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

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

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

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (i): Web technology in statistical information systems

WEB-SUPPORTED STATISTICAL DISSEMINATION PROCESS SERVING STATISTICAL DATA USERS

Supporting paper

Submitted by Statistical Office of the Republic of Slovenia[1]

I.   INTRODUCTION

1.   This paper presents two aspects of using web technology in statistical information systems: web technology in the statistical process and web technology in dissemination of statistics. Both uses will be presented on practical cases, developed in the Statistical Office of the Republic of Slovenia.

2.   The corporate (statistical) data warehouse is used in SORS as a target common data management concept. This concept is very closely connected with the presentation of the statistical data and metadata on the web. We have developed some web-supported solutions both for the statistical process and for dissemination.

II.   COMMON DATA MANAGEMENT ARCHITECTURE IN SORS

A.   Main elements of the architecture

3.   The Statistical Office of the Republic of Slovenia has developed main strategic objectives in the field of data dissemination: equal access to results of statistical surveys, professionalism and transparency, user friendliness, especially availability of the statistical information on the Internet. In order to achieve and keep objectives, several projects have been developed in the past five years.

4.   In the project called Klasje, a classification database was built and put into the production at the beginning of 2000. In the next very large project StatCOP98, a corporate data warehouse concept of data management was introduced. Several dimensionally modelled internal databases and a first release of the

---

[1] Prepared by Matjaz Jug (matjaz.jug@gov.si).

central metadata repository were built. The project was completed at the end of 2001. In 2002 the STAT2000 project introduced a new dissemination process to centralize and modernize the preparation of final statistics in SORS. Further development was made on the central metadata repository. Finally, in 2003 SORS rebuilt and modernized its internet site.

5.   The new SORS's web page is closely related to metadata systems and statistical data warehouse developed in the mentioned projects. The classification database KLASJE contains all classifications used in SORS and provides automatic extraction of classification related metadata (categories, levels, hierarchies, etc.) which are used in codelist tables, star scheme dimension tables and cube dimensions of the statistical data warehouse. The classification database was built on the basis of New Zealand classification model (CARS) and has been in production since November 2000.

6.   The meta database METIS contains structured metadata about surveys, questionnaires, questions, observation objects, variables, their valuesets and values, publications, activities, etc. The database is finished and is in the process of deployment. Some parts of METIS (for example advance release calendar) are in production.

7.   In the STAT2000 project a new dissemination database, based on Swedish pseudo-OLAP file format PC-AXIS, was built. The database has marketing name SI-STAT on SORS's Internet site and enables users searching for published data, selecting the variables and variable values, pivoting the table layout and downloading data in various formats. For the organization of statistical process in SORS, it is very important that PC-AXIS has been selected as an internal dissemination standard. Multidimensional PC-AXIS matrixes have reference role and are used as a source for different dissemination channels like SI-STAT web database, publications, data shooting, etc. The general scheme of the new dissemination process in SORS is presented in Picture 1.
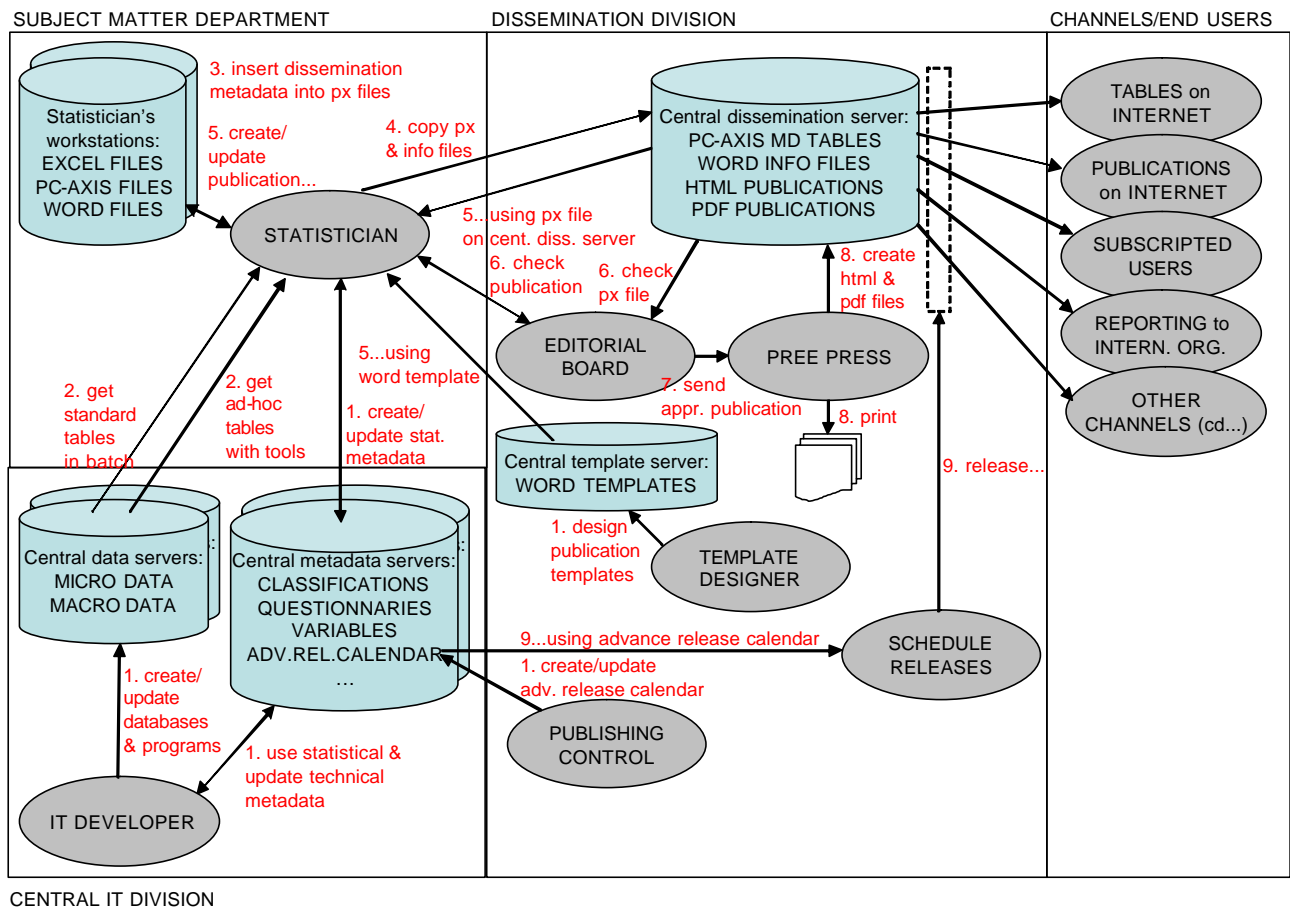
III.   WEB TECHNOLOGY IN THE STATISTICAL PROCESS

A.   Statistical process from tabulation to the publication on the web

8.   Data extracted from the statistical data warehouse and other sources are aggregated into final multidimensional matrixes, stored in PC-AXIS format. PC-AXIS file format includes many important types of metadata, for example administrative metadata (refresh date, matrix name, etc.), methodological metadata (footnotes, indicators, etc.). The preparation of the dissemination matrix should be supported with a powerful user interface, so we are using standalone application PX-Edit, developed in Statistics Finland, with broad functionalities for importing, manipulating and maintaining the multidimensional matrixes (cubes).

B.   Web form as a tool for submitting final statistical data

9.   Finished matrixes have to be submitted by the responsible subject-matter staff on the central dissemination server into the corresponding folder. The web form used for this purpose offers some very useful possibilities. Firstly, upload of the matrix using web form is easy and fool-proof because of predetermined fields, checked by the metadata already in the database. Some administrative metadata, which are not included in the PC-AXIS format, (for example the list of related printed publications) could be entered through the form. Files with multidimensional data and related metadata are placed in the corresponding directory and important metadata are extracted from the PC-AXIS file and placed in the relational database together with metadata, which were entered directly on the form.

Picture 1: STAT2000 Scheme of the dissemination process

SUBJECT MATTER DEPARTMENT   DISSEMINATION DIVISION   CHANNELS/END USERS

Statistician's workstations:
EXCEL FILES
PC-AXIS FILES
WORD FILES

3. insert dissemination metadata into px files

5. create/ update publication...

4. copy px & info files

STATISTICIAN

Central dissemination server:
PC-AXIS MD TABLES
WORD INFO FILES
HTML PUBLICATIONS
PDF PUBLICATIONS

TABLES on INTERNET

PUBLICATIONS on INTERNET

SUBSCRIBED USERS

REPORTING to INTERN. ORG.

OTHER CHANNELS (cd...)

5.. using px file on cent. diss. server
6. check publication

6. check px file

8. create html & pdf files

EDITORIAL BOARD

PREE PRESS

7. send appr. publication

8. print

9. release...

5...using word template

1. create/ update stat. metadata

2. get standard tables in batch

2. get ad-hoc tables with tools

Central data servers:
MICRO DATA
MACRO DATA

Central metadata servers:
CLASSIFICATIONS
QUESTIONNARIES
VARIABLES
ADV.REL.CALENDAR
...

Central template server:
WORD TEMPLATES

1. design publication templates

TEMPLATE DESIGNER

SCHEDULE RELEASES

9...using advance release calendar
1. create/update adv. release calendar

1. create/ update databases & programs

1. use statistical & update technical metadata

IT DEVELOPER

PUBLISHING CONTROL

CENTRAL IT DIVISION

10. Using web forms in the statistical process has two main advantages: forms are accessible for all staff and the deployment of new versions is easy. In our case the same form is used for both uploading the dissemination materials to the dissemination server and filling the meta database with metadata. Using this approach we try to follow golden rules of the development and maintenance of metadata systems, particularly to capture metadata at their natural sources and to make metadata-related work an integrated part of the business process of the organisation (Sundgren, 2003).

IV. WEB TECHNOLOGY IN DISSEMINATION

A. Presentation of different types of information on the web

11. SORS's Internet site (http://www.stat.si) offers different types of presentation of statistical data (presentation channels). First releases are presented as news on the main page with the link to full version with more detailed figures, charts and comments. Time series for main figures (topical indicators) are presented in a very simple but powerful user interface with the short methodological comment, with the possibility to export the figures in Excel or web query and do the dynamic recalculation between different time periods. Printed publications (Rapid Reports, Results of Surveys, etc.) are presented mostly using PDF and static tables easy to download and print. More detailed data are presented as multidimensional matrixes (cubes) with the possibilities to filter the data according to variables and their values, pivot the table structure and download the resulted table in various file formats. Some advanced charting possibilities are

possible and the presentation of data in maps, there are direct links to the documents with the methodological explanations and some additional formalized metadata.

12.   There are also different types of presenting methodology and metadata. The Guide Through Statistics offers a glossary of statistical terminology in a very descriptive and user friendly way. Methodology explanation documents are linked to database matrixes. Classifications from the classification database Klasje are presented trough a web application that offers searching, navigating and exploring classification and downloading codelists. The new website offers new services such as periodical subscription to news, publications and data matrixes, direct web queries, etc.

B.   Putting all together: linked dissemination services

13.   A lot of statistical information in combination with a lot of different types of presentation should be carefully organized on the web to prevent users from getting lost searching for wanted information. There are different levels of organizing and linking different website contents:
   a) General navigation of the SORS's site is the same regardless of different presentation types and is arranged in 28 statistical fields joined in 4 more general topics. Users can find the content in a similar way in the publications, databases, etc.
   b) Different presentation types of statistical data from the same statistical fields are linked directly. Users can, for example, jump from the publications page to the database and see the contents for the same statistical field.
   c) Inside the statistical fields data are further organized in statistical products. The same statistical product could be presented as news (first release of data), publication and one or a couple of matrixes in a database. However, links between the same products in the different presentation channels are not as easy as those for statistical fields because not all products are presented through all presentation types. On the other hand, direct jumps to the same products can save users a lot of unnecessary clicking.
   d) "Same product" links are useful, but they can still return more than one result because of a different data organization in different presentation channels. While many data periods are assembled together in the same database matrix, there is only one data period in the first release. One publication, of course, contains more than one table and a jump in the database returns multiple results. In some cases "smart direct shot" returning only one corresponding result is possible. Such an example is links to documents with the methodological explanations.

14.   Closely integrated websites with smart links to related topics and contents are state of the art today. Here is the point where different phases from the statistical process find its expression: metadata collected using standalone applications and web forms serve as a source for automatically managed links on the interconnected website. Links should be exactly where the users need them to navigate effectively between different presentation types of data and different types of metadata.

15.   Within the SORS's metadata systems and dissemination process, a great effort was made to enable connectivity between content in different dissemination channels. The classification of 28 statistical fields is built in metadata applications and a classification database to enable the same navigation and links on the website. The statistical product code was introduced in the STAT2000 project and this code is contained in short names of publications and matrixes. In a PC-AXIS upload form, the list of related publications is entered by subject–matter staff so there is no need to maintain the links later manually. PC-AXIS format holds the name of the document with the methodological explanations.

16.   The current connectivity on the website is presented in the table below. Most presentation channels have the same navigation structure (a). The exception is the Guide Through Statistics, which is organized as a glossary of statistical terminology. In a general navigation only a limited number of most frequently searched terms are listed. The "Same statistical field" links (b) are offered on the "news" pages, we are planning to add them to the publications and database pages also. The "Same product" links (c) will be

possible between publications and database. The direct links (d) are for now only established from the database to methodology.

From/To	News	Time series	Publications	Database	Guide	Methodology	Classifications
News		a, b	a, b	a, b	b	a, b	a, b
Time series	a, b		a	a		a	a
Publications	a	a		a, (b), (c)		a	a
Database	a, b	a	a, b, (c)			a, d	a
Guide	a, b	a	a	a		a	a
Methodology	a	a	a	a			a
Classifications	a, b	a	a	a			

F.   Technical aspects

17   Technologies, used for web forms and the internet site, are listed in the following table:

Category	Description
Web server	Microsoft Server 2003, Microsoft IIS
Web RDBMS	Microsoft SQL Server 2000 Enterprise Edition
Web Applications	ASP, ASP .NET
Web Database	PX-Web, PC-AXIS file format

V.   CONCLUSION

18.   What are the future plans? Web services are a very good basis for building a well-interconnected data service both internally and over the Internet. SORS is planning to develop further the dissemination process and to integrate different contents on the website.

References

Križman I., 2002: The Slovenian National Statistics, Trustworthy and User-oriented. Statistical Office of the Republic of Slovenia, Ljubljana.

Sundgren B., 2003: Developing and Implementing Statistical Metadata Systems, MetaNet WG3 Deliverable

Jug M., Šnuderl K., 2003: Data Tables on SORS Website – Changes in the Method of Preparing and Using Statistical Data, 13th Statistical Days Radenci, 24.-26. November 2003

Arnic A., Jug M., Klep J., Kutin J., Seremet T., 2003: Metadata production systems within Europe: The Case of Statistical Office of the Republic of Slovenia, Metadata Production and Exchange Workshop, Luxembourg, 3-4 April 2003

STAT2000 Final Report, Annex 5: Description of Dissemination Database and Process, Phare project "Alignment of Slovene Statistics with the Acquis Communautaire", June 2003

- - - - -