

Distr.
GENERAL

CES/AC.71/2005/27 (Summary)
4 February 2005

Original: ENGLISH

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

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)

ORGANIZATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Bratislava, Slovakia, 18-20 April 2005)

Topic (iii): XML and web services

eSTATISTIK.CORE - an XML based approach to data collection from ERP systems

Supporting Paper

Submitted by the Federal Statistical Office, Germany¹

Summary

I. INTRODUCTION

1. In cooperation with renowned software producers and respondents, the German statistical offices have developed the internet-based Common On-Line Raw Data Entry, eSTATISTIK.CORE. Its principal objective is to facilitate the use of business data management systems - such as ERP systems (enterprise resource planning) - of businesses and organizations for generating and transmitting raw data messages. eSTATISTIK.CORE consists of several infrastructure and software components: a single point of delivery on the web, standardized cross-survey XML-based document formats, electronic survey descriptions and free software, namely the software library *CORE.connect* and the stand-alone application *CORE.reporter*. It is envisioned that eSTATISTIK.CORE will lead to "push-button" solutions where so-called *statistics modules* provided by software producers automatically compile raw data, and generate and transmit messages.

II. eSTATISTIK.CORE

2. eSTATISTIK.CORE relies on three XML-based document types. *DatML/RAW* is used for sending raw data messages. A *DatML/RAW* document may contain any number of messages, which may differ from each other in survey, reference period, respondent, the collecting office, and so on. *DatML/RES* is a response message format complementary to *DatML/RAW*. It is used for storing protocols of the tests that are performed on ingoing *DatML/RAW* documents. Respondents can access and download these protocols and use them for internal purposes. Finally, *DatML/SDF* is the format in which machine-readable survey descriptions are supplied on the web for controlling statistics modules and *CORE.reporter*.

¹ Prepared by Michael Schäfer, michael.schaefer@destatis.de.

3. The library CORE.connect implements the basic functionalities for generating and transmitting DatML/RAW documents and for downloading DatML/RES protocols and DatML/SDF survey descriptions. The stand-alone application CORE.reporter is built on CORE.connect and addresses the needs of smaller businesses, enabling users to map data sets onto survey data models.
4. Respondents use either statistics modules (as part of their business data management system) or CORE.reporter to generate raw data messages and to transmit them to a single point of delivery on the web. The associated DatML/SDF survey description file controls basic client-side plausibility checks and helps keep track of changes in surveys.
5. On the data collection server, ingoing DatML/RAW documents are checked for validity and then pass a distribution layer that dissects them into single reports with definite contexts, i.e. collecting office, survey, reference period, and so on. The distribution layer may be two-phased, with a distribution by collecting office in phase one and a distribution by survey in phase two, and can itself be distributed over the central data collection server and servers of individual offices. Valid raw data messages are, if necessary, automatically transformed into the specific raw data interface of the respective survey.
6. The metadata that controls eSTATISTIK.CORE is generated by tools that subject-matter specialists use in their routine work, particularly for creating specifications of variables, data edits, data sets and tabulations. These tools have been developed over the last years, with the explicit objective of making metadata available in machine-readable formats.
7. An evaluation version of eSTATISTIK.CORE is available since early 2004. In the first quarter of 2005, a production version will be released and used for regular application in selected wage statistics.

III. CONCLUSION

8. Standardized, cross-survey document formats and the availability of definition metadata make generic, efficient and secure procedures for collecting raw data from business data management systems feasible. However, any such solution would fail to succeed without support from the targeted user community.
