

**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)
(Luxembourg, 9-11 April 2008)

Topic (ii): Part A: Advocating for statistical metadata in a corporate context

STRATEGY FOR THE DEVELOPMENT OF A NEW INTEGRATED METADATA DRIVEN STATISTICAL INFORMATION SYSTEM IN THE STATISTICAL SERVICE OF CYPRUS

Submitted by the Statistical Service of Cyprus¹

I. EXECUTIVE SUMMARY / INTRODUCTION

1. The Statistical Service of Cyprus (CYSTAT) is facing increased number of users on both national and international level. Furthermore, great challenges for CYSTAT are changing users' needs and calls for better quality of statistical information. At the same time, however, CYSTAT is requested to increase efficiency of statistical production and to reduce burden on statistical respondents. Such development has a significant impact on the scope, contents and quality of observed statistical information and, consequently, on the methods, tools and techniques used for collection, processing and dissemination of statistical information.
2. In accordance with the CYSTAT strategic goals, a redesign of the existing information system was launched after the accession process to the EU had been successfully completed in 2004. A global architecture of an Integrated Statistical Information System in CYSTAT (CYSTAT-ISIS) was designed.
3. An integral part of CYSTAT-ISIS is the Statistical Metainformation System (SMS). Effectiveness of CYSTAT-ISIS depends directly on effectiveness of its SMS. The goal of SMS is to ensure internal and external integrity of CYSTAT-ISIS from both statistical content and technology point of view. Making a business case for CYSTAT-ISIS and SMS management was indispensable precondition for SMS development and implementation.

¹ Prepared by Costas K. Diamantides, cdiamantides@cystat.mof.gov.cy

II. CURRENT SITUATION

A. Statistical Service of Cyprus

4. The Statistical Service of Cyprus is the competent authority responsible for the compilation and the dissemination of most of the official statistical data in Cyprus. CYSTAT functions under the Ministry of Finance however the Statistical Service maintains its autonomy in technical matters and has exclusive responsibility for the choice of methodology, technique, definitions and procedures for the realization of the programmes of statistical activities, as well as for the publication of the statistical data produced.

5. CYSTAT is organised by subject matter on the basis of 7 main Divisions and 4 supporting Sections. In addition to the main office in Lefkosa, CYSTAT has also offices in Lemesos, Larnaka and Pafos. CYSTAT currently employs 145 permanent staff. A number of casual staff is also employed to assist in the various censuses and sample surveys, primarily at the data collection phase.

6. Over the last decade, the demand for the Cyprus's statistical data has grown exponentially. This is mainly due to three important factors. The first factor is the increased demand for data and metadata by EU institutions, mainly Eurostat and the European Central Bank, due to the fact that Cyprus joined the European Union in 2004 and adopted the Euro in 2008. The second factor concerns the increased usage of official statistics by several groups of users like journalists, students and enterprises. Finally, the third factor is the easy access to the official statistics that the Internet offers to users. Nowadays, users do not only ask for data but for metadata as well. Users are interested to know about methodologies, quality of data etc.

B. Statistical production

7. The current statistical production in CYSTAT faces several bottlenecks as follows:

- The technical infrastructure and the current organization of statistical production are not very efficient in meeting the growing demands of the users. Although in recent years, some new technologies have been implemented (e.g. BLAISE for the data capture of households surveys, palmtops for small scale surveys, website technology for dissemination, etc), the basis of the technical structure and organization of computerized work requires urgently upgrading. For example, VAX - introduced in CYSTAT in 1980 - is still used for the data entry of some surveys.
- Most of the IT work is carried out by the individual subject matter Divisions/Sections of CYSTAT. The support of IT Unit is not sufficient, due to the fact that this unit disposes two permanent staff members only. For the processing of statistical data, diverse application software packages are used. The data from statistical surveys are stored individually on the personal computers of the subject matter statisticians, responsible for their processing. It is difficult (or even not possible) to access such stored data by other users within CYSTAT.
- The subject matter statisticians have to spend, under the current organization of statistical production, a significant amount of time by performing technical work. The consequence is that, they are lacking time and capacity for the substantial, content oriented, statistical and methodological activities.
- The CYSTAT's website is the main dissemination tool. It is, therefore, imperative to introduce new web technologies in order to offer more data and metadata to users in a simple and user friendly way.
- A great challenge for CYSTAT is to identify methods for reducing the response burden of the enterprises. Extensive use of administrative sources can contribute to solve this problem significantly. Although the use of administrative sources has increased in CYSTAT during the last years, their use is still limited (mainly due to differences in timing and methodology).

C. Statistical metadata

8. Metadata on statistical data are presented in statistical questionnaires, tables, publications and on the website. No unified and /or formalized metadata system exists. Metadata are maintained manually in the individual subject- matter Divisions/Sections.
9. There is a lack of methodological coordination in metadata definition and description of statistical variables/indicators in CYSTAT. Furthermore, there is no sufficient methodological coordination in metadata definitions and description of data from administrative sources.
10. Statistical classifications are maintained in the responsibility of individual subject matter Divisions/Sections. There is no central administration of statistical classifications in CYSTAT. Classifications are kept in flat file form, without indicating links between individual classifications domains.
11. Sample surveys are prepared under the responsibility of the individual subject matter Divisions/Sections. There is limited central evidence on sampling methods for statistical production in CYSTAT.
12. In the time being, a coherent statistical metainformation system does not exist in CYSTAT. Its development is a key prerequisite for the innovation of CYSTAT - ISIS and its better functioning in future.

III. CYSTAT - ISIS

A. Strategic study

13. In order to improve the current situation, the top management of CYSTAT agreed that a study should be prepared for the Development of an Integrated Statistical Information System in CYSTAT. The study was completed and approved in April 2007.
14. The major goals of the study were the following:
 - To identify problems and shortcomings of the current information system of CYSTAT by analysing the current situation;
 - To describe the proposed integrated statistical information system by first developing Business System Options (BSO) for SMS, data collection, data storage and processing, data dissemination, registers and data confidentiality;
 - To describe the technical solution by first developing Technical System options;
 - To develop a management strategy and organization for the CYSTAT – ISIS;
 - To specify requirements for human resources;
 - To estimate costs for the CYSTAT – ISIS development and implementation.

B. Expected benefits

15. CYSTAT – ISIS is expected to achieve the following benefits:

i. Modernization

16. The proliferation of statistics, produced by CYSTAT, calls for efficient methods and techniques for storage and maintenance of statistical data. Advantages of data warehouse technologies meet CYSTAT requirements. Uniformity of data storage and related metadata description, easy and quick access to data, powerful possibilities of data export towards users and a potential for data protection are examples of such important data warehouse features.

ii. Statistical Metainformation System

17. The corporate SMS will contribute to upgrade methodological work in CYSTAT, to increase efficiency of statistical activities and to enhance quality of statistical data/information.

18. The SMS will service and drive statistical production in all stages of a statistical production life cycle (from the design of statistical tasks, definition of input and output variables/indicators, questionnaires, validation controls, statistical methods, to the dissemination of the results, evaluation of quality of statistics produced and management of the Registers maintained in CYSTAT).

iii. Improved service to users

19. CYSTAT-ISIS will assign subject matter statisticians and methodologists more time for the professional statistical work. It will contribute to increased quality of statistical final products and professional services to statistical users.

20. In addition, there will be a greater transparency in the methods used to produce statistical data, since users will have access to all relevant metadata.

21. Undoubtedly, the increased availability of statistical data and metadata will benefit the government departments, agencies and educational and scientific institutions, as they are the main users of the statistics in Cyprus.

iv. Increased quality of statistical information

22. CYSTAT-ISIS will focus on improving quality of statistical information. The SMS will provide key tools for this task. In this respect, namely the following should be highlighted:

- SMS will maintain and update qualitative characteristics and methods for statistical data assessment,
- permanently updated SMS central database will ensure correct interpretation of statistical data and processes,
- metainformation, maintained on statistical processes, will contribute to the evaluation of efficiency of statistical activities, conducted in the framework of those processes.

v. Easier access to data and metadata

23. Nowadays the main dissemination channel of statistical information of CYSTAT is the website. Although the site contains a large volume of data, the format of available data (files in Excel format) is not

satisfactory, as it limits the volume of information and, in addition, the process of preparation and uploading of files is time consuming.

24. Online output databases, proposed in CYSTAT-ISIS, will enable users to search and download, via Internet, all requested statistical data and use them for further processing in any format. In this way, there will be no restriction on the amount of information that can be provided.

vi. Broader use of administrative sources, reduction of respondents' burden

25. Broader use of administrative sources, aims to lighten the respondents' burden. To solve this problem is a methodological issue, dealing with definition of administrative data content, benchmarking of administration data quality, timing and purposes for administrative and statistical data production. The CYSTAT-ISIS aims to assist solving this problem. Use of administrative data remains to be a great challenge for CYSTAT. This task should be explored and solved jointly with governmental and other bodies, responsible for administrative data. It is a long term goal but, its solution will have a significant impact on the cost efficiency of the statistical production process.

26. SMS will play an important coordinating and integrating role in all above mentioned CYSTAT-ISIS activities. Bearing in mind, however, a diversity of statistical respondents, the goal "to diminish respondents' burden" will remain a longstanding endeavour for CYSTAT. It calls for efficient planning and systematic monitoring of those activities.

C. Main components

27. The main components of CYSTAT-ISIS are the following:

- Content and methodology
- Statistical metainformation system
- Unified statistical production process and upgraded dissemination

i. Content and methodology

28. At present, there is no specific body for statistical methodology in CYSTAT. Individual subject matter Divisions are responsible for the development and implementation of statistical methodology.

29. Clearly, CYSTAT-ISIS requires a central monitoring of statistical methodology issues, related to the statistical production (e.g. preparation of statistical surveys and samples, definition of statistical variables etc). It would ensure the development and use of unique methodology for individual content components related to statistical production, their central monitoring, storage, administration and use. SMS will provide necessary tools for those activities.

ii. Statistical metainformation system

30. The SMS is a core part of the CYSTAT-ISIS. The SMS will ensure a systematic use of metainformation in CYSTAT and drive statistical production process. The SMS strategy must be an integral part of the CYSTAT strategy. The proposed SMS strategy is focused on the statistical production process. To develop SMS is a content methodological task dealing with definition, modelling and formalized description of statistical objects (classifications, variables/ indicators, samples, surveys etc). Necessary precondition for SMS functioning is a centrally organized administration.

31. A model for metadata definition and description of statistical variables will be applicable in all stages of the statistical life cycle.

32. The development and implementation of SMS will require a special effort from subject matter statisticians and methodologists (namely in its initiation stage). The use of SMS will contribute significantly to increased quality of statistics. It will be a core component of central methodological activities in CYSTAT.

33. With regards to the users' requirements, the SMS should ensure namely the following functions:

- To be a source of information for quality evaluation. Based on the EU recommendations, national statistical organizations have adopted important criteria for quality benchmarking (e.g. comparability of statistical concepts, accuracy of statistical estimates, timeliness of provided information, accessibility and clarity of information, etc.). For each selected criterion, a corresponding set of metadata has to be created that will enable the quality of the provided information to be evaluated;
- To be a key tool for providing statistical information to users. Users of statistical information need high-quality and reliable metadata for searching, accessing, interpreting and further processing of statistical information.
- To provide tools for the integration of CYSTAT-ISIS with other governmental information systems. Growing utilization of administrative sources for statistical purposes becomes imminent. To enable data sharing, the integration and sharing of metadata between CYSTAT and other governmental bodies is indispensable.
- To provide tools for the integration of CYSTAT-ISIS with the information systems of international organizations. Key international organizations (Eurostat, ECB, IMF and others) impose increasing demands on metadata from national statistical organizations. The aim is to make the flow of statistical information more effective, and to ensure its comparability and unified interpretation at both national and international level.
- To provide tools for unification and standardization of data collection, processing and data dissemination and thus be more effective.
- To provide tools for effective management of statistical data processing.
- To provide tools for better monitoring of the CYSTAT's performance, namely via monitoring users satisfaction. To contribute to a better evaluation of the costs and benefits of the CYSTAT-ISIS.

iii. Unified production processes and upgraded dissemination

34. Current production of statistical data in CYSTAT is decentralized in the subject matter Divisions and/or in the IT Unit. Different software is used and tailor made applications are prepared for production of individual surveys. There is no central statistical data repository in CYSTAT. Statistical data from various statistical surveys should be more efficiently shared and made easily available for analytical purposes.

35. The implementation of CYSTAT-ISIS will radically change the processes of collecting, processing, storage and dissemination of statistical data.

36. The CYSTAT-ISIS will operate with a single data warehouse, interlinked with SMS database. Data warehouse will ensure an easy access to data and data export for analytical purposes. Furthermore, it is envisaged to unify and standardize application software for statistical production wherever possible (statistical questionnaires, data validation, data correction, data aggregation, creation of statistical tables, publications, etc).

37. CYSTAT has a long history in dissemination and, especially in dissemination via Internet. The website has become the major tool for dissemination. However, the dissemination activities should be continuously improved and conform to the changing needs of both users and technology.

38. SMS and data warehouse are the core components of CYSTAT-ISIS. All other operations should take into consideration rules, principles, functionality and capabilities of those two components. This is the reason why the CYSTAT-ISIS development strategy envisages a parallel implementation of those two components

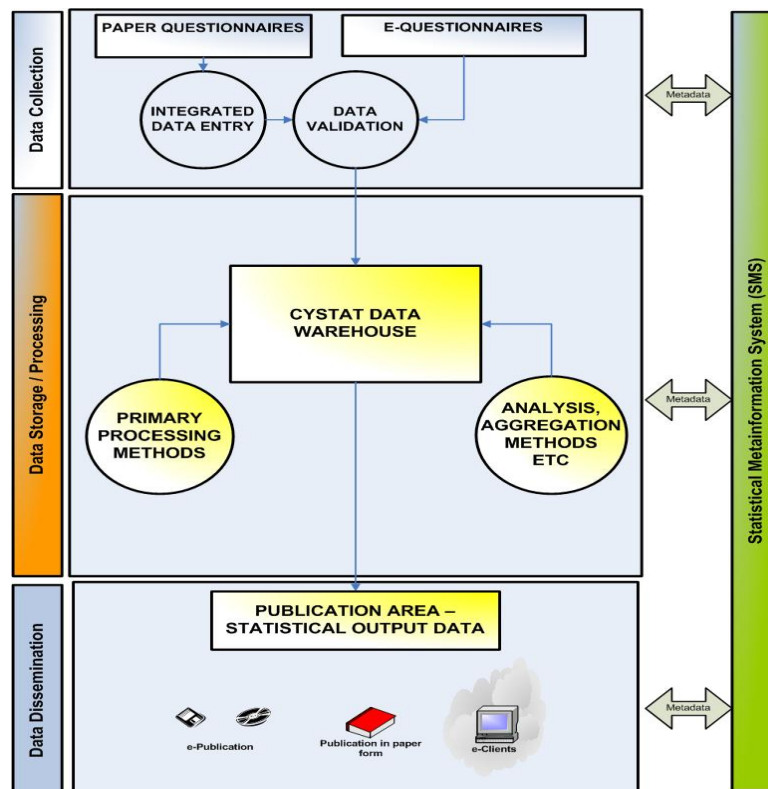
39. Based on the CYSTAT's needs, however, some statistical application software (e.g. Blaise for VAX replacement) has to be implemented prior to the CYSTAT-ISIS implementation. Such statistical application software should be integrated later on with SMS and data warehouse as well.

40. The statistical production process in CYSTAT is composed mainly of the following five phases:

- Definition of statistical tasks, surveys etc;
- Collection of statistical data: methods/techniques used for data collection, data entry and data validation;
- Primary Processing of statistical data: methods/techniques used for data imputation, non-response, computation, blowing-up etc;
- Analysis, aggregation and other methods/techniques for preparing output;
- Dissemination of statistical data: methods/ techniques used for the dissemination of statistical data to the diverse groups of statistical users;

41. All above mentioned phases, encompassing the use of international statistical standards (e.g. SDMX), will create a solid basis for smooth adaptation of CYSTAT-ISIS to users requirements and international data sharing needs (namely with EU).

42. The scheme below demonstrates a CYSTAT-ISIS global technical architecture:



IV. PLANS, MANAGEMENT AND ORGANIZATION

A. Implementation plan

43. The CYPSTAT top management identified priorities for the CYPSTAT-ISIS development and implementation. The implementation plan is based on those priorities. According to the plan there are four implementation phases as follows:

- Phase 0: Preparatory
- Phase 1: Analysis and Design
- Phase 2: Implementation
- Phase 3: Migration

44. Phase 0 contains activities that should be completed before the start of Phase 1 as their output is required for the Analysis and Design Phase. These activities include the analysis of the current status of statistical classifications, the analysis of the current processes needed for the execution of statistical tasks and the identification of duplicates in variables/indicators. Furthermore Phase 0 contains activities that are not directly linked to the implementation Phase 1 such as, the use of Blaise for data entry instead of the VAX and the development of a GIS system for the needs of the 2011 Population Census. Based on the schedule, Phase 0 is expected to commence in 2008 and be completed 2010.

45. Phase 1 is very important as the aim is to define WHAT is required and expected. Phase 1 includes the definition of the functional specifications and design of the first 4 SMS subsystems (statistical tasks, statistical surveys, statistical classifications, statistical variables/ indicators, the technical specifications and design of CYPSTAT-ISIS, the specification of CYPSTAT-ISIS hardware equipment and the functional and technical specifications for statistical application needed for CYPSTAT-ISIS. The launch of Phase 1 is scheduled for 2010 and its duration will be for 1 year.

46. During Phase 2 the CYPSTAT-ISIS hardware will be installed and the development and testing of the system, based on the specifications specified in Phase 1, will be carried out. In Phase 3 the current system will migrate to the CYPSTAT-ISIS statistical production. The implementation is expected to be completed in 2014.

B. Management and organization

47. The implementation of CYPSTAT-ISIS requires the restructuring of CYPSTAT. The current organization of CYPSTAT reflects decentralized nature of statistical activities in subject matter domains, methodology and technology of statistical production.

48. The CYPSTAT-ISIS implementation should comply with the following major new requirements:

- Central monitoring of content oriented methodological tools;
- Management of SMS development and implementation;
- New role of IT Section in the statistical production process, as a consequence of increased responsibilities and workloads requested by the proposed CYPSTAT-ISIS strategy (principle to unify and standardize methods and techniques of statistical production).

i. Central monitoring of statistical methodology

49. CYSTAT-ISIS requires a central organization of content oriented methodological tools. Such organization should ensure development and use of unique methodology for individual content components related to statistical production, their central monitoring, storage, administration and use.

50. To cope with this requirement, it is recommended to install a central “ISIS Methodological Unit” (ISIS-MU) in CYSTAT, directly supervised by the CYSTAT’s Director.

51. The ISIS-MU should take over, step-by- step, a conceptual and coordination role of CYSTAT-ISIS. Its functions and activities, however, will differ in individual phases of the CYSTAT-ISIS project implementation life cycle. The core framework and tool for ISIS-MU activities will be, after its accomplishment, the SMS.

52. The ISIS-MU members should have expertise in statistical methodology with ability for system designing and modelling. General knowledge of IT statistical applications is highly desirable. The ISIS-MU should cooperate closely with subject matter statisticians from CYSTAT’s Divisions/Sections and IT experts from both CYSTAT’s IT Section and external software producers.

ii. SMS organization

53. A basic characteristic of the SMS project is its cross-cutting nature. The SMS is important for the following CYSTAT-ISIS activities: management, methodology, activities of the subject matter Divisions/Sections, dissemination of statistical information and statistical data processing.

54. As it is linked with virtually all of the statistical activities in the CYSTAT, it will require a multi-professional structure, both on the side of the project’s management as well as on the side of experts working in the project teams. An interdisciplinary team approach must be the main way how to organize the SMS project.

55. The SMS organization structure includes the Managing Committee (MC), the SMS project team and the project teams for the individual SMS blocks.

56. The teams should encompass experts from the following fields: methodology, the subject matter Divisions/Sections, IT, and users of statistical information and the SMS. The composition of the teams should be flexible and will depend on the specific issue being solved.

57. A number of project teams have to be appointed by the CYSTAT top management. Similarly, all the results of the work of the teams have to be approved by the CYSTAT top management.

iii. New role of IT Section

58. The proposed organization of IT activities aims to contribute increasing quality of subject matter statistical work via decreasing workloads linked to the purely technological aspects of statistical production.

59. The proposed organization of statistical production is based on the principle to unify statistical operations, appearing in the production of all statistical tasks (validation controls, corrections, aggregations, production of final outputs etc). Standard software should be developed for those operations. The IT Section will carry out such standard operations, for every statistical task, centrally.

60. The subject matter statisticians will perform statistical analysis, via downloading the data from the central data warehouse and using analytical statistical software packages.

61. The proposed organization requires substantial changes in the role and resources of the IT Section.

V. CONCLUSIONS

62. The strategy for the development of an Integrated Statistical Information System in the Statistical Service of Cyprus is an ambitious and long term project. In order to implement the strategy successfully it is important to have the full support of the top management of CYSTAT as well as the commitment of the personnel. In addition, since in the implementation of CYSTAT-ISIS there is another government department involved (the Department of Information Technology Services (DITS) is responsible to provide high quality Information Technology Systems and Service to the Government), a close and productive cooperation between the two departments is essential for the successful completion of the project. It is important to stress here that DITS's involvement is significant since in CYSTAT there is currently no technical knowledge and expertise.

63. The expected benefits of CYSTAT-ISIS, mentioned earlier in this paper, mainly concern the improvement of quality of data and metadata processed and analysed within CYSTAT and as a consequence offered to users. The strategy does not provide any cost benefit analysis, i.e. there is no comparison between the current and the proposed new system as regards savings in costs. It is very clear that although CYSTAT has responded very well in the previous challenges (e.g. accession to EU, adoption of Euro) however, the current information system is not sufficient to meet the new challenges.

64. A significant aspect of the strategy is the big bang approach meaning that a new platform will be built at once and it will not be a step by step approach. This approach was selected because there is currently no common platform within CYSTAT. In order to avoid possible future technological conflicts and in addition, due to the significance of metadata in the statistical production, it has been decided to implement both pillars of the strategy, namely data warehouse and SMS, at the same time.