Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

67th plenary session
Geneva, 26-28 June 2019
Item 6 (a) of the provisional agenda
Emerging role of national statistical offices as offices for statistics and data -
Session 1: Emerging data system opportunities and issues

Emerging roles of national statistical offices as offices for statistics and data[1]

Prepared by Statistics Denmark

Summary

Statistical agencies are transforming themselves from 'factories' of data to modern service organizations that provide not only data but also services related to data access and use of microdata. This document views both opportunities and challenges from the viewpoint of a small statistical agency with a well-integrated statistical production system based to a high degree on data from administrative registers – and with a high degree of trust from the population towards state agencies.

Statistics Denmark presents its experience in becoming an agency for both statistics and data approximately 30 years ago with giving access to microdata for researchers, and developing a production system linking civic registration information with data on education, labor market participation, income, health, etc. There are increasing demands from other parts of the public sector to use the competences of the national statistical office to ensure the best data foundation also for administrative decisions. This poses a challenge but also an opportunity for statistical agencies to use their competences for the public good in a broader sense than producing official statistics.

This document is presented to the 2019 Conference of European Statisticians seminar on "Emerging role of national statistical offices as offices for statistics and data", session 1 "Emerging data system opportunities and issues" for discussion.

---

[1] The present document was submitted late due to resources constraints.

I. Introduction

1.	The role of statistical agencies is changing as data and statistics become more and more important as the basis of societal development and decisions both in politics and in enterprises worldwide. Statistical agencies are already transforming themselves from 'factories' of data to modern service organizations that provide not only data but also services related to data access and the use of microdata for statistics and analysis performed by users from outside the statistical agencies.

2.	This development is an excellent opportunity for statistical agencies to increase their usefulness as important societal institutions in the age of data – but it is an opportunity not without its challenges.

3.	This document views both opportunities and challenges from the viewpoint of a small statistical agency in a small country with a well-integrated statistical production system based to a high degree on data from administrative registers – and with a high degree of trust from the population towards state agencies like e.g. the statistical agency. One the one hand this kind of experience might be of interest for countries and agencies in a different situation – and on the other hand, it is understood that the experience of each individual country is special.

II. The Data Center at Statistics Denmark

4.	Statistics Denmark developed into an agency for both statistics and data approximately 30 years ago when access was given to use microdata for researchers and research projects.

5.	The background for this was a development from the introduction of a unique identifier for every person born or living in Denmark in 1968. Based on this Statistics Denmark built a register based statistical production system through the 1970's. The last survey based census was conducted in 1970, and in the early 1980's the development of a production system linking civic registration information with data on education, labor market participation, income, health etc. was finalized.

6.	In the beginning researchers had to conduct the analysis on the premises of Statistics Denmark (an on-site arrangement). In 2001, an on-line system was established and it soon became the most popular way of accessing the data. The on-site arrangement was therefore closed in 2008.

7.	The data center has been immensely popular as shown in figure 1. The number of authorized institutions has increased from way below 100 in 2001 – when the on-line service were established – to close to 500 today.

8.	Access is given to authorized institutions, which in most cases are research departments and units at universities, but access is also given to analytical units in government departments and municipalities and to private sector consultancy firms.

Figure 1
Development in total and new authorized institutions at the data center at Statistics Denmark 2001-2017
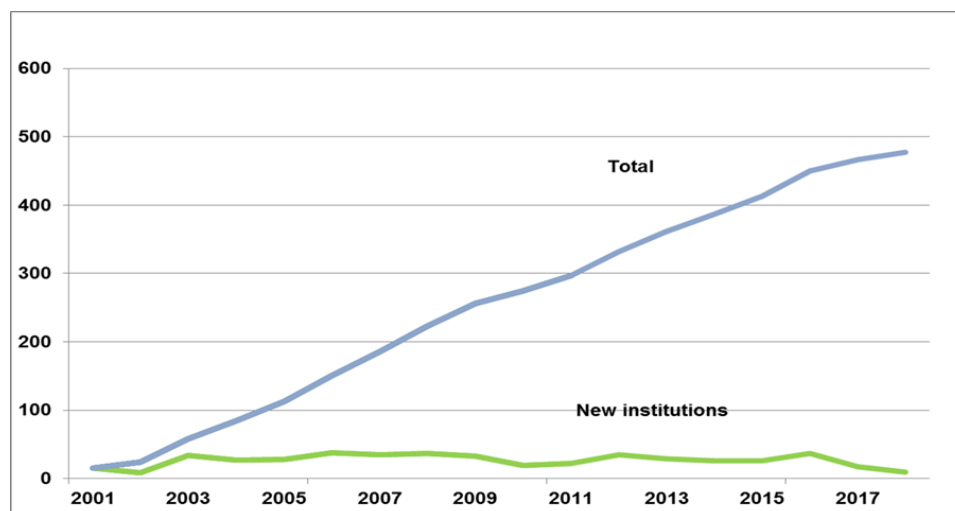
9.　To become an authorized institution one must have a stable analytical environment, with responsible managers, researchers and analysts with experience in handling microdata. The National Statistician grants authorization.

10.　The legal basis for access is granted in the law of Statistics Denmark. It states that on top of its obligations to produce statistics it can give access to the use of microdata as long as the users cover the costs of setting up this operation. The Danish Act on Protection of Data 2018 underlines that information collected for statistical purposes can be processed for the use of other statistical or scientific purposes. The Public Administration Act 1985 states that individuals who have access to microdata for research purposes are sworn to secrecy, and the Danish Penal Code set the punishment for a breach of secrecy to a fine or imprisonment for up to 6 months.

11.　The data security set up from the side of Statistics Denmark has the following features:

- Microdata are de-identified – the user does not get access to the unique personal identifier but to a number that contains no information about the induvial

- Microdata never leaves Statistic Denmark – they are contained on special dedicated servers that are separated from the statistical production system at Statistics Denmark

- Only results i.e. output where it is impossible to identify persons and companies must be sent out from the research machine/server

- Table cells have to contain more than 3 observations (minimum)

- Each output is scanned for microdata before it is sent out

- Researcher receives a warning if the system suspects the output to contain microdata

- If users take away output that violate the rules, their access to microdata are closed for a month and they have to deliver a plan that ensures that violation will not happen again.

12.　For each research project, the data center creates a dedicated data set for the specific project. The data sets are created based on the content of the project description in dialogue with the researchers based on a 'need to know' principle. Larger research institutions with many projects running at the same time can get the right to create a database that can cater for more projects. The goal of the data center is to deliver the agreed upon data no later than a month after agreement is reached.

13.　In recent years the data center arrangement for researchers and private company users has been extended with a new arrangement for government agencies that gives a broader and faster access to data.

14. This development is a reaction to a tendency from core users who want extended – close to unlimited – access to data. The background and challenges in this development is covered in the following section of this paper.

15. The demand for data is reflected in the organization of the personnel of the data center as they are divided into four teams:

- Ministry-customer team

- Health-customer team

- Private-customer team

- General public customer team.

16. Some data are more popular than others. As can be seen in figure 2, data on population, education and income are in the top three most demanded.

Figure 2
Demand for data organized register by register for the Statistics Denmark Data Center in 2017

Registers	Long name	Ordered by number of extracts in 2017
BEF	Population (status 31. december)	1
UDDA	Education (status)	2
IND	Personal income	3
AKM	Working classifications module	4
LPRDIAG	Hospitalized patient – diagnoses	5
RAS	Register–Based Labour Force Statistics	6
LMDB	Prescription Medicine data base	7
IDAP	Integrated data base on labour market research	10
IDAN	Integrated data base on labour market research – Employment and salary data	11
FAM	Family formation	12

III. Potential future developments – challenges and opportunities

17. The demand for data and the demand for fast data access seems to be limitless. Statistics Denmark has a huge community of dedicated users of the Data Center – and even though they usually praise the access to data given to them – they are also demanding customers who want more data and faster.

18. As stated earlier in this paper the data access for individual research and analytical projects created some 30 years ago has recently been supplemented with a broader and faster access to data for analysts at government departments. This group of users does not prepare project descriptions as researchers do. Often waiting for data for a month is not a viable option when the results are to inform political negotiations and decisions. Therefore, the access to data are broader but the restrictions in the use of data mentioned in para 11 still apply.

19. Currently about five ministries have signed up for the new arrangement. One of these has taken it a step further and has asked Statistics Denmark to build a data warehouse that combines data from Statistics Denmark with data from the ministry. The data warehouse is contained inside the firewall of Statistics Denmark. The rules for the use of data and the output that can be taken home are the same for the data warehouse users as for other users of the data center.

20. This last development is potentially just the first in a row of demands to use data stewarded by statistical agencies for administrative purposes – which everyone in the statistical community know is not allowed. But instead of rejecting the request statistical agencies could take pride in the fact that their data are seen as the gold standard of well documented data. Of course, statistical data cannot be used for administrative purposes for

many good reasons. But that does not mean that statistical agencies cannot be of service for government agencies who want to make use of the fact that statistical agencies are good at collecting, processing, checking and disseminating data. Therefore, we need to find solutions so that the knowledge behind that ability can be used also for non-statistical data.

21.   If and when statistical agencies start to change their concept of themselves from being factories for the production of statistics to becoming data service facilities for the public sector and for the public good, this has to be done explicitly and communicated to the public. There has to be full transparency of what is done with the data, how they are still treated in a secure way, and why statistical agencies are engaging in a new activity like this.

IV. Conclusions

22.   In a data driven world it is natural that the role of statistical agencies is challenged and changing as a response to the challenges.

23.   It becomes more and more obvious that statistical agencies not only hold data that are in demand for new usages like access to micro data – but that they also have capabilities in the collection, processing and documentation of data. These will be increasingly in demand from other parts of the public sector that want to use those competences to ensure the absolutely best data foundation also for administrative decisions. This poses a challenge for the agencies but also an opportunity for statistical agencies to use their competences for the public good in a broader sense than originally.

_____