UNITED NATIONS

E

Economic and Social Council

Distr.
GENERAL

CES/2003/20
21 May 2003

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-first plenary session
(Geneva, 10-12 June 2003)

CHALLENGES FOR TRADITIONAL APPROACHES TO CONFIDENTIALITY PROTECTION – THE DANISH EXPERIENCE

Supporting paper submitted by Statistics Denmark[1]

I. THE DANISH CONCERNS

1. Confidentiality protection is one of the top priority policy issues of Statistics Denmark because of the conflicting interests between, on the one hand, making best use of the data and, on the other hand, protecting personal data of citizens and enterprises who have, directly or indirectly, entrusted their data to us. Therefore, the Board of Statistics Denmark made a number of important decisions on the issue in the past year, and will continue discussions.

2. It should be noted that because of the way statistics are produced in Denmark, making intensive use of data from administrative registers, Statistics Denmark possesses an immensely rich database with billions of individual data. Most of the data are endowed with precise identifiers and can thus be linked. These data are never passed on to outsiders, not even for research purposes ("the One Way Street"). However, the needs of researchers and analysts are so strong and qualified that they may be allowed, under specific circumstances and strict control, to access unidentified micro data.

[1] Prepared by Lars Thygesen.

II. DATA USERS VERSUS DATA SUBJECTS

3. The conflict between people who want to make use of data and respondents is a real one. But in recent years in Denmark, the data users have been much more prominent in the public debate than the data subjects.

4. It is obvious that the total data warehouse of Statistics Denmark is a virtual gold mine for researchers and others – including the Government and Parliament – who want to generate a strong and detailed basis for their decisions. The richness of the data on individual citizens and enterprises allows for innumerable valuable analyses. Statistics Denmark recognises this and has for many years sought ways to accommodate the needs in ways that do not threaten confidentiality.

5. There has been continuous and strong pressure, especially from researchers, to obtain access to "everything". Why should there be restrictions at all, since the researchers only want to identify structures in the data, not to snoop into the private lives of persons or enterprises? Danish researchers claim that practices of NSOs in other countries are more "research friendly". It is difficult to argue that Statistics Denmark must be more confidentiality conscious than other NSOs – and for this reason international discussions like this one in the Conference are important. The result of the pressure has been a steady move towards more and more access.

6. But at the same time, we have to take the perception of respondents into account. They haven't been very active in the debate, maybe because they don't know exactly what is going on. Yet we see the risk of a major publicity disaster. This, together with the European data protection legislation and its Danish implementation, makes it necessary to keep up strong measures. And the Board, having businesspeople as well as researchers as members, is very much aware of the importance of protection, as well as the user perspective.

III. THE DANISH SOLUTION

7. The 2001 negotiations between Statistics Denmark, the Ministry of Research and the Research environment resulted in the signing of a contract on the establishment of a special unit (the Research Service Unit) in Statistics Denmark with the particular duty to improve researchers' access to micro data through a better infrastructure and to lower the costs of using the data. The budget for the Research Service Unit is 6 million D.kr. per year (approx. 800,000 Euro). Some of the money is used to upgrade the special Unix computers, cf. below.

Principles for access

8. Statistics Denmark has created an advanced solution allowing researchers and policy makers to have access to micro data – always unidentified – based on the Need to Know Principle. This means

that they can only obtain the data needed for their purpose, and Statistics Denmark has to make the judgement on that. Micro data must never leave the NSO, but users obtain access to "their" data sets in Statistics Denmark via a virtual private network, and the use is under strict control and observation. A log will be used for investigation in case of confidentiality violation.

The technical solution

9. Since 1986 Statistics Denmark has given access to researchers to analyse micro datasets from work stations in Statistics Denmark ("the in-house researcher arrangement").

10. In 2001, Statistics Denmark launched a new system granting access for specially authorized research and analysis environments to approved datasets from their own workplaces. A research or analysis environment can apply for an authorization from Statistics Denmark. As of 15 March 2003, 43 environments had been granted authorization. The wording of the authorization appears in Appendix 1.

11. The technical solution is based on a virtual private network on the Internet, see the chart in Appendix 2.

12. The relevant micro data are produced by the staff in Statistics Denmark and the de-identified micro data are transferred to the disk storage connected to special Unix servers dedicated to the researchers. These Unix servers are separated from the production network.

13. Communications via the Internet are encrypted by means of a so-called RSA SecurID card, a component that secures Internet communications against unauthorized access. In practice the researcher rents a password key (a token) from Statistics Denmark. The token ensures that only the authorized person obtains access to the computer system.

14. A farm of Citrix Servers ensures that the researchers from their own workplace can "see" the Unix environment in Statistics Denmark. All data processing is actually done in Statistics Denmark and data cannot be transferred from Statistics Denmark to the researcher's computer. The researcher can work with the data quite freely and can make new datasets from the original data sets.

15. All results from the researchers computer work can be stored in a special file and such printouts are sent to the researchers by e-mail. This is a continuous process (every five minutes) and has proven to be quite effective. The advantage to Statistics Denmark is that all e-mails are logged at Statistics Denmark and checked by the Research Service Unit. If the unit find printouts with too detailed data, contact is made with the researcher in order to agree on details of the level of output. No severe violation of the rules, established in the authorisation formula, has taken place.

16. In December 2002 the Board of Statistics Denmark decided to consider the on-site scheme and the remote access scheme as equivalent concerning data security and as a consequence of this decision all data sets which can be accessed from on-site can also be accessed from remote. This has led to an explosion in the use of the remote solution, and it is envisaged that the in-house arrangement will soon become obsolete and be closed down.

17. With this decision it has been very important to revise the rules for granting authorisation to micro data. The new rules can be seen in appendix 3

IV. PERSONAL DATA VERSUS BUSINESS DATA

18. In Denmark we have found it expedient to distinguish between micro data on citizens and on enterprises, the latter containing more problems than the former, since it is difficult to avoid the fact that people obtaining access will immediately recognize well-known companies even without identifiers. For this reason, we have decided to restrict access to business micro data: data have to be at least one year old, and access cannot be granted to research departments of private companies.

V. AUTOMATED CONFIDENTIALITY PROTECTION

19. It has been argued in international discussions that protection may be performed using automatic tools, e.g. giving access to scrambled micro data or using tools like a “statistical fire wall”, giving people access to process micro data without revealing the identity. We haven’t found evidence that this is a fruitful path, partly because it may distort the inference which the users are looking for; if this risk is avoided, we fear the risk that there may be ways to circumvent such tools, e.g. by multiple data requests.

20. Consequently, we haven’t found good solutions to giving general access to extremely detailed data via the web, e.g. geographical information systems where users would be able to compose their own detailed geographic break-down “on the fly”. Our database on the Internet, www.statbank.dk, is extremely detailed and rich, but the levels of geographical break-down are predefined and fixed.

APPENDIX 1: AUTHORIZATION FORM

Statistics Denmark

AUTHORISATION

Statistics Denmark hereby grants

[Institution] represented by [Chief Researcher]

Authorisation for

Remote electronic access to selected datasets at Statistics Denmark

Remote Access via the Internet is subject to the following terms:

1. A project description must be submitted, which states the project objectives and renders it possible to select the data required for successful project execution.
2. Based on the project and data description, Statistics Denmark decides whether external electronic access to data can be granted for the specified project. If the authorisation is not granted, the researcher is referred to use the ordinary scheme for the on-site arrangement for external researchers at Statistics Denmark.
3. The researcher to whom external electronic access is granted shall sign a special agreement with Statistics Denmark, cf. appendix.
4. All datasets are confidential, cf. §27(3) of the Danish Public Administration Act and §152 of the Danish Criminal Code.
5. The researcher obtains access to make batch runs on Statistics Denmark's special researcher machines (UNIX system) from one or more PCs specially assigned for that purpose in the research/analysis environment. Access is denied for batch runs from remote PCs, PCs at home or PCs which cannot be properly supervised.
6. Only the client software assigned by Statistics Denmark may be applied in connection with the RSASecurID card provided. A PC connected to Statistics Denmark may not be made available to unauthorised persons, and when the user leaves the PC, the PC must be either shut down or disconnected, i.e., protected from any unauthorised use.
7. The password of the individual researcher is personal and strictly confidential.
8. The researcher may not, directly or indirectly, download the dataset or any datasets derived there from. All transfers of output for printing or further statistical processing (in spreadsheets or similar) must be executed in accordance with the guidelines and methods laid down by Statistics Denmark. Statistics Denmark will create a log file of such authorised transfers. Furthermore, individual records

may not be printed, and all output must be aggregated to an extent that eliminates any risk of direct or indirect identification of persons or enterprises. The researcher may not attempt to make such identification.

9. Statistics Denmark shall be entitled at unannounced visits to check that the rules of this agreement are observed.

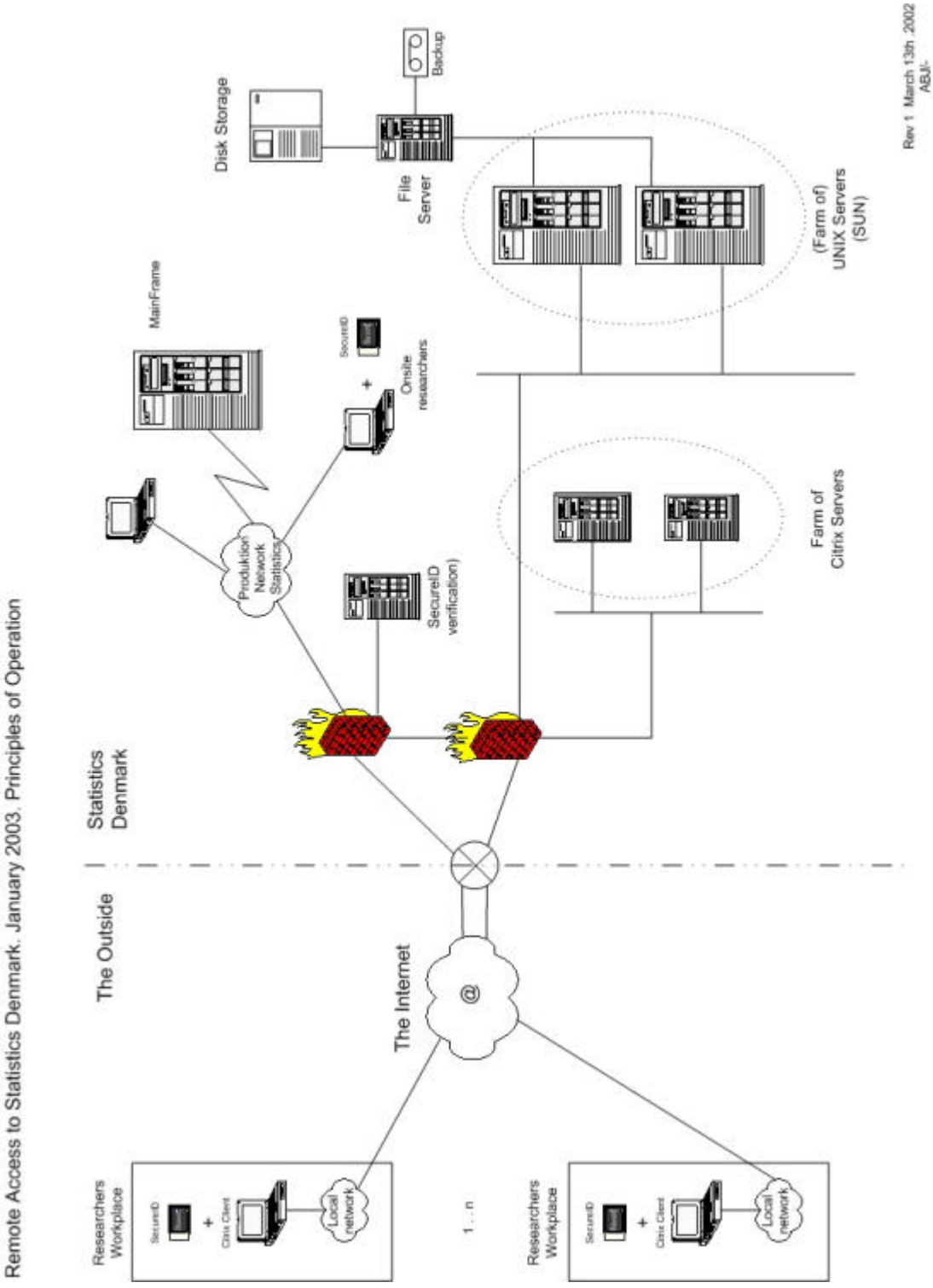
10. The person signing this agreement on behalf of the research/analysis environment shall ensure that publications by the environment do not contain any information that may identify individual persons or individual enterprises.

11. The person signing this agreement on behalf of the research/analysis environment undertakes personally to supervise or to appoint a person to supervise that the provisions of this agreement are observed.

12. In case of breach of the provisions of this agreement, the researcher in breach will be excluded from using any researcher schemes of Statistics Denmark permanently or for a period of not less then three years. Furthermore, in the case of breach hereof, this authorisation will be withdrawn for a period.

This agreement, which is signed in two copies, enters into force on [date] and may be terminated by either party at three months' notice.

APPENDIX 2: THE TECHNICAL SOLUTION



APPENDIX 3: RULES ON ACCESS TO DE-IDENTIFIED MICRODATA UNDER STATISTICS DENMARK'S RESEARCH SCHEMES

At its meeting on 2 April 2003 the Board of Governors laid down the following rules on access to de-identified microdata[2] under Statistics Denmark's research schemes:

Who can obtain access?

Access is only granted to authorised research and analysis environments. Only research and analysis environments of a more permanent nature with a chief researcher and several researchers/analysts can be authorised, as sanctions in case of violation of the rules would otherwise have limited effect. Authorisation is granted by the Director General. The authorisation form is shown in Appendix 1.

Prior to granting the authorisation, Statistics Denmark makes a concrete assessment of the applicant's reliability as a data recipient. In respect of non-governmental organisations and enterprises it is relevant to examine the ownership, the staff (qualifications) and the assignments handled for public-sector clients in particular. The examination may include enquiries to such clients to in order obtain a statement.

When a research or analysis environment has been authorised, agreements may be concluded with specific researchers/analysts.
The following environments can be authorised:

1. *The user group defined under the framework agreement between Statistics Denmark and the Ministry of Science, Technology and Innovation* can be authorised and thus obtain access. This group comprises all employees in government funded research projects, employees in public research and analysis environments (i.e., universities, government research institutes, ministries, government agencies, etc.) and researchers employed with non-profit foundations in Denmark.
2. In the private sector, the following organisations with research and analysis environments of a more permanent nature are eligible for authorisation:
   a. *Non-governmental organisations*
   b. *Consulting firms* may be authorised, but cannot generally obtain access to microdata containing business data. The Director General may grant exemption to consulting firms that carry out investigations or research for a public authority, or to a non-governmental organisation that would be eligible for authorisation if its client guarantees, in writing, the correct use of data in terms of security.
   c. *Other individual enterprises* may be authorised, but cannot obtain access to microdata containing business data.

[2] There is no access to microdata with identification.

3. Danish researchers *who are working abroad for a period*, but who are attached to an authorised Danish research environment, may obtain external electronic access from their place of research abroad. In these cases the responsibility lies with the Danish research environment.
4. *Foreign research institutes* may, in exceptional cases and following a concrete assessment, be authorised to use the on-site research scheme for external researchers for as long as this scheme exists. They cannot use the external electronic access.
5. *Foreign researchers* who are working in Denmark in a Danish research environment may achieve external electronic access for the duration of their stay under the authorisation of the particular research environment, which assumes the overall responsibility.

What data can be accessed?

Access can only be granted to de-identified data, i.e., data for which all identification details such as name, ID number and address have been removed.

Access is granted according to the need-to-know principle, which implies that researchers/analysts can obtain access to the data required for the specified purpose[3]. Accordingly, the applicants have to document a reasonable relationship between the requested data content and the project description. If the issue requires total population coverage, access may be granted to total data material, otherwise a sample will be made available. In addition, data may be limited in the form of grouping or segments for confidentiality reasons.

Generally, authorised persons have access to all types of personal and business data with the limitations following from the above rules on consulting firms and other individual enterprises, and the need-to-know principle. However, anonymised business data cannot be accessed until one year after the reference period. Detailed product data on individual enterprises are anonymised before they are made available.

In certain cases, Statistics Denmark may deny the requests of a researcher or analyst because of insufficient data quality, primarily in connection with compilation of information from different statistical fields. This applies to both personal data and business data.

5.1 Cases of doubt

[3] This is in accordance with the principles of the Danish Act on Processing of Personal Data, particularly section 5(3): "Data which are to be processed shall be adequate, relevant an not excessive in relation to the purposes for which the data are collected and the purposes for which they are subsequently processed"; and section 10(1): "Data as mentioned in section 7(1) or section 8 may be processed where the processing is carried out for the sole purpose of carrying out statistical or scientific studies of significant social imp ortance and where such processing is necessary in order to carry out these studies."

Acting on the recommendation of the heads of divisions, the Director General decides any cases of doubt resulting from interpretation of the rules.