DGINS 2005/91/2.3/EN

91ST DGINS CONFERENCE

COPENHAGEN, 26 + 27 MAY 2005

National Statistical Systems and Researchers

Svante Öberg Director General, Statistics Sweden

National Statistical Systems and Researchers

Svante ÖBERG Director General, Statistics Sweden

In this paper, I discuss the relations between national statistical systems and researchers from the point of view of the statistics producer. I argue for free access to publications and databases on the Internet, for developing systems for online access to microdata for researchers, and for improving cooperation with researchers. I refer to the situation at Statistics Sweden as an example. I offer these thoughts for discussion without pretending to know how well they would suit individual statistical institutes.

1. Introduction

The research community is an important group of users of statistics. For us at National Statistical Institutes (NSIs) it is of utmost importance to provide researchers with statistics and microdata as a basis for their research. In this paper I present some thoughts on how in the future to better serve researchers by free access to databases over the Internet and in particular better access to microdata, which I see as the next step in disseminating statistics in an even more flexible way. I also present some thoughts on how to develop the cooperation with researchers at the NSIs and on the European level.

2. Free access to publications and databases on the Internet

Twenty-five years ago printed tables in paper publications were the normal way of disseminating statistics. They provided little in terms of output flexibility. Typically statisticians had to decide on what tables to publish and design the survey accordingly. If researchers wanted to use the statistics for further calculations they had to transfer the numbers by hand or have them transferred to their own programs.

In the mid 1970s, Statistics Sweden began to build up databases. They were made available on a commercial basis via terminals over the telephone network. We charged the users to be able to finance the work connected with the databases. In January 1997, the databases were made available over the Internet.

In January 2000, we changed our dissemination policy and made the databases available free of charge. Before that we charged each user a yearly fee of SEK 6 000 (about EUR 650). We had about 350 paying users and some 100 non-paying users (libraries etc). The Government compensated us for the loss of revenues (about EUR 200 000).

1

Within three months we had more than 10 000 users. The number grew successively to about 30 000 by the time we stopped counting because users did not any longer have to register on our website to get access to the databases. This last change in procedures resulted in a doubling of the number of database withdrawals.

In January 2000 we also decided that all our publications should be made available in electronic form on our website and that we should not charge users for downloading publications. One important reason for not charging users for downloading publications was that it would not worsen our finances. Although our subscription revenues for printed publications would probably be reduced, so would our costs by about as much. The revenues covered only the costs for paper, printing, distribution and administration.

The website is now our main way of disseminating statistics. Last year we had 2.7 million visits on our website, not counting visits from search engines.

This experience shows the benefits of free access over the Internet to publications and databases. There are also good theoretical arguments for considering statistics as a public good, something that should be financed by a general tax and made available free of charge. The main reason is that the marginal cost for disseminating an extra copy of a publication or a database series is very small, much smaller than for a paper copy with the same content, and typically less than the benefit for the user.

At the same time NSIs have to find ways to finance development costs. If it is impossible to get funding by the Government it might be necessary to charge users. Another possibility to reduce the costs for making publications and databases free of charge, is to rationalise the production system by focusing on dissemination over the Internet as the main dissemination channel. When publications are prepared, they should be prepared in a way that allows both printing and downloading from the website without extra costs.

I strongly believe in free access to publications and databases over the Internet. I am particularly pleased that Eurostat recently has changed its dissemination policy in this respect. Statistics on developments in the European Union and in the Euro Area are becoming more and more important as Member States grow more integrated and European policies become increasingly developed.

In this area I have two suggestions. Firstly, I would suggest that the OECD Secretariat changes its policies in the same direction. The OECD Secretariat has very high quality publications and a wide range of very valuable databases that would be used to better advantage if they were made available free of charge. In Europe there has been much focus on harmonisation of statistics over the last ten

2

to fifteen years. It is now time to focus more on the comparability between Europe and other parts of the world, and in particular with other OECD countries. This would be supported by free access to the OECD's extensive data archive.

Secondly, I would also suggest that NSIs that still charge users for publications and databases on their websites to review their policies, assess the arguments for and against free access and consider possible ways of making publications and databases free of charge on the Internet.

3. Systems for online access to microdata for researchers

Anonymised microdata files, i.e. data files with records on individual persons or companies but without identification numbers or similar, can be seen as the next step for NSIs to serve its users with a more flexible output. Databases available over the Internet make it possible for the user to compile a wide range of tables. Anonymised microdata files provide even more flexibility. Technological advances have also made it easier for researchers to use microdata in their research and the demand for microdata is increasing.

Statistics Sweden has a long tradition of collecting administrative data and transforming these data to registers applicable for statistical uses. Swedish statistics are to a great extent based on administrative registers - from 85 to 95 % depending on how we measure. The register system also includes a number of survey-based registers, e.g. results from the Labour Force Surveys. Over a number of years, Statistics Sweden has distributed anonymous microdata to a large number of research institutions and authorities using magnetic tapes, CD-Rom discs, DVD discs or other formats. The volume has increased at the same time as the number of releases/assignments has increased. Around 170 releases took place last year.

Confidentiality protection of individual and business data is one of the main principles in official statistics and must be addressed when discussing microdata. The individual is entitled to be protected by unacceptable intrusion into personal privacy. The use of statistical information is therefore normally regulated in the legislation and/or in a code of practice.

In Sweden, the legislation regulating the use of statistical information states that all data, including anonymous data, obtained for statistical purposes are confidential irrespective of the source. Data collected for statistical purposes may be used only for the production of statistics or for research purposes. Confidential data may only be disclosed to authorised people. Access may be granted in forms, which do not allow direct or indirect identification of individuals or of other data subjects such as enterprises. In practice, Statistics Sweden only provides access to microdata without name, address and identification number. Legislation in Sweden, as well as in other Nordic countries, does

not contain any specific rules that restrict the way of releasing microdata. As long as the general requirements in the legislation are fulfilled, the most suitable method can be chosen.

At Statistics Sweden, we are now developing procedures that would make it easier, cheaper and more secure to use microdata in research. The main idea is that we should keep microdata in physical form at Statistics Sweden instead of distributing microdata to research institutions using CD-Rom discs etc, develop techniques to make it easy for researchers to make calculations over the Internet on microdata at Statistics Sweden, and seek funding for developing and managing the system to make the marginal cost for the individual researcher very low.

A unit was established last year to take charge of this initiative. It is called Register Coordination and Microdata Access and is placed at the Department of Research and Development. One of the objectives is to build a comprehensive register system, the Statistics Sweden Data Warehouse, in which variables and populations in different registers are standardised and interlinked, responsibilities are clarified for each register and for different stages in the coordination work.

A system for online access to microdata at Statistics Sweden has been developed. The system, which has largely been inspired by a similar system at Statistics Denmark, is called MONA (Microdata Online Access at Statistics Sweden). It allows researchers to have online access to specific servers at Statistics Sweden. A desk top with relevant software (e.g. SAS or SPSS) and with access to microdata is made up for the researcher, who can work with the data quite freely, compile his/her own data sets, carry out calculations and the like. However, all data processing will be carried out on the server at Statistics Sweden and no downloads are allowed. The results are frequently sent by e-mail to the researcher as tables.

One important aspect in future development is to compile several new thematic registers tailored to better meet the needs of the research community (e.g. LISA – an integration register on illness leaves and employment). To accomplish this, considerable work is needed, engaging both methodologists and subject matter experts. Another future trend is to develop techniques that allow linkage of data from different sources, both within and outside Statistics Sweden.

However, improved access to microdata involves relatively high costs. Without funding from the Government, which we do not have, costs must be borne by researchers. Because of this, Statistics Sweden has approached The Swedish Research Council suggesting funding from the Council of a system of microdata access. Such funding would give researchers lower marginal costs when using microdata. This would foster an increased use of microdata in research. Experiences from other areas where basic financing has been arranged and researchers only pay for low marginal costs have

been very positive. Such a solution would include full IT-support for online access via Internet, a front office to serve and advice the researchers, and thematic databases.

Statistics Sweden gets many requests for microdata from universities and research institutions in the United States, who want to enrich their own research databases with, for example, census data and register data from us. For several reasons, including legislatory restrictions, we are not able to satisfy such requests, if they require data to be physically moved abroad. However, there seem to be technical possibilities to develop the Danish/Swedish model described above into a more general, international network solution, where all data could physically stay where they belong, while making them at the same time available to researchers all over the world. On a national level, this kind of network solution is being implemented in Australia at present, the so-called National Data Network (NDN), where the Australian Bureau of Statistics is one of the nodes in the network, which also contains other agencies and research institutions.

In this area I have three suggestions. Firstly, I would suggest that NSIs that are ready to develop systems for online access to microdata form a network to exchange ideas and experiences. I am convinced that research community would benefit substantially from access to such facilities. Statistics Sweden is moving in this direction and we would certainly want to take an active part in such a network.

Secondly, I would suggest that in the international context, we take a step-by-step approach, first agreeing on some guidelines on confidentiality and microdata. Such guidelines are now being developed under the auspices of the Conference of European Statisticians by a task force led by the Australian Chief Statistician Dennis Trewin. Attempts last year in both the Statistical Programme Committee at Eurostat and the Committee on Statistics at the OECD to rush developments did not obtain strong support.

Thirdly, I would suggest that we explore the possibilities to create international microdata networks, enabling researchers from other countries than our own to get access to our data in a safe way.

4. Organisation of the cooperation with researchers

Most if not all National Statistical Institutes have some kind of organised cooperation with researchers. This cooperation may take different forms depending on the size of the country, traditions etc.

At Statistics Sweden, we have twelve user groups for different subject matter areas: economic statistics, welfare statistics, demographic statistics, regional statistics etc. These user groups consist

5

of external experts including researchers and they advise us on the development of statistics in their areas of interest. We also have a Scientific Council, a Consumer Price Index Committee and a Building Index Committee with representatives from the research community. This year we established a National Accounts Committee to give us advice on methods used in that area, of course within the framework of the European System of National Accounts.

Furthermore, there is an organised cooperation with the Universities of Stockholm and Örebro in particular, the cities where we have our offices. A cooperation agreement with the University of Örebro was signed in 2001. It includes a shared professor in statistics, shared financing of PhD studies, and regular courses, seminars and summer schools. I am a member of the Board of the University of Örebro and several professors at the University of Örebro are members of the different groups I mentioned above. A professorship at Stockholm University is specially dedicated to official statistics, and the professor spends part of his time at Statistics Sweden.

In addition to these more formalised and extensive types of cooperation, there is widespread cooperation between different parts of Statistics Sweden and researchers in many areas. Some of our staff members are teachers at universities and some university researchers work part time with us. Finally, we conduct regular user satisfaction surveys that are directed towards researchers, among others.

Comparing Statistics Sweden with other NSIs, I have noted that we do not have the benefit of such an extensive cooperation with universities as in the United States. NSIs in some other European countries such as the United Kingdom and France are also in a much better position in this respect than Statistics Sweden. Also, we do not have much in-house research and analyses like NSIs in other countries such as Canada, Australia and Norway. And we do not have in-house higher education in statistics like INSEE in France.

In this area, I also have two suggestions. Firstly, I would suggest that in most NSIs we probably could and should strengthen the cooperation with researchers. I believe in benchmarking and in learning from good examples in other countries. In fact many of the changes that are now taking place at Statistics Sweden originate from visits to other countries. I do not envisage that we should adopt the same kind of organisation of the cooperation, but that we could pick up ideas that would suite our particular national systems.

Furthermore, I have also been struck by the difference between NSIs and Eurostat in how we treat relations with users. In most NSIs we have very elaborated systems for dialogue with users. They vary from country to country. Some have National Statistical Councils that cover the whole

spectrum of official statistics. Some, like Statistics Sweden, have a number of user groups for different areas of statistics. Most NSIs have frequent contacts with branch organisations etc.

Most of this is lacking on the European level. We do have the CEIES, which is now under review. I think it is not unfair to say that CEIES has not played an important role as a user group. Instead, user views have been channelled through the different General Directorates of the Commission and the working groups and committees of Eurostat. However, the seminars CEIES has organised have been of high quality and much appreciated.

My second suggestion in this area is therefore that Eurostat takes an initiative to assess how to develop its relations to users including researchers, taking into account experiences in member countries. Would the European Statistical System benefit from user councils for European Statistics as a whole or for different areas of statistics? Would it benefit from a Scientific Council or from special Councils for National Accounts, Consumer Prices, or other areas of statistics? Would it benefit from user satisfaction surveys covering different areas of statistics?

5. Conclusion

In this paper, I have discussed the relations between national statistical systems and researchers from the point of view of the statistics producer. I do see possibilities for us to improve our output and make it more valuable to researchers. In particular, I argue for free access to publications and databases on the Internet and for developing systems for online access to microdata for researchers. Also, I do see possibilities to improve cooperation with the research community. I offer these thoughts for discussion understanding that other solutions might be more relevant for individual national statistical institutes.