1.1	01.08.2009	09.09.9999	Produkce		
CZ-COFOG_KL	Klasifikace funkcí vládních institucí (CZ-COFOG)	1	01.01.2002	09.09.9999	Národní účty		
CZ-COICOP_KL	Klasifikace individuální spotřeby podle účelu (CZ-COICOP)	1.2	01.09.2000	09.09.9999	Národní účty		
CZ-COPNI_KL	Klasifikace služeb neziskových institucí sloužících domácnostem podle účelu (CZ-COPNI)	1.1	01.09.2006	09.09.9999	Národní účty		
CZ-CPA_KL	Klasifikace produkce (CZ-CPA)	1	01.01.2008	09.09.9999	Produkce		
CZ-ICSE_KL	Klasifikace postavení v zaměstnání (CZ-ICSE)	1	01.01.1998	09.09.9999	Zaměstnání		
CZ-NACE_KL	Klasifikace ekonomických činností (CZ-NACE)	1	01.06.2007	09.09.9999	Činnosti		
CZ-NUTS_KL	Klasifikace územních statistických jednotek	1	01.01.2008	09.09.9999	Území		
ISCED97_OBORY	Mezinárodní klasifikace vzdělání ISCED 97- obory vzdělání	1	01.01.2008	09.09.9999	Vzdělání		
ISCED97_UROVEN	Mezinárodní klasifikace vzdělání ISCED 97- úrovně vzdělání	1	01.01.2008	09.09.9999	Vzdělání		

Stránka: 1 2 ≥ (z 2) zobrazit: 10 20 30 řádků na stránku

Zobrazit klasifikace k datu:
 všechny (bez ohledu na datum platnosti)
 platné k datu: 19.08.2009

Prohlášení o přístupnosti

Český statistický úřad, Na padesátém 81, 100 82 Praha 10; tel.: 274 051 111 (ústředna)
 Verze: 0.2.1 © ČSÚ 2009

Aktualizováno dne: 19.08.2009

http://deneb.ab.czso.cz:7779/biosSMS/klaslst.jsp?kodcis=80004

Date selection:

- all classifications
- classifications valid on:

Results' page selection

Number of items per page

Apart from the filters and searches mentioned above, if need be, it is also possible to change the sorting of classifications. They can be assorted by abbreviations, names, versions, the beginning and the end of validity and also by families.

Once users have found the classification they wanted, they can display its structure (by clicking on its name), respective information (by clicking on the ⓘ symbol) or they can go directly to the download page (by clicking on the ⬇️ symbol). Whatever possibility they choose, they get to the second level of the classifications' presentation.

2.1.2. Classification's detail

On this level one has access to all the information related to the classification chosen (as a whole). The page here is introduced (after the common heading) with the basic identification of the classification, which is, in fact, the information from the classifications' list. Below there is a submenu enabling users to display:

- the classification in tree type structure format
- the classification in the linear form with search engine
- available additional information
- download page

Regardless of what option of these four is chosen, the upper part of the page always remains the same. It is only the detail below the submenu that changes. Thus it is at first glance obvious that all these four pages represent the same level of information about a classification.

Image 3 – Classification's submenu

Identification of a classification (as in the list of classifications):
[Abbreviation](#) | [Name](#) | [Version](#) | [Valid from](#) | [Valid to](#) | [Family](#)

Classification's submenu:
[Structure](#) | [List of items](#) | [Additional information](#) | [Downloads](#)

At present, we only two of these options are working. These are "List of items" and "Additional information". For the two remaining there are just model pictures showing how these other pages are planned to look like.

Under the “Structure” a classification will be displayed in a form of so called tree type structure. It has been inspired by the tree of directories in Windows Explorer. On opening only the first level of the classification will be displayed. By clicking on pluses one will be able to get down the classification to the level of detail needed.

It will be also possible to choose the language of the data set. The option will be available below the classification tree. This will not be the same as the choice of language on the upper right hand side of the page. While the latter changes the interface language between Czech and English (menus, table headings, descriptions...) the first one gives the possibility to retrieve data in a language desired. It is counted obligatorily with Czech and voluntarily with any of English, German and French.

Image 4 – Structure of a classification – example

Tree type structure format of presentation

Klasifikace ekonomických činností (CZ-NACE)

Akronym	Název	Verze	Platnost od	Platnost do	Rodina
CZ-NACE_KL	Klasifikace ekonomických činností (CZ-NACE)	1	01.06.2007	09.09.9999	Činnosti

Struktura klasifikace:

- 1 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
- 2 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
- 3 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
 - 31 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
 - 32 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
 - 321 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
 - 322 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
 - 323 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
 - 33 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
- 4 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
- 5 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
- 6 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
- 7 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
- 8 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX
- 9 XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX

Jazyková verze:

Prohlášení o dostupnosti

Aktualizováno dne: 19.08.2009

Český statistický úřad, Na padesátém 81, 100 82 Praha 10; tel.: 274 051 111 (ústředna)

Selection of the language version of data

The second option from the submenu “List of items” returns a classification in a linear form, as table. While the tree type structure is useful for top-down searching, the linear form is more suitable for displaying a list of codes found by a string search.

Items here can be “filtered” by selecting how many levels of the classifications are to be displayed (i.e. all | 1 | 1-2 | 1-3 | etc.).

Image 5 – List of classification’s items / headings

Linear format of presentation – list of items

The screenshot shows the 'Statistický metainformační systém' interface. The main content area displays the classification 'Klasifikace územních statistických jednotek' in a linear format. A table lists the classification items with columns for 'Úroveň', 'Kód', and 'Název'. Below this table, there are search fields for 'Kód' and 'Název'. Annotations with arrows point to the 'Úroveň' dropdown menu and the search fields, with labels: 'Selection of number of levels to be displayed' and 'Search fields for codes and names'.

Akronym	Název	Verze	Platnost od	Platnost do	Rodina
CZ_NUTS_KL	Klasifikace územních statistických jednotek	1	01.01.2008	09.09.9999	Území

Úroveň	Kód	Název
vše		
1	CZ	Česká republika
2	CZ0	Česká republika
3	CZ01	Praha
4	CZ010	Hlavní město Praha
3	CZ02	Střední Čechy
4	CZ020	Středočeský kraj
3	CZ03	Jihozápad
4	CZ031	Jihočeský kraj
4	CZ032	Plzeňský kraj
3	CZ04	Severozápad

The iSMS presents not only classifications themselves but also some other relevant information related to them. It is the third option of the submenu where this kind of information is accessible.

In general, there are the following types of additional information that may be available for a classification:

- definition of a classification
- list of classification's levels
- correspondences with other classifications and nomenclatures (i.e. record that items of the classification have correspondences with items of the classification or nomenclature in this list)
- list of classification's attributes (characteristics)
- specific external files related to the classification (e.g. methodology, overviews, schemes, legal acts etc.)

Naturally, not every classification has to have all the possible information recorded so the range of information here may vary from classification to classification.

Image 6 – Additional information about a classification

Statistický metainformační systém

nacházíte se: [Home](#) > [Klasifikace](#) > [CZ_NUTS_KL](#)

Klasifikace územních statistických jednotek

Akronym	Název	Verze	Platnost od	Platnost do	Rodina
CZ_NUTS_KL	Klasifikace územních statistických jednotek	1	01.01.2008	09.09.9999	Území

Struktura | Seznam položek | **Doplňující informace** | Ke stažení

Číselníky úrovní klasifikace

Úroveň	Název číselníku
1	ZC: NUTS 2004, úroveň 0 - stát
2	ZC: NUTS 2004, úroveň 1 - území
3	ZC: NUTS 2004, úroveň 2 - oblast
4	ZC: NUTS 2004, úroveň 3 - kraj

Vazby

Akronym	Název	Verze
CISPOU	Správní obvody obcí s pověřeným obecním úřadem	
CISORP	Správní obvody obcí s rozšířenou působností	
OKRES_LAU	Číselník okresů	
NUTS4-2004	ZC: NUTS 2004, úroveň 4 - okres	
NUMNUTS	Zpracovatelský číselník NUTS	

Prohlášení o přístupnosti [?](#)

Český statistický úřad, Na padesátém 81, 100 82 Praha 10; tel.: 274 051 111 (ústředna)
Verze: 0.2.1 © CSÚ 2009

Aktualizováno dne: 19.08.2009

Hotovo

Classification's levels:
[Level](#) | [Name of the level \(a separate nomenclature\)](#)

Links:
[Abbreviation](#) | [Name](#) | [Version](#)

The last option of the submenu is reserved for downloads. Users will be given the possibility to download the classification itself, its correspondences or any of the external files discussed on the previous page.

Before downloading a classification or its correspondences the presentation will enable users to adapt the data to their individual needs. Thus it will be possible to indicate whether the file with the classification will contain just the classification's structure or even explanatory notes and / or attributes.

Similarly the correspondences could be downloaded just in form of respective codes or accompanied with texts. For both there will be the chance to choose the language of all the texts.

Users will have the choice between the XML and DBF file. As a standard, the presentation will prefer the XML format than can provide full detail of information. However, not to discriminate users with older systems the application will be able to return DBF files as well, even though, the format has its limitations reflected in the absence of longer texts.

Together with these two downloads there will be the users will be allowed to download external files with methodology, overviews, schemes, legal acts etc. These files could have various formats like for example doc, xls, pdf, jpg.

Image 7 – Classification’s downloads – example

Download of a classification itself:
[Levels](#) | [With definitions](#) | [With attributes](#) | [Language](#) | [File format](#)

Classification’s correspondences:
[Correspondence with...](#) | [Valid on](#) | [With texts](#) | [Language](#) | [File format](#)

Related external files:
 Methodology, overviews etc.

2.1.3. Information on a classification item / heading

The third (lowest) level of the presentation is represented by the detailed information about a single classification’s item. Again, below the common heading there is the identification of the classification and below it the chosen item is identified. Further information like explanatory note, attributes, corresponding items from related classifications or nomenclatures may follow if available.

Image 8 – Information on a classification item / heading

Identification of a classification:
[Abbreviation](#) | [Name](#) | [Version](#) | [Valid from](#) | [Valid to](#) | [Family](#)

Identification of an item (heading):
[Level of the classification](#) | [Code](#) | [Name](#)

The screenshot shows the 'Statistický metainformační systém' interface. The breadcrumb trail is: Home > Klasifikace > CZ_NUTS_KL > Střední Morava. The main heading is 'Klasifikace územních statistických jednotek'. Below it is a table with columns: Akronym, Název, Verze, Platnost od, Platnost do, Rodina. The row for 'CZ_NUTS_KL' is highlighted. Below this is the 'Střední Morava' section with a table for 'Úroveň', 'Kód', 'Název' showing '3', 'CZ07', 'Střední Morava'. There is also an 'Atributy' section with 'Kód RSO' 272 and a 'Vazby' table showing relationships with 'Česká republika', 'Střední Morava', 'Olomoucký kraj', and 'Zlínský kraj'. At the bottom, there is contact information for the Czech Statistical Office and a date 'Aktualizováno dne: 19.08.2009'.

Item's attributes

Item's correspondences:
 Classification /nomenclature | Code | Name

2.2. Nomenclatures

The iSMS presentation of nomenclatures looks similarly to that of classifications. However there are some basic technical differences.

- In the database, classification is just a virtual object build upon respective basic physical nomenclatures and their links creating the classification's structure. Therefore, nomenclatures are identified by codes and abbreviations but classifications do not have their codes.
- Change in a classification (new items added, old deleted) is made by creation of its new version. A version cannot change in time. It is static. When it is desired to get an older classification, it is done by finding the version that was valid then. On the contrary, nomenclatures can be dynamic. Items of one nomenclature can develop over time. To find a nomenclature valid on certain date one just selects the desired date and the application returns all the items valid that day. A nomenclature has no versions.

Despite these differences, the presentation of nomenclatures is built upon the same foundations as the presentation of classifications. The menus, options, tables, selection principles are all used the same throughout so that to avoid confusion and make work with the presentation as comfortable as possible.

2.3. Indicators

A new sub-system SMS so-called "UKAZ" is presented on the CZSO's Internet. The system includes statistical indicators (statistical variables) from all areas of official statistics. Search in the system currently uses standard procedures. In future, the presentation will be available also in English version and will allow printing of the lists of indicators and their structure.

2.3.1. Current state of the Internet Presentation

The presentation of variables is based on the Variables Documentation System (UKAZ). The system is not built in isolation and is linked to other relevant metadata sub-systems like Classifications and Nomenclatures Documentation System (KLAS) and Tasks Documentation System (ULOHY) and plays important role in data storage in data warehouse (DWH). The presentation will play its role as a harmonizing tool for names of variables and their definitions.

For the reasons of historical continuity and habits of the majority of external users the subsystem of variables is called "Indicators".

Currently, there are the following three possible ways of how to **search for an indicator** – by alphabet, by full-text or by a topic. An Example: I can find "Evidenční počet zaměstnanců" (Number of employees) under letter "E" or by full-text search "Eviden" and/or under topic Economic statistics.

Image 9 – Indicator selection

Indicators

Statistický metainformační systém - Mozilla Firefox

Soubor Úpravy Zobrazit Historie Záložky Nástroje Nápověda Seznam Slovník Slovník Internet Intranet KLAS KLAS_tst Picasa

http://deneb.ab.czso.cz:7779/biosSMS/ukazvyb.jsp

CSU Statistický metainformační systém Home | Kontakty | Tisk | English

Klasifikace Číselníky **Ukazatele**

nacházíte se: Home > Ukazatele

UKAZATELE

Statistický ukazatel vymezuje obsah statistikou sledovaného jevu. Struktura ukazatele obsahuje texty a atributy. Každý ukazatel má přidělen unikátní identifikační kód. Mezi texty patří „plný název“, „definice“ a „synonyma“ pro využití v dotaznících a výstupních tabulkách. Atributy nám mohou říci, zda je ukazatel za období nebo k datu, numerický z čísel nebo znakový s písmeny, kvantitativní o počtu nebo kvalitativní „vdaná“, byl primárně zjištěn nebo odvozen výpočtem, lze jej počítat nebo ne a je dále zařazen do nějaké skupiny statistických zjišťování např. majetku nebo pojištění.

[více informací >>](#)

Výběr ukazatele:

podle abecedy

A B C Č D E F G H CH I J K L M N O P Q R S Š T U Ú V W X Y Z vše

vyhledání textu v ukazatelích

Hledej

podle tématu

1. Demografické a sociální statistiky
2. Ekonomické statistiky
3. Životní prostředí a průřezové statistiky
4. Metodika sběru, zpracování, diseminace a analýzy dat
5. Strategické a řídicí problémy oficiální statistiky

Prohlášení o přístupnosti

Český statistický úřad, Na padesátém 81, 100 82 Praha 10; tel.: 274 051 111 (ústředna)
Verze: 0.2.1 © ČSÚ 2009

Aktualizováno dne: 24.08.2009

Hotovo

Selection of an indicator:

- by alphabet
- by full-text search
- by a topic

A statistical indicator defines content of a phenomenon being searched for. **An indicator structure** contains texts and attributes (characteristics). Example here is the indicator: “Emisní ážio” (“Share Premium”) is the amount by which the price at which a company sells a share exceeds its par value. Most shares are issued at a premium to their nominal value.)

Image 10 – Detail of an indicator

Indicator „Emisní ážio“

Structure of an indicator:

- full name
- code
- definition
- synonyms
- attributes

Each indicator has its unique code. **Texts** are divided into “full name”, “definition” and “synonyms”, which are used in questionnaires and output tables. **Attributes** qualify if an indicator describes a period or a date, it is a numerical or character with letters, it is a quantitative (number) or qualitative (married), it is a result of a survey or is estimated, it is an accountable or not and it is a member of a survey group e.g. property or insurance.

2.3.2.Next steps

In the nearest future, there will be some other features added to the website such as English texts, prints (Czech and English) and choice of topics will be possible from categories arranged in a tree type structure into several levels. Also, under the structure of indicators, there will be displayed a list of related indicators.

In farther future we expect to add a possibility to extend the above shown search possibilities with search by attributes, definitions and synonyms.

3. Conclusion

Our past experience showed benefits of a dynamic website presentation. Former presentation was heavily used not only by CZSO's staff, but also by students and many others who needed explanation of a concept used in CZSO's statistics (e.g. to the "Share premium").

Having learned from the advantages as well as from drawbacks of the old metadata presentation, we tried to develop a tool that will be able to satisfy most of the users' needs in hope that the new Internet presentation of metadata will be even more successful than the previous one.