Distr.
GENERAL

ECE/CES/SEM.54/13 (Summary)
3 April 2006

Original: ENGLISH
ENGLISH AND RUSSIAN ONLY

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

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD Seminar on the Management of Statistical Information Systems (MSIS)
Sofia, Bulgaria, 21-23 June 2006

Topic (ii): Dissemination and client relations

A STRATEGY FOR CONTROLLED AND SECURE ACCESS TO MICRODATA AT STATISTICS NETHERLANDS

Invited Paper prepared by Olav ten Bosch and Frans Hoeve, Statistics Netherlands

Summary

1. This paper describes the strategy developed at Statistics Netherlands to offer trusted researchers from Dutch universities and (governmental) research institutes access to well-documented microdata. It highlights the general policies for microdata access at Statistics Netherlands and it describes the actual microdata services being developed both from an organizational as well as from a technical point of view. It describes some of the infrastructural, organizational and security issues that come about when developing such a microdata dissemination strategy, the resulting microdata services that have been realized so far and the directions for extending these services in future.

2. In addition to the aggregated statistics produced by Statistics Netherlands, disseminated via its main website and via other publications, there is an ever increasing demand for dedicated analyses to be performed. These analyses cannot be answered from regular statistics and require access to the underlying microdata. The increasing interest in microdata is also due to the fact that the focus for data collection at Statistics Netherlands shifted from carrying out surveys to connecting to registrations, which further extended the opportunities to perform advanced research on data collections combined from different sources. Although for reasons of privacy Statistics Netherlands cannot disseminate microdata

directly, it does offer research institutes other ways to access microdata. First of all, it offers *on site* facilities for analysing microdata within the premises of Statistics Netherlands. Second, researchers may send in scripts to be executed on well-defined sets of microdata. This is called *remote execution*. Finally, from mid-2005 a *remote access* facility is in pilot phase, which makes it possible for researchers to analyse microdata through a secure connection from workstations in their own institute. In all cases, researchers may bring in data from their own institute to be combined with data from Statistics Netherlands. This combined data set may only be used for their own specific research.

3. Recently, all services for microdata research, both for access to business statistics as well as social statistics, have been concentrated into one organizational unit, the *Centre for Policy Related Statistics*. As part of this process, Statistics Netherlands developed an overall strategy for offering microdata services. An important aspect of this strategy is that external researchers are allowed to use Statistics Netherlands microdata only under secure and confidential conditions. Furthermore, the services are open tp researchers of trusted institutes only, as specified by Dutch law or of institutes that have special permission to access microdata by approval of the Central Commission for Statistic s of Statistics Netherlands.

4. Using either one of the microdata facilities, researchers may work with statistical tools such as Spss, Stata, Gauss and Ox on microdata that were documented especially for use by external researchers. This documentation standard was developed specifically to annotate microdata from social statistics, but it is currently being generalized to document microdata on business statistics as well. In addition, international metadata standards such as DDI and SDMX are taken into account when refining this standard in future.

5. (Intermediate) research results are checked for confidentiality. Unfortunately, this is a labour intensive task, which cannot be automated easily, since it requires knowledge of the research field and an intelligent interpretation of the research results. This aspect needs special attention for the remote access service, since the number of remote users may grow heavily. The solution we are heading toward is to provide the means to write research reports via the remote access infrastructure, so that intermediate research results need no longer be checked. We expect this approach to keep the remote access facility scalable from an organizational point of view.

6. From a technical point of view, the remote access facility is built up of a number of microdata servers, Citrix servers and a web server. These servers are behind firewalls and disconnected from the production infrastructure of Statistics Netherlands. Applications can be launched from the web interface to run microdata that has been imported to the (shared) workspace on the remote access data server beforehand. Authentication is implemented using biometric identification (fingerprints) and PKI certificates.

7. Statistics Netherlands has the policy that all output is available to everyone free of charge. There is no reason *not* to apply this principle to access to microdata. However, access to microdata differs from access to aggregated statistics in the fact that it requires additional work. Therefore, the pricing policy of Statistics Netherlands is that, although we do not charge for the microdata itself, the costs for using the microdata services are charged. This is inevitable, since the use of microdata services requires additional resources from Statistics Netherlands for using

the technical facilities, for creating additional documentation, for advice on the use of the microdata and for checking the results of researchers on disclosure risks.

8. In future, facilities for secure access to microdata could be integrated more closely with the dissemination facilities for aggregated statistics. For example, everyone can browse all information in the statistical output database *StatLine*, and in addition authorized researchers can "zoom in" to the underlying microdatasets that were used to create these results (top down). In addition, microdata researchers working on a specific dataset could easily retrieve the official statistics that were published based on these and other datasets (bottom up). Of course, researchers are only allowed to perform this type of operation if they have explicit permission to do so and when their identity is checked using biometric authentication. Unauthorized users cannot browse to the underlying microdata, but they do have access to the metadata of this microdata.

9. First of all, this would make Statistics Netherlands a more transparent organization to its customers. Second, this makes it possible for researchers to browse the metadata of the microdata archives for data that might be interesting for their research area. As a fist step in this direction, Statistics Netherlands decided to make the documentation of well-documented microdata sets available on the Internet and integrate them as much as possible with the thematic entrance to aggregated statistics on the website. This is only a first step to increased integration of micro and aggregated data in the statistical dissemination process. Even more powerful services, such as calculating statistics at run-time based on the interest of the user, may be achievable in future.
