UNITED
NATIONS

E

Economic and Social
Council

Distr.
GENERAL

CES/AC.71/1999/21
7 December 1998

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 15-17 February 1999)

Topic (iii): Integration of statistical activities at the national and international levels, including data modelling strategies and standards needed for statistical data integration

MODELLING OF DATA PROCESSING AT THE CENTRAL STATISTICAL BUREAU OF LATVIA

Submitted by the Central Statistical Bureau of Latvia[1]

I. INTRODUCTION

1. The Integrated Data Management System (IDMS) at the Central Statistical Bureau (CSB) of Latvia has the following objectives:

- fully automated statistical data activities;
- prevention of routine programming;
- optimally automated data processing and dissemination process.

2. The main objective is to design a Data Warehouse system as a software platform for IDMS. The fundamental principles of the Data Warehouse approach correspond to the specific demands for a statistical information system. In practice, it means working with several databases simultaneously, accessing and coordinating complex information interactively. IDMS should allow to manage extremely large databases with a widespread content: an extraordinarily

[1] Prepared by Karlis Zeila and Dzintra Dragune.

large number of different kinds of data with highly complex inter-relations.

2. First, it is necessary to determine the complete structure of the Statistical Database System. In order to make this task easier, it is necessary to create a comprehensive description of CSB as a system. One possibility is to use graphic modelling. For many reasons (mainly because specialists from the Latvian University are participating in the design of the GRADE) GRADE Modeller was chosen as a tool-set for system modelling.

3. A detailed description of the structure and functionalities of the existing system was created. The following items had to be identified:

- how the CSB operates;
- how the data processing is carried out;
- what is the task of each CSB structural unit;
- what are the data flows within the organization;
- what (if any) are the links between different surveys;
- external data links.

4. It is not possible to formally describe the staff initiative and creative approach but it can be seen in data flows. Such graphical descriptions or models should be comprehensible and should not be misunderstood.

5. In order to start the data flow modelling, it is very important to know the exact scope and volume of the statistical activities, i.e. the state administration’s and other clients’ (including international organizations) requests for social and economic data. The activities are forecast in the annual CSB national statistical data programme in accordance with the Latvian Law on Statistics. The national programme covers the most important statistical data that is collected, aggregated and published by CSB, as well as the data received from other institutions and published in the statistical publications within the frames of the allocated state budget. The national programme also includes separate activities, which are fully or partly financed by the PHARE national programme. All these activities are vital for Latvia’s integration into the European Union.

6. Experts from the developed European countries and international statistical organizations have participated in the preparation of survey programmes. PHARE funds are utilised because the interviewers’ services are expensive and the CSB budget cannot cover this activity. The national programme consists of 29 themes which describe the different statistical reports. 154 initial reports were used in their creation.

II. GRADE MODELLER AS A TOOL-SET FOR SYSTEM MODELLING

7. GRADE is a tool-set that operates in the business modelling language GRAPES-BM. They were created simultaneously. The business modelling language in this case GRAPES-BM, is a semi-formal graphical language aimed towards

modelling and simulation of complex business systems (production process, offices, information systems, etc.). It contains convenient facilities for modelling of data flows, and of the organizational structure of a system. Such modelling is a starting point for the understanding and re-engineering of any business system.

8. A system in GRADE represents a collection of subsystems and models. In GRADE, the information contained within a model is summarized via two different lists of contents: the model tree window and the model dictionary window. The basic elements of the model tree are diagrams. The basic elements of the model dictionary are names, which occur in the model. Each occurrence of a name is classified either as a definition or as a reference.

9. GRADE model of the CSB of Latvia contains data on current information processing data flows. It includes the following information: volumes of data flows; time of receipt of information on document production (array, table, etc.); person days needed for particular steps of information processing; the used program environment; and, in some cases, other relevant information.

10. All this information is represented with business modelling tool GRADE 4.0 in the form of diagrams. The information is grouped according to organizational units involved (i.e. for each unit there is a separate BUSINESS PROCESS (BP) diagram). However, this grouping has only a technical character. It permits to split the models into diagrams of more or less manageable size. At the same time, from the GRADE semantics' point of view, it can be considered as a single model for all CSB. In total, the model contains information about approximately 3500 data objects.

11. The current GRADE model depicts the overall situation in the Central Statistical Bureau of Latvia for the year 1997. It is intended to give a complete picture of all information processing activities that occur in CSB during a one-year period. However, since the model was constructed over a period of several months and the situation is continuously changing, it may be slightly incorrect for any specific moment.

12. **The structure of the GRADE model** contains the following diagrams (provided in the Annex):

- CLASSES (CL) diagram which formally describes objects contained in the model and the static structure of a business system during the first stages of modelling (Annex, figure 1);

- ORGANIZATION (ORG) diagram which describes the organizational structure of CSB, external information providers and receivers, and also several "artificial" performers that permits to group outgoing documents and publications in a more convenient way (Annex, figure 2).

- BUSINESS PROCESS (BP) diagram which depicts organizational units of CSB, data flows between these units, and those incoming documents that are

being processed by several units.

- A separate BP diagram for each organizational unit which depicts incoming documents processed by this unit, outgoing documents prepared by it, intermediate data arrays and work tables used during information processing, and data flows which describe the way how the incoming documents are used to produce corresponding outgoing documents. In addition, the diagram contains timers which indicate when a particular data object is being received or produced (Annex, figure 3).

- A TASK DETAILS (TD) diagram for each of the data objects that appears in BP diagrams mentioned above, showing attribute values of particular data object (volume, effort in person days, programme, environment, etc.). The TD diagrams also show the immediate environment of corresponding data object, i.e. sources of incoming and recipients of outgoing data flows (Annex, figure 4).

13. In addition, the model contains several more technical diagrams. In general these are needed only for simulation of the model. The model also includes diagrams that are automatically produced by GRADE modelling tool.

14. **Model tree**: all GRADE diagrams are included in a so-called model tree. Figure 5 in Annex shows a part of a sample model tree relevant to the GRADE model of CSB.

15. As mentioned above, many BP and TD diagrams have a tree-like structure attached to the Business Model (BM) diagram, as is the case with the highest level BP diagram “Valsts Statistikas Komiteja”.

17. Below the highest level BP “Valsts Statistikas Komiteja” are attached BP diagrams for each of the CSB organizational units described by this model, and BP diagrams corresponding to incoming documents which appear in the highest level BP. Similarly, below BP’s describing organizational units are attached BP diagrams for all objects (incoming documents, data arrays, work tables, outgoing tables) appearing in BP for this particular organizational unit.

18. A TD diagram, corresponding to each BP diagram, is shown at the right side of it. In this model, TD diagrams are used which correspond to particular data objects (incoming documents, arrays, tables, outgoing documents) and they contain attribute values for these data objects. These TD diagrams are generated automatically from BP diagrams.

19. **Diagram layouts**: some of the BP diagram layouts can also be used to obtain information that initially is not very visible. **Vertical line** and **horizontal line** layouts group tasks by performers within a particular BP diagram. A reduced size sample diagram is appended (Annex, figure 6).

20. Oblique lines layout permits to split into clusters diagram objects that are relatively weakly connected. A reduced size sample diagram is appended (Annex, figure 7).

21. The GRADE diagrams enable easy navigation using joint data groups (surveys in statistics) or data objects (tables, requests etc.). Navigation can be also done according to synonyms or object colours, as well as at different aggregation levels or via multidimensional system levels of the CSB graphic model. Dictionary filters can also be used for this purpose.

22. The **data dictionary** included in the model permits selecting and navigating between diagrams, satisfying a particular property. This possibility is not very powerful, since the required properties can be described only by filters; however it still allows extraction of a satisfactory volume of data.

23. The dictionary has the following functions:

- Overview of all names which occur in the model (or a part of them: filtered via name patterns and/or name types;
- Overview of the diagrams where a name has been defined;
- Overview of all diagrams in which a name is used (i.e. referenced);
- Easy navigation to the diagrams found (and conversely - easy navigation from a name which occurs in a diagram to the relevant dictionary information);
- Overview of errors by the syntax analyser;
- Global and selective name changes in the model;
- Copying the contents of the dictionary to a text file;
- Deletion of names that are no longer used.

All these functions of the dictionary can be reached through the dictionary window.

24. The provided example (Annex, figure 8) shows that data is accessed from the labour statistics diagram via a dictionary filter about the external users which is exported to a text file, and then to an Excel file (Annex, figure 9).

25. **Export/import facilities**: GRADE diagrams, complete model or some part (sub-tree) of it can be exported and imported in a so-called EIF format (in files named *.eif). EIF format can be used for the exporting of GRADE model to external applications.

26. Another kind of export is offered in HTML format. It enables the production of a *.html file that allows to view the model with an HTML browser and offers navigation facilities similar to GRADE. Note that no zoom facilities will be offered. Therefore, before exporting, the scale in which diagrams will be exported must be chosen (it will be the same as the current zoom factor for diagrams). Even so, the export of large diagrams can be

problematic – it will either give diagrams that are too reduced in size, or diagrams that are very large and difficult to view with ordinary HTML browsers.

27. It is not necessary initially to create a concrete structure for the model. It is possible to start the work and allow the system to develop in the desirable direction. The data object approach should include not only evaluation but also modifications.

28. In order to acquire more information about the GRADE tool, see Internet, Home page: http://www.latnet.lv/LU/MII/Grade/default.htm or Hot line service: e-mail to grade@cclu.lv.

III. CONCLUSION

29. The development of the Integrated Data Management System usually requires a significant investment of resources. Each new situation requires a new solution. Use of a standard modelling tool can be a quick and economical solution as the local research work will be simplified. The GRADE tool might need to be enhanced to enable model simulation. This would allow to spot problem areas in the statistical data processing which need extra attention.

30. The graphic CSB model will considerably facilitate the following tasks:
(a) the creation of the metadata base structure (the model helps to determine the order in which it should be created and developed according to user requirements);
(b) using metadata for correct interpretation and use of the data;
(c) searching for existence and availability of data;
(d) evaluating the quality of data.

31. GRADE model enables to view the CSB organization of work from a “bird’s eye”. Therefore, the GRADE model will be used not only as a basis for the design of the data management system. It will also simplify administrative decision-making as it gives an unquestionable insight into the organization of data flows in each substructure and in CSB in general. It will, of course, require regular testing and upgrading of the data in this graphic model.

32. Further collaboration with GRADE Modeller designers would be desirable in order to continue the upgrading of this graphic tool so that the current outputs are not lost and the use of technical innovations in this area are optimized.

ANNEX
(Figures)

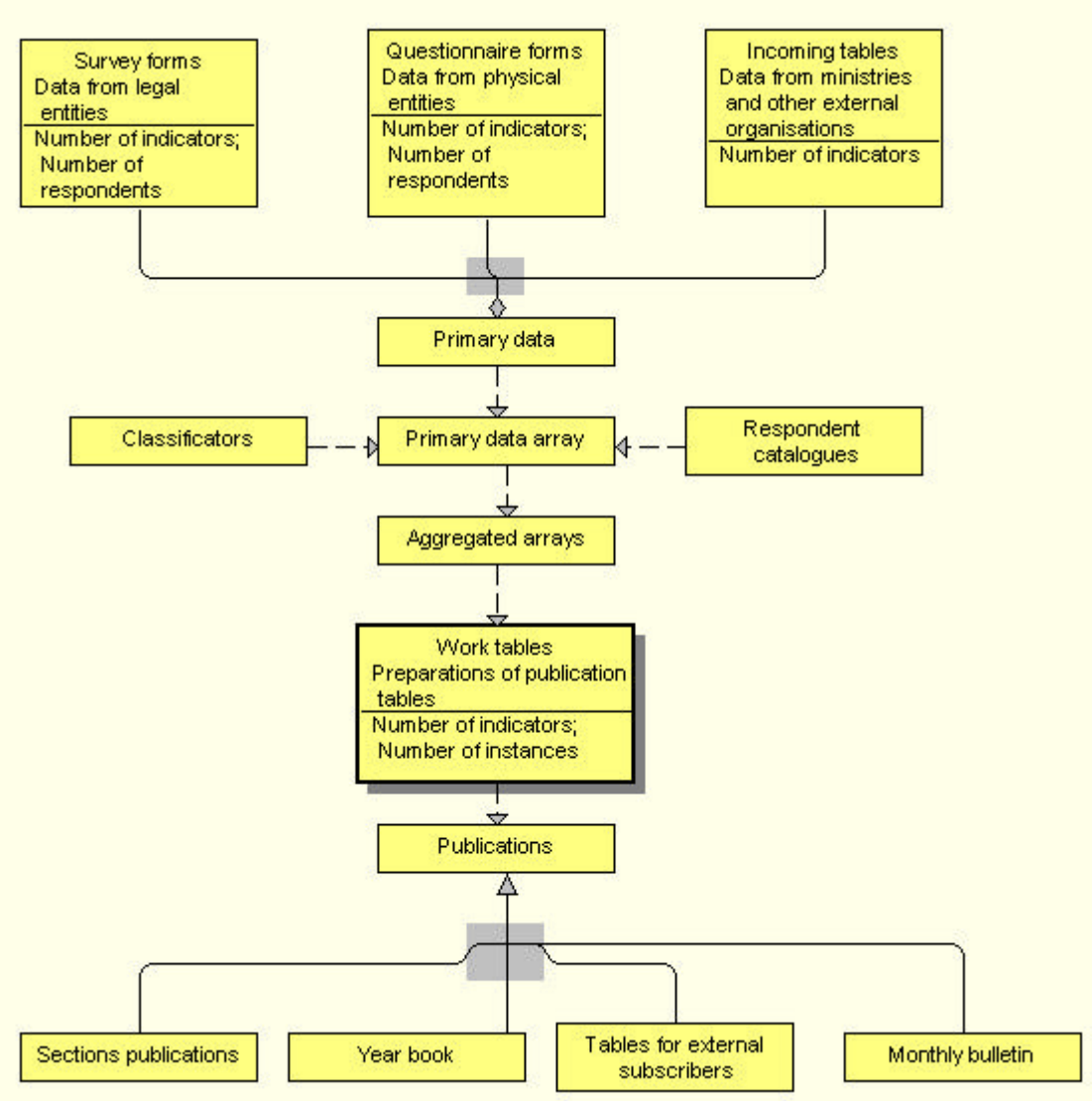

Figure 1. CLASSES diagram

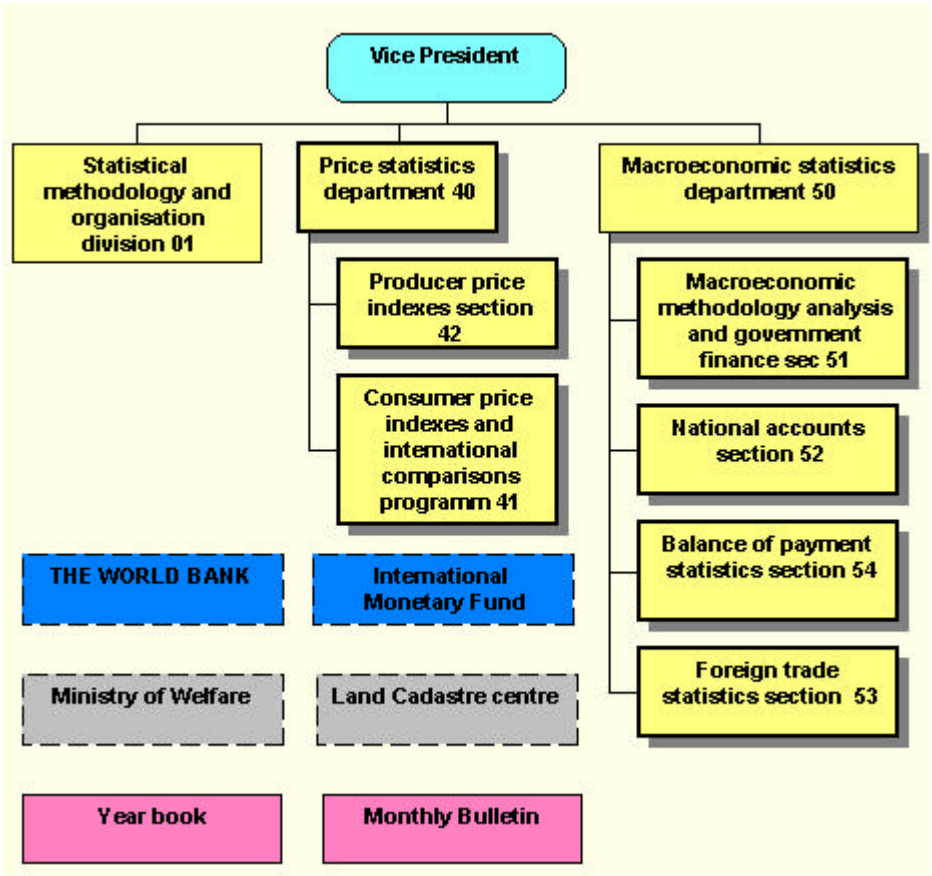

Figure 2. Fragment of the ORGANIZATION diagram

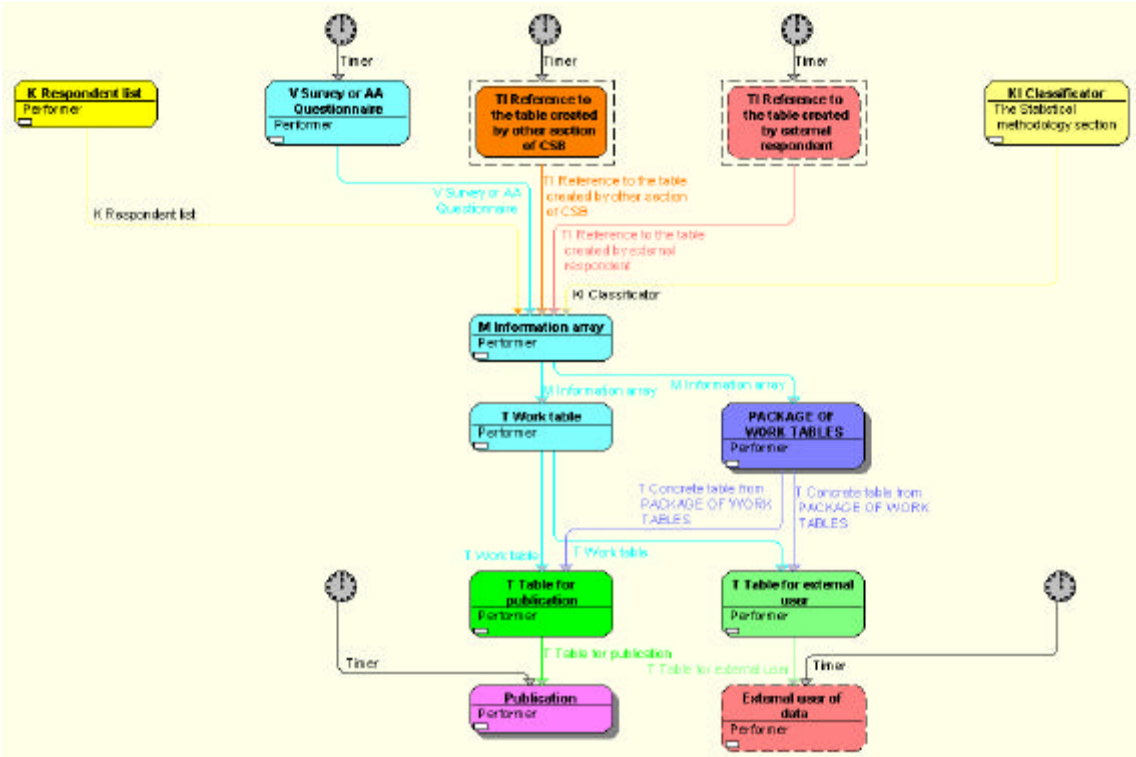

Figure 3. A simplified example of the BUSINESS PROCESS diagram

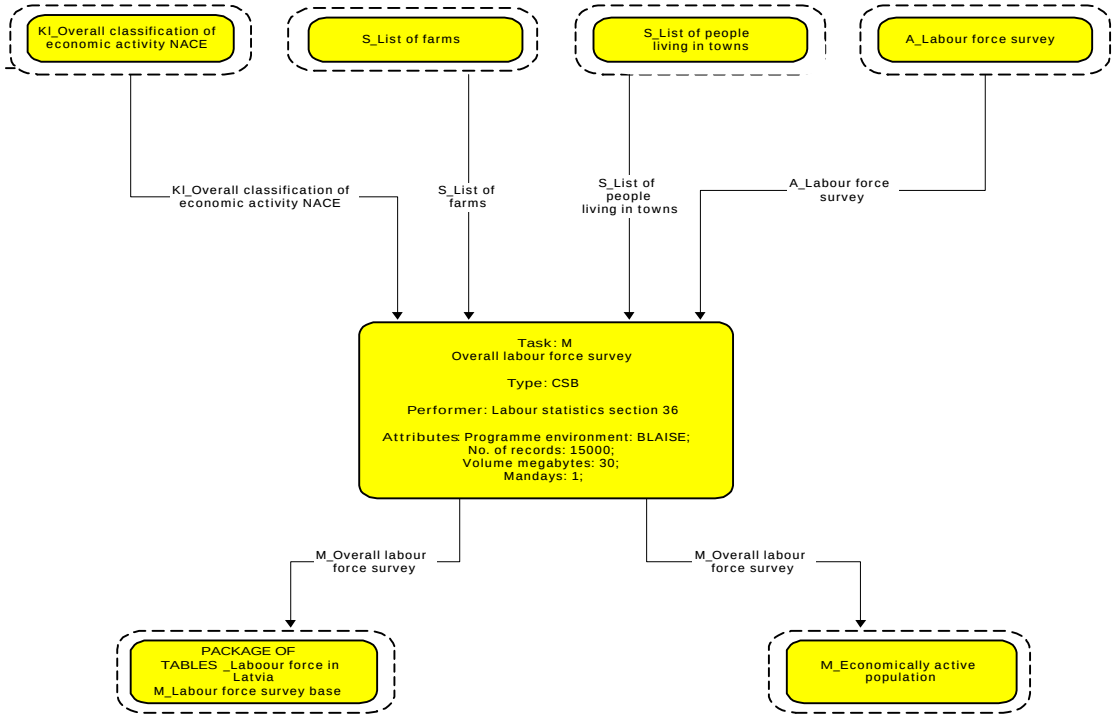

Figure 4. The TASK DETAILS diagram

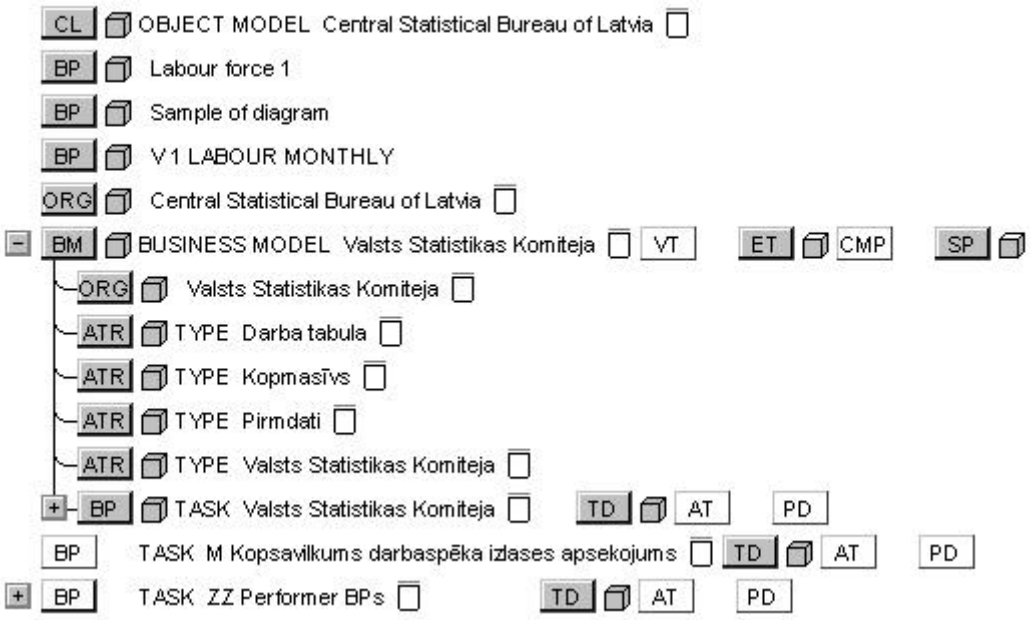

Figure 5. Fragment of Model tree of CSB

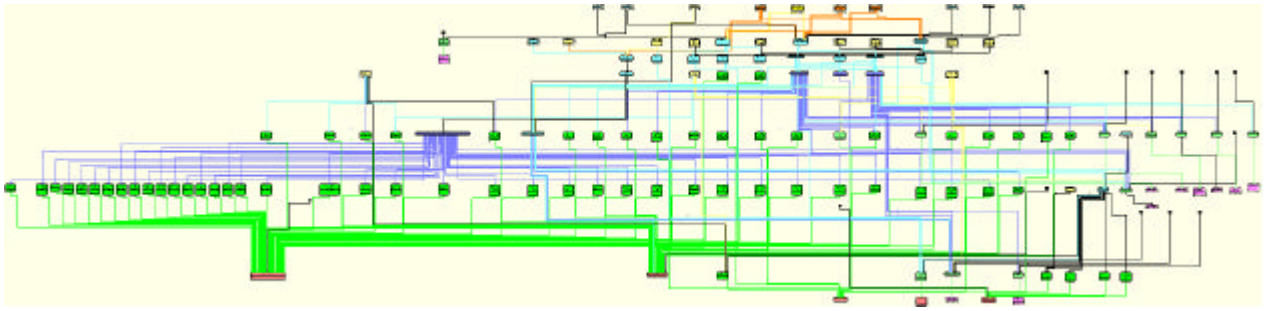

Figure 6. BUSINESS PROCESS diagram with vertical line layout

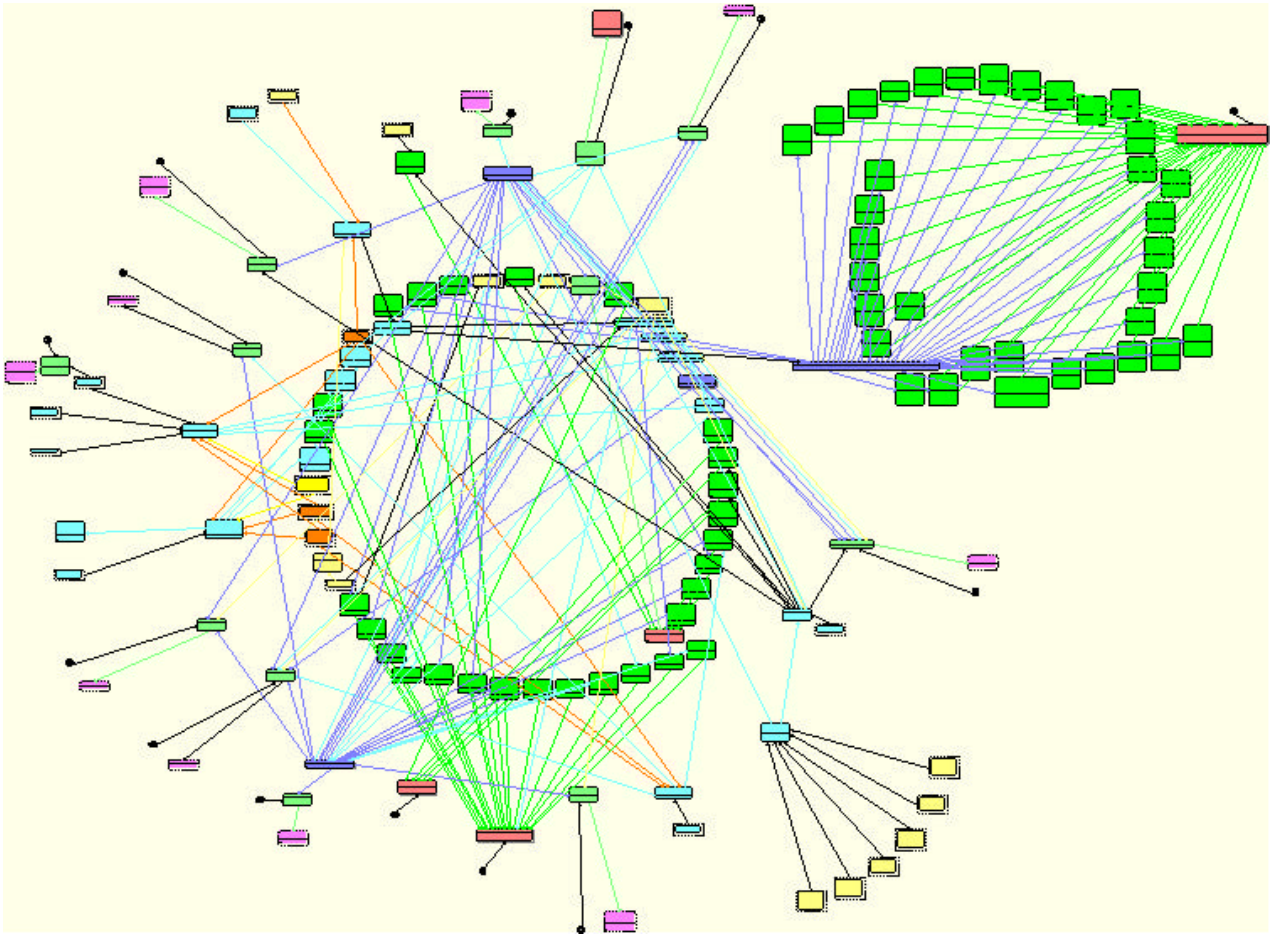

Figure 7. BUSINESS PROCESS diagram with Oblique lines layout

```
  DICTIONARY CSB_darba_0122
------------------------------------------------------------------------
FILTER SETTINGS
Scan: names
Scope: current subtree
Current diagram: BUSINESS PROCESS Darba statistika
Name pattern: *dati*
Show definitions: yes
------------------------------------------------------------------------
------------------------------------------------------------------------
Name type: *dati*          8 name(s)
------------------------------------------------------------------------
------------------------------------------------------------------------
T_ANO_36_dati      2 def(s)
Defined as  Task in TASK DETAILS T ANO 36 dati     21 Jan 98 10:58
Defined as  TASK DETAILS T ANO 36 dati     21 Jan 98 10:58
------------------------------------------------------------------------
T_EUROPEAN_PARLAMENT_36_dati      2 def(s)
Defined as  Task in TASK DETAILS T EUROPEAN PARLAMENT 36 dati     19 Dec 97
12:07
Defined as  TASK DETAILS T EUROPEAN PARLAMENT 36 dati     19 Dec 97 12:07
------------------------------------------------------------------------
T_EUROSTAT_36_dati      2 def(s)
Defined as  Task in TASK DETAILS T EUROSTAT 36 dati     19 Dec 97 12:07
Defined as  TASK DETAILS T EUROSTAT 36 dati     19 Dec 97 12:07
------------------------------------------------------------------------
T_OECD_36_dati      2 def(s)
Defined as  Task in TASK DETAILS T OECD 36 dati     21 Jan 98 10:58
Defined as  TASK DETAILS T OECD 36 dati     21 Jan 98 10:58

------------------------------------------------------------------------
------------------------------------------------------------------------
End of list
```

Figure 8. Dictionary filter example

Name type	Name	DEF/REF	Type	Diagram	Update time	Editor	Manager	Design status	Change level	Unique ID
dati	T_ANO_36_dati	REF	External task	BUSINESS PROCESS Darba statistika	28 Jul 98 12:02	User_ITgrupa		In Work	2	901312925+101944
dati	T_EUROPEAN_PARLAMENT_36_dati	REF	External task	BUSINESS PROCESS Darba statistika	28 Jul 98 12:02	User_ITgrupa		In Work	2	901312925+101944
dati	T_EUROSTAT_36_dati	REF	External task	BUSINESS PROCESS Darba statistika	28 Jul 98 12:02	User_ITgrupa		In Work	2	901312925+101944
dati	T_OECD_36_dati	REF	External task	BUSINESS PROCESS Darba statistika	28 Jul 98 12:02	User_ITgrupa		In Work	2	901312925+101944

Figure 9. Dictionary filter example