

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

**ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 9-11 February 2004)

Topic (iii): Metadata models and terminology

Which came first – terminology or models?

Contributed Paper

¹Submitted by the Czech Statistical Office (CZSO)

I. INTRODUCTION

1. Creation and updating of so called “description of indicators” is one of the basic instruments, which contribute to the transparency and to unambiguous and unified definition of the Czech state statistical service. At the same time it is one of basic conditions for mutual linkage of data and metadata and for metadata propagation of statistical information.

II. What is currently included in the CZSO METIS?

2. At present there are some parts of the METIS system created in the CZSO (according to the UN ECE terminology and not according to the historically originated names inside of our office). It concerns especially the following parts:

- METIS indicators (statistical characteristics)
- METIS code books (classifications)
- RES (including statistical units).

3. The following is in the stage of creation:

- database of forms’ identification codes (IKF)
- database of outgoing tables from processing
- public database (called PDB)
- catalogue of publications.

¹ Prepared by Miroslava Brchanova, brchanova@gw.czso.cz

4. We suppose to create:

- additional terminology vocabulary
- list of important users (national and international).

5. In the instrument for preparation of surveys, creation of reports and projects PROJEKTMAN there is information stored concerning:

- units of statistical report forms
- mathematical methods for surveys and grossing up.

6. Mutually linked can only be parts of the indicators, code books and prepared PDB, Linking of the indicators to PDB may be done by the way which is substitutional and it artificially implants into the PDB data, which should be linked automatically during the stage of processing.

7. That is why at present the PDB, which works with metainformation of outgoing data, must supply metainformation from the outside and in artificial way. On the other side you may conclude from it that the present stage of indicators as such is corresponding with the present CZSO surveys' technology of preparation and processing and in fact is in convenience with present requirements.

III. METIS indicators (their description)

8. Description of indicators provides:

- control of content and grouping of statistical indicators (statistical characteristics), variables and indices,
- detailed identification of statistical information content inevitable for securing content unification of the information in the process of surveying, processing and for analytical activities,
- the base for creation of the data fund description,
- creation of unified metadata concentrated at one accessible place.

9. Objects of the description are:

- statistical data and statistical indicators (statistical characteristics), variables and indices, which are the part of statistical surveys organised by CZSO according to the State statistical service law.

10. Procedure in the process of submitting, approving and presentation of indicators descriptions:

- the participants of the above mentioned process of creation and updating of indicators description are besides "Subject matter departments" also database administrators for "Indicators" and "Code Books"; updating is made according to needs of subject matter departments continuously;
- proposals for description of new statistical indicators (statistical characteristics), variables and indices and for any changes are done by the subject matter departments;

- metadata unit then makes assessment; it decides if they are accepted in accordance with the current methodology for doing indicators description and in the case of accepted proposals it gives instruction to the database administrator to include the indicator into database;
- the database administrator continuously presents the statistical indicators (statistical characteristics), variables and indices and enables the access to described the statistical indicators (statistical characteristics), variables and indices.

IV. Problems with terminology of derived statistical indicators (statistical characteristics)

11. Filling in the PDB has been started by creation of active form of CZSO yearbook. The problem, we have encountered, is the problem of specification, calculation, names and definitions of derived indicators.

12. Besides standard basic indicators, as we have “Number of marriages”, “plant production” or “GDP”, in the tables of the yearbook (and of course in other places) there are indicators as “Number of marriages per 1000 inhabitants” or “Proportion of plant production in the gross agriculture production” or “GDP per 1 inhabitant in PPP, where the EU = 100”. Are these new, other indicators? Of course. But is it desirable to have them in the “Catalogue of indicators” as any new, other ones?

13. The principle is a rational consideration on the topic that it is not necessary to store in the base besides absolute indicators also relative indicators and indices, because they always originate as a function of absolute indicators.

14. The problem starts in the moment of practical realisation of this idea. It is really possible to create a relative indicator as a function of absolute indicators stored in the database, in a way so that you define their corresponding function (usually their proportion). But you cannot in this way create the names (and often also even definitions) of relative indicators. E.g. the indicator which has originated as the proportion of production and number of employees, is called “Work productivity”. You cannot create for it only a technical name like “Proportion of production per 1 employee”. Similarly it concerns terms “Population density”, “Specific emissions” etc.

15. Moreover during the creation of indicators’ names there might occur problems that such indicators have not existed before and the new ones (according to the name) would not link with up to now existing time series of derived indicators.

16. Because of the fact that there is quite a big number of derived indicators, which have their established and specific names (Work productivity, Stock turnover time, Profitability, Costs, Ageing index, Economic burden index, Investment rate, Flexibility index, etc.) it would be desirable to maintain besides the database of absolute indicators also database of indicators with derived names.

17. The chemistry workers who are in comparison followed by the statisticians with 200 years handicap use a system similar to ESA 95 with clear codes. They describe hydrogen as H_2 and oxygen as O_2 . From the point of view of PDB technologists it has been supposed that originated H_2O (water) may be described by single definition of hydrogen and oxygen

properties. It may be partially done in this way. But moreover after putting H₂ and O₂ together there will appear a completely new quality (H₂O) which requires its own definition of water. Moreover the terms H₂, O₂ and H₂O are being used rather in the field of specialists. Common people prefer the use of the name “water”. And PDB specialists should serve both to specialists and also the broad public.

Example:

Name of the indicator	Definition of the indicator	Calculation of the indicator
Energy process efficiency rate	It concerns measuring of the energy use efficiency	$\frac{\text{Total energy output} \cdot 100}{\text{Total energy input}}$ (%)

V. Statistical indicator (statistical characteristics) versus index

18. A statistical indicator is a specific sort of statistical characteristics, as it utilises only limited number of functional regulations (in most cases the total). Thus the indicator (statistical characteristics) is a quantity variable. Each variable gets certain value in dependence to its definition and indicators are always spoken in relation to numeric values. There is then the question coming, in which way to obtain from the variable any number, i.e. concrete value that is data.

19. Statistical indicator is a statistical characteristics and it is apparent that this statement anticipates in implicit way, that the statistical set is in general defined in the space and time. If we take into consideration for example an indicator “Time worked”, then this indicator is defined in the methodology regulations as a total of the time worked by the staff (or labourers) of the company (or its outlet) in some month (or quarter, or year). If we precisely define the time and space (e.g. February 2003, enterprise Alpha), then we obtain concrete value of the indicator for concrete object.

20. Indicator is thus a variable and the indicator’s value is the value of this variable, which may arise from concrete definition of time and space.

21. **Primary indicators** are the indicators surveyed directly, not derived, e.g. time worked, number of employees at certain date, stocks volume etc. It concerns indicators where you can simply specify the sort of the characteristic (e.g. average, total, standard deviation, meridian), statistical units or objects (e.g. school, enterprise, person) and also variables (e.g. age of a student, production, height). You can do some more mathematical operations with these indicators (e.g. division, subtraction).

22. Another sort of indicators are the **secondary, derived indicators**, which may occur three ways:

- as a function (usually the difference or proportion) of various primary indicators – e.g. profit, value added, etc.;
- as a function of various values of the same primary indicator – here we can name all type averages, structure indicators like gross turnover;
- as a function of more than two primary indicators (i.e. combination of preceding proceedings); as an example we can mention relative indicators, where at least one is an arithmetic average (work productivity per 1 labourer) or the function of more primary indicators (value added proportion in the total production).

23. In connection with the indicators' specification to primary and secondary there is a question, where to classify indices, absolute differences and other similar differences. Are these quantities also indicators or not?

24. Indices, absolute differences and other similar differences are instruments for comparisons and instruments for analysing the results of comparisons. The indicators themselves can speak about some reality, but they do not make their assessment, while the indices and absolute plus differences measure the difference of two values of the same indicator, analytical measures or difference make the evaluation of such differences.

- **Note:** If some authors consider indices, absolute difference and analytical measures of differences as indicators, they usually specify for them a special place, e.g. by classification of indicators to authentic and non – authentic. Those authentic ones (which are primary and secondary) describe certain economic phenomena, certain reality; non – authentic indicators serve as comparison of values of authentic indicators and for assessment of discovered differences. Non – authentic indicators are then the instrument for comparing and analysing of value differences of authentic indicators. Non – authentic indicators are always secondary ones.
- Besides the indicators' classification to primary and secondary, the classification of indicators to absolute and relative is also important. **Absolute indicators** express the extent of certain phenomenon without any relation to other phenomenon. To this category belong all primary indicators and also some secondary indicators (time averages, indicators of gross turnover, etc).

25. **Relative indicators** express the size of one phenomenon per one measurement unit of other phenomenon. Relative indicators are always secondary, as they are created as proportion of absolute (primary and also secondary) indicators.

26. The basic problem is how to differentiate exactly, what is a relative indicator and what is an index. In this case we can differentiate **three basic groups of relative indicators** :

- Relative indicators originated as a proportion of two different indicators, the values of which are expressed in different measurement units (work productivity as a proportion of production and number of workers, prices, GDP per 1 inhabitant, etc.). Here you can put also relative indicators which have average character (average rain falls, average water flow, etc.).
- Relative indicators originated as proportion of two different indicators, the values of which are expressed in the same measurement units (production costs, investment rate, etc.).
- Relative indicators which have a character of structural figures, i.e. proportion of the part in the total (proportion of protected territories in the total territory area, etc.).

27. Every relative indicator is bringing qualitatively new information, which is not supposed to be comparison, although each proportion of two indicators means certain form of

comparison of values. **But the objective is not to make comparison itself as such** (e.g. value of the numerator is 20 % higher than the value of denominator) but the value itself of the relative indicator, with its qualitatively new speaking ability.

28. **Index** is a relative measurement unit of difference. It is dimensionless figure showing how many times is value in the numerator higher (or lower), than the value in the denominator. **Absolute increase** than shows the difference value of minuend and subtrahend. E.g. if the profit in the company Alpha was in 2002 180 million CZK and in 2003 200 millions, then we know that the profit of Alpha company has increased 1,1 times, or in fact by 11% (index), i.e. by 20 million CZK (absolute difference).

VI. Resume

29. The paper deals with the problems of metadata as they are solved in the CZSO. It describes terminology needed in this field (primary, secondary, absolute and relative indicators and the question of “indicator versus index”). Also mentioned are problems with measurement units.