

TWINNING CONTRACT

CRIS 2015/370-467

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics



STUDY VISIT REPORT

on

COMPONENT B: MICRODATA SERVICES TO RESEARCHERS

Activity B5

Study visit to NSI's proving remote access to microdata for researcher purpose in two member states Organizational structure, procedures and processes

Date: 06 – 09 February 2017
Final version

Table of contents

| | |
|--|----|
| List of Abbreviations..... | 4 |
| Authors' names, addresses, e-mails | 5 |
| Summary of the study visit | 6 |
| 1. General comments | 9 |
| 2. Lessons Learned | 10 |
| 2.1 Step-by-step procedures from proposal to publication | 10 |
| 2.1.1 Institutional authorization..... | 11 |
| 2.1.2 Project request | 12 |
| 2.1.3 Intake Meeting and data extraction protocol | 14 |
| 2.1.4 Confidentiality statement..... | 14 |
| 2.1.5 Price calculation and financial contract..... | 15 |
| 2.1.6 Secure remote access environment | 18 |
| 2.1.7 Data extraction/creation of data views..... | 22 |
| 2.1.8 Analysis | 22 |
| 2.1.9 Output control..... | 23 |
| 2.1.10 Publications | 24 |
| 2.2 Organizational framework | 24 |
| 2.2.1 Internal organisation within the National Statistical Institution | 24 |
| 2.2.2 Organisation of the Research service unit | 25 |
| 2.2.3 Advisory council for research services..... | 26 |
| 2.3 Legislation, confidentiality and security | 26 |
| 2.3.1 Data confidentiality policy and interpretation of the “need to know principle” | 28 |
| 2.4 Supporting tools and documentation | 28 |
| 2.4.1 Data catalogue | 28 |
| 2.4.2 Data library to be used by the Research Service Unit | 28 |
| 2.4.3 Metadata and documentation..... | 29 |
| 3. Conclusions and recommendations | 30 |
| 3.1 Recommended basic steps for simplified procedures and responsibility | 30 |
| 3.2 Recommendation for organization | 30 |
| 3.3 Recommendation for data and documentation | 31 |
| 3.4 Recommendation for data for researchers | 31 |
| 3.5 Recommendation for output control..... | 32 |
| 3.6 Recommendation for IT and administration..... | 32 |
| 3.7 Next steps | 33 |

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

Annexes (external to the report):

Annex B5 – 1 Terms of Reference B5

Annex B5 – 2 Meeting program for B5

Annex B5 – 3 Persons met at B5

Annex B5 – 4 Comparative table for research services in Denmark and the Netherlands

Presentations by Israeli Central Bureau of Statistics (ICBS)

Annex B5 – 5 Current Status and future plans for Research Services in Isarel

Annex B5 – 6 The future administrative management system in for Research Services in Isarel

Presentations by Statistics Denmark (DST)

Annex B5 – 7 The High Qulaity Documentation System (HKD)

Annex B5 – 8 SAS data library

Annex B5 – 9 Cooperation with Subject Units

Annex B5 – 10 The price structure

Annex B5 – 11 Data confidentiality and output control in practice

Annex B5 – 12a Authorisation Contract with Statistics Denmark

Annex B5 – 12b Confidentiality agreement with the researcher at Statistics Denmark

Annex B5 – 13 FSESYS and standardisation - part I

Annex B5 – 14 FSESYS and standardisation - part II

Annex B5 – 15 Introduction to the Danish system for access to microdata

Annex B5 – 16 Output guidelines

Annex B5 – 17 FSE – SLA reports

Annex B5 – 18 SIEM log management

Presentations by Statistics Netherlands (CBS)

Annex B5 – 19 Statistics Netherlands Facts that matter

Annex B5 – 20 Legal issues on-site facilities and remote access and execution

Annex B5 – 21 Remote Access Microdata Infrastructure

Annex B5 – 22 Remote Access Innovations

Annex B5 – 23 Microdata availability

Annex B5 – 24 Output checking

Annex B5 – 25 Remote Access – from proposal to publication

Annex B5 – 26 Metadata and data management Data Service Centre (DSC)

Annex B5 – 27 Service catalogue

Annex B5 – 28 Output guidelines

List of Abbreviations

| | |
|------|--|
| CBS | Statistics Netherlands |
| DANS | Data achieving and Network service (The Netherlands) |
| DST | Statistics Denmark |
| EU | European Union |
| ICBS | Israeli Central Bureau of Statistics |
| MUC | Microdata under contract |
| NSI | National Statistical Institute |
| RS | Research Service Unit |
| RTA | Resident Twinning Advisor |
| RTAA | Resident Twinning Advisor Assistant |
| SDC | Statistical Disclosure Control |

Authors' names, addresses, e-mails

Ms. Sigalit Mazeh (Shmueli)

Director

International Relations and Statistical Coordination Department

Israeli Central Bureau of Statistics (ICBS)

Deputy Project Leader and Component Leader for the B Component of the Twinning

Email: sigalit@cbs.gov.il

Mr. Genady Etin

Head of Sector

IT Department

Israeli Central Bureau of Statistics (ICBS)

Email: genadye@cbs.gov.il

Ms. Anat Katz-Avram

International Relations and Statistical Coordination Department

Israeli Central Bureau of Statistics (ICBS)

Email: anatk@cbs.gov.il

Ms. Tamar Rand

Resident Twinning Adviser Assistant

The Twinning Project on Statistics (CRIS 2015/370-467)

Email: tamarra@cbs.gov.il

Ms. Charlotte Nielsen

Resident Twinning Adviser Assistant

The Twinning Project on Statistics (CRIS 2015/370-467)

Email: charlotten@cbs.gov.il

Acknowledgement:

Input given by all participants of the study visit at meetings after returning to Israel is highly valued by the authors of this report. Ms. Batia Attali, International Relations and Statistical Coordination Department is thanked for her assistant and useful comments for the current report.

Summary of the study visit

The study visit served as a necessary step to proceed with the initiated work at Israeli Central Bureau of Statistics (ICBS) related to implementing microdata services to researchers.

In February 2017 ten ICBS staff members, the Resident Twinning Advisor (RTA) and the Resident Twinning Advisor Assistant (RTAA) visited Statistics Denmark (DST) and Statistics Netherlands (CBS) - two of the most experienced MS countries in providing microdata to researchers with high standards for confidentiality and security. The ICBS delegation consisted of representatives from units that currently provide services to the research communities and who's cooperation is essential for improving research services: subject matter units, IT, the Confidentiality Committee, the Computer Security Office, the Legal Department as well as the International Relations and Statistical Coordination Department which is in charge of Research Services at ICBS. Besides getting technical input from the MS countries, the study visit also ensured a common understanding between ICBS staff members of the challenges and required steps towards providing microdata and services to researchers in a secure, professional, user-friendly and competent manner.

Both CBS and DST have a long tradition of giving authorized researchers access to de-identified microdata. In both countries microdata access for researchers started out as "on-site" access from designated research rooms located on the property the statistical institutions. In the year 2000 DST launched an encrypted secure remote access system to microdata placed on servers at DST so that researchers could work on data from their own computer through the Internet (remote access). Today access to microdata in Denmark is only granted as remote access and on-site facilities no longer exist. In the Netherlands's CBS offers authorized researcher access from research rooms in their offices in The Hague and in Heerlen¹ ("on-site"). Since 2005 CBS have also provide access from the researcher's own institution via a secure remote access. Since then the amount of research projects carried out at the "on-site" facility has decreased since the remote access is the preferred way of working with data for the researchers.

Both in Denmark and the Netherlands researchers are granted access to de-identified 'Secure-use files' meaning de-identified confidential data to which no further methods of statistical disclosure control (SDC) have been applied (SDC = local suppression e.g. aggregation into age groups, adding noise, synthetic data)². In both countries data are made available to researchers under strict conditions for statistical research ensured by law.

In addition researchers in the Netherlands have the option of working with selected 'Microdata under contract' (MUC files) to which methods of SDC have been applied to a level where there are no the risk of disclosure of individual persons, businesses, institutions or households. These files are distributed by the Data Archiving and Network Service (DANS) and can be stored on the researchers own computer.

In both countries authorized researchers can conduct their own research using the enormous amount of data produced by the National Statistical Institutions (NSI) as well as administrative data from other public agencies. In the Netherlands many of the datasets dates back to 1995 and in Denmark multiple datasets dates back to the 1980's, which make them very relevant for e.g. longitudinal research projects. The content of the registers also covers many fields of research such as labour market research, sociology, health and business. Research access to national administrative data from other agencies is ensured by either law or

¹ The special on-site working rooms, can be used every working day from van 8:30 am to 5:00 pm. A booking system is in operation.

² Commission regulation (EU) 557/2013 on access to confidential data for scientific purpose. Please consult this link: [Access to confidential data](#)

by bilateral agreements between the national agencies and DST and CBS. In addition these data can also be combined with the researcher's own data.

The strength of the system in both countries is that the identification keys (person number, address, central business register number and property title number) enable correlation between datasets both within a specific year, between subjects and, as mentioned, longitudinally across several years.

In both the Danish and the Dutch model a specialized research service unit as well as a specialized IT section have been created in order to professionalize and standardize the service provided. Currently DST provides service to ~500 authorized institutions, 2000 researchers and 1200 research projects. The Research service unit consists of the Head of the Unit as well as 20 academics (generalists), 1 IT-specialists, 1 secretary and the IT service desk has 2 employees. At CBS services are provided to 130 authorized institutions, 530 researchers and 700 projects. The services are provided by 16 staff members of whom 7 of them have an academic degree. The Data Service Centre has 8 employees. In order to match these ratios ICBS will need 3-4 full time staff members to provide services to the research community if assuming ICBS will have the equivalent IT and management structure (ICBS 75 projects and 100 researchers).

In order to be able to construct customized dataset the research service units in both countries maintain a data library for their research service unit. In Denmark the data library is constructed as a SAS library with direct identifiers such as CPR numbers. However, before data are placed on the research server data are always de-identified by replacing all direct identifiers by project specific random numbers allowing merging of dataset within a specific project but not across projects. At CBS the Data Service Centre maintain a data library with ASCII files where all direct identifiers are de-identified. Access to data is given to the researchers as views to full datasets with a common de-identification key for all projects

In both countries researchers will only be given access to the data which are needed for their research, the so-called "need to know principle". This means that researchers only can get access to the data required for the specified purpose of a given project. Accordingly, researchers have to document a reasonable relationship between the requested datasets and the objective of the project. In Denmark the "need to know principle" applies at the level of population, registers and register year as well as at the variable level. In contrast to Denmark, the Netherlands applies the "need to know" principle to the register and register year since access is always granted as views to full populations and registers. However, in the Netherlands each register usually contain fewer variables per register compared to Denmark.

In both the models applied by the CBS and DST a secured working environment is provided for the researcher in which they can store draft versions, syntaxes and outputs. Results from the researcher's analyses can, in both models, be exported to the researchers own computer as long as they keep the rules of confidentiality in each of the countries.

In Denmark outputs are not checked prior to export, but all outputs sent home by the researchers are logged and stored by DST. However, only a random sample of the exported outputs is checked by the Research Service Unit. The legal responsibility regarding the confidentiality of the output rests on the authorized institutions and its researchers. The obligation of DST is solely to provide guidance on confidentiality. Thus, in the event of a data breach, the institutions as well as the researchers are sanctioned. In Denmark output control is provided free of charge for the researcher since the cost are ensured by a yearly grant from the Danish Agency for Science and Higher Education. In the Dutch model all outputs are checked before they are mailed to the researchers. The output check is performed by at least two controllers from the Research Service Unit of the CBS. In case the researchers outputs cannot be accepted, CBS will contact the researcher in order to make sure that they understand the rules for outputs. CBS distinguish between "output light" that is usually approved on the same day and normal outputs that are approved within two to three working days. CBS charges the researcher € 200 for an output.

Both countries provide extensive output guidelines. In addition CBS requires all researchers to pass an online test the first time they enter a new research project in order to ensure that they do understand the rules for outputs. In addition questions about confidentiality will automatically pop-up at random intervals every time the researcher is logged on to their project.

An important aspect in both the Dutch and the Danish model is that there is a clear link between cost and income related to services. In the Netherlands, all costs are covered by the researchers and in Denmark the cost is shared between the researchers and external funding from the Danish Agency for Science and Higher Education.

After the study visit, it became clear that the future process of developing improved research services in the ICBS can benefit from consulting both Statistics Netherlands and Denmark, since each country has elements of interest for ICBS. In brief the expertise knowledge of particular interest from Statistics Netherlands is their procedures and rules for output control as well as their technical set-up of the remote access system based on virtual servers. When it comes to the interpretation of the 'need to know principle', Statistics Denmark seems to meet the ICBS's needs and legal requirements best. In other aspects such as e.g. the approval process, data catalogue, codebook, data library, documentation, mode of access, type of data files provided for researchers, price structure as well as organization and consulting a combination the Dutch and the Danish model will ensure that the vision of ICBS is fulfilled.

At the study visit the MS countries stressed the need for the ICBS to put more resources into getting general pre-approvals for using administrative data from Ministries and Agencies for research purposes. The situation in Israel today is that even though administrative data are highly used in Statistics published by the ICBS these data cannot be used by the researchers without approval of each individual research project by the provider of the administrative data – a process that might take up to six months and sometimes even longer.

1. General comments

This study visit report was prepared within the Twinning Project "Support to the Israeli Central Bureau of Statistics (ICBS) in enhancing the Quality of Official Statistics" and organized under component B: Microdata Services to Researchers. The study visit was the fifth activity in the component.

During the study visit the Israeli delegation³ met with staff members from Statistics Denmark (DST) in Copenhagen and Statistics Netherlands (CBS) in Den Haag. The study visit served as a necessary step to ensure further progress in the already initiated work at ICBS related to microdata services to researchers. The main topics covered were:

- *General setup and workflow*
- *Legislation, confidentiality and security*
- *IT – Technical and security setup*
- *IT – Management*
- *Responsibility and cooperation within the National Statistical Institution (NSI)*
- *Guidelines, training and other supporting materials*
- *Business plan and price structure*

Due to the broad spectrum of subjects and in order to ensure in depth knowledge would be obtained for all professionals, parallel sessions in IT and general research services took place part of the time in both countries.

Some of the topics listed above had already been presented and discussed in previous Twinning Missions with Danish experts. Therefore, the study visit in Denmark aimed to give the participants practical demonstrations of the set-up implemented in Denmark for researcher's access to microdata whereas the study visit in the Netherlands had more character of an overview of their set-up. However, for comparability some of the topics already covered at previous missions in Israel by the experts from DST are also summarized in the current report. In addition knowledge obtained at the study visit has been supplemented with information from written materials handed out at the study visit as well as from the websites of the two institutions:

- *Denmark*
<http://www.dst.dk/en/TilSalg/Forskningservice>
- *The Netherlands*
<https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research>

1Ms. Sigalit Mazeh (Shmueli), BC Project Leader Deputy of the Twinning Project and Director of International Relations and Statistical Coordination Department, ICBS.

Mr. Brian Negin, Legal Advisor of Israel's Central Bureau of Statistics (ICBS)

Mr. Ahmad Hleihel, Chairman of the data confidentiality Committee.

Ms. Anat Katz-Avram, International Relations and Statistical Coordination Department

Mr. Itzik Goldstein, Deputy Director of IT Department.

Mr. Genady Etin, Head of Sector, Technology in the IT Department.

Mr. Haim Plachinsky, Data Security Officer, Information Security Department.

Ms. Hadas Yaffe, Demography and Census Department.

Ms. Orly Furman, Micro-Economic Statistics Department.

Mr. Adnan Mansur, Head of Sector, Education and Teaching Statistics Department.

Ms. Tamar Rand, Resident Twinning Adviser

Ms. Charlotte Nielsen, Resident Twinning Adviser

Whenever relevant hyperlinks to specific information provided on the internet by the two intuitions has been inserted in the current report

The study visit gave the Israeli delegation a wealth of valuable ideas and inspiration for future development of a future microdata service to researchers in Israel. The staff of ICBS would like to express their thanks to all officials and individuals met for their kind support and the valuable information which they provided before, during and after the stay in Denmark and the Netherland's which highly facilitated the work.

This views and observations stated in this report are those of the participating ICBS staff and do not necessarily correspond to the views of the EU, DST or CBS. Any errors or misunderstandings are the responsibility of the participating staff alone.

2. Lessons Learned

In the following sections you will find a detailed review of lessons learned by the participants in the study visits to the CBS and to DST. A table summarizing the most significant points can be found in *Annex B5 - 4 Comparative table for research services in Denmark and the Netherlands*. The report is based on presentations and discussions given during the study visit and supplemented with additional information extracted from materials handed out by the National Statistical Institutions at the mission.

2.1 Step-by-step procedures from proposal to publication

Both at the CBS and at DST the process from proposal to publication is relatively simple (Figure 1). In both countries the formal process from initial request until the researcher can start working with data, takes in most cases only a few days to a few weeks depending on the complexity of the project. Although the basic steps in many aspects seem similar some differences in how each step is implemented in the two countries occur. In the following sections each of the steps are described in more details for each country. Several of the processes at DST were already presented and discussed in details under Mission B3. However, for the sake of comparability, the processes are also summarized in the current report.

| <i>Statistics Netherlands (CBS)</i> | | <i>Statistics Denmark (DST)</i> | |
|-------------------------------------|----------------------------------|---------------------------------|------------------------------------|
| 1 | • Institutional authorization | 1 | • Institutional authorization |
| 2 | • Project request | 2 | • Project description and approval |
| 3 | • Intake meeting(s) | 3 | • Data protocol |
| 4 | • Confidentiality statement | 4 | • Confidentiality Statement |
| 5 | • Price calculation and contract | 5 | • Price calculation and contract |
| 6 | • Create secure environments | 6 | • Create secure environments |
| 7 | • Data views | 7 | • Data extraction |
| 8 | • Analyse | 8 | • Analyse |
| 9 | • Output/invoice | 9 | • Output |
| 10 | • Publication | 10 | • Publication |

Figure 1: The basis steps from approval to publication at Statistics Netherlands (Blue) and Statistics Denmark (Green) respectively based on slightly modified figures presented by the two countries.

2.1.1 Institutional authorization

In both the Netherlands and in Denmark access to microdata can only be granted to researchers from authorized research environments after approval from National Statistical Institution.

Denmark

At DST access to microdata can be granted to researchers and analysts from research environments pre-approved by the National Statistician of Statistics Denmark. Authorizations can be granted to public research and analysis environments (e.g. university departments, sector research institutes, ministries, etc.), to research organizations as a part of a non-profit foundation as well as to private companies under certain conditions. Statistics Denmark will not grant authorization to single persons.

In order to obtain an authorization an application needs to be submitted to the Research Service Unit at DST. The application should include 1) a description of the research environment; 2) name and contact information for a responsible manager of the authorization, in most cases the head of the Department.

- *Departments of Danish Universities*
- *Public Institutes for scientific research*
- *Danish Ministries (for policy advice or policy analysis)*
- *Danish research organizations as a part of a non-profit foundation*
- *Non-governmental organisations*
- *Consultancy firms*
- *Enterprises. However, single enterprises cannot have access to microdata with enterprise data*

DST evaluates the application carefully and makes a concrete assessment of the applicant's reliability as a data recipient (e.g. ownership, educational standard among the staff and the research experience). If the institution is approved for access to microdata an authorization contract between Statistics Denmark and the research institution/environment is signed by the National Statistician and the responsible research manager at the institution (For contract please consult Annex B5 – 12a).

The most important point in the authorization contract is that the person signing the authorization agreement on behalf of the research/analysis environment undertakes the role of personally supervising that the confidentiality of data is always ensured. The researcher may not, directly or indirectly, download the dataset or any datasets derived from the datasets. Furthermore, individual records may not be printed, and all outputs 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. Unlike many other countries the authorization in Denmark is usually done at Department level and not University level. The rationale behind this is that researchers working under the authorized institution should be personally known to the responsible manger of the institution.

In case of breach the researcher will be excluded from using any research schemes of Statistics Denmark for a period of time or permanently excluded. Furthermore, in the case of breach hereof, the authorisation will be withdrawn from the responsible institution for a period, meaning that all researchers from that specific institution will be excluded for a period of time. This is done since the person signing the authorization agreement has not fulfilled the role of personally supervising the confidentiality rules stated in te agreement. In total DST has presently recognized and authorized ~500 institutions.

The Netherlands

At the CBS access to microdata can be granted to researchers and analysts from research environments pre-approved by director-general of the CBS. The purpose of the application is to check whether the applicant complies with the criteria for access to microdata at CBS Two important conditions are: that the primary mission of the applicant institution should be to conduct statistical or scientific research and that the results

of the research should be published. An online [Application form for organisations](#) can be found on the CBS webpage.

Authorizations can be granted to the following organisations:

- *Dutch universities - under the Law on Higher Education and Scientific Research*
- *Institutes for scientific research*
- *Organisations for policy advice or policy analysis*
- *Statistical authorities in other EU countries*
- *Other research institutions authorised to work with the microdata*

For Governmental Institutions (Ministries, Autonomous Agencies, Municipalities etc.) it is a mandatory requirement that the research activities are concentrated in a ‘statistical enclave’ as a separated part of the institution with its own management. The authorisation is only granted to that part of the institution, while the head of the research department as well as the head of the overall institution (secretary general, director, city manager) are responsible for ensuring that the data are not used for purposes of regulations or enforcement.

In contrast to Denmark the authorization is done at the University level in the Netherlands and the authorized institution does not have any obligation to supervise the associated researchers in the confidentiality rules and disclosure control. At CBS, the prevention of disclosure entails a shared responsibility of Statistics Netherlands and of the individual researcher who has been granted permission to work with microdata whereas the authorized Institution does not have contractual obligations in regards to disclosure. At CBS 130 research institutions has been authorized and include 30 foreign researchers from 10 different countries

2.1.2 Project request

Denmark

In Denmark the next step in the process is to submit a project proposal using a word template provided by Statistics Denmark or alternatively downloaded from the internet. However, currently the application form is only available [on-line](#) in Danish and not in English. The project proposal covers the following paragraphs:

- *Description of purpose of the research project*
- *Population to be studied (Need to know)*
- *DST registers included in the study (Need to know)*
- *External data (survey data or data from external registers)*
- *Register year (Need to know)*
- *Who will have access to data (Names, contact information and title)*
- *End date of the project*

At the study visit DST stressed that in order to fulfil the legal requirements data is always provided according to the so-called “*need to know principle*”, implying that the researchers can only be granted access to the data needed to fulfil their research purpose. This applies at the level of population, registers and register year as well as at the variable level.

In order to assist the researchers in selecting the relevant registers for their project, a data catalogue with comprehensive descriptions of most of the registers and variables, including code lists, is available on-line (for further details please consult Section 2.4).

If researchers want to combine DST data with their own data or other external data not in the possession of DST, permission must also be granted by the [Danish Data Protection Agency](#) and by the data owner. For further information on the Danish Data Protection Agency and its role and responsibility please consult their webpage. It is the researcher's responsibility that the permission is granted. However, DST requests the researchers to provide copies of the relevant permissions. If a project only contains data from DST no approval from the Danish Data Protection Agency is needed since DST holds a general approval for these data when used for research purposes.

All research projects have to be approved by the Head of the Research Service Unit of DST.

Netherlands

As in Denmark, the CBS request that an application form is filled out for all research projects. The [form](#) is an excel form that can be downloaded from the internet.

The required fields cover the following paragraphs:

- *Facility*
 - *on site/remote access*
- *Project information*
 - *Project title*
 - *Purpose of project*
 - *Start date (First day of month)*
 - *Expiration date (Last day of month; max. 1 year after start date)*
- *Information on the contractor: (= institution that will execute the research)*
- *Information on the commissioner of the research (If commissioner and contractor are two different parties)*
- *Publication of results*
 - *Please note that a requirement for obtaining access to microdata is that the results of the project will be published. The researchers need to indicate how and when the results of the project will be published.*
- *Project description - Send as an addendum a project description, containing at least:*
 - *Title of the project*
 - *Goals and backgrounds of the project*
 - *The intended research method (globally)*
- *Billing address*

As in Denmark, researchers will only be granted access to the data which are needed for their research, the so-called “*need to know principle*”. This means that researchers can get access to the data which is required for the specified purpose. However in contrast to Denmark this only applies to register and register year and not when it comes to populations and variables, since access is given as “views” to full datasets and not as customized dataset. This means that access is always granted to full populations and full registers. However, in the Netherland each register usually contain fewer variables per register (approx. 10 variables per dataset) compared to the situation in Denmark (up to 100 variables in some datasets – rare thought).

At CBS the research proposal also assists in finding the right staff members from the subject units to participate at the intake meeting that are mandatory in the Dutch model.

2.1.3 Intake Meeting and data extraction protocol

Denmark

In Denmark researchers are invited to an “intake meeting” only if it is requested by the researcher or the research group. Intake meetings only takes place occasionally and primarily with first time users of the research services provided by DST. In those cases, the purpose of the meeting is primarily to give the researcher an overview of services provided, data available, an introduction to documentation systems and to the confidentiality rules that applies. In other cases intake meetings are requested by experienced users of the microdata services. In those cases staff members from the subject unit are often invited also since the researcher often request more detailed information that what is provided by the on-line documentation or can be provided by the staff of Research Services Unit that have a more general knowledge.

However, for all projects the Research Service Unit will, in cooperation with the researchers, prepare a detailed data extraction protocol based on the research proposal, including detailed lists of the variables to be included. In order to assist both the researchers and the Research Service Unit in selecting the relevant registers and variables, an interactive [data catalogue](#) linking to the [documentation system](#) can be found on the internet although only in Danish.

The Netherlands

For researchers that want to use the services for microdata access, an “intake interview” is mandatory for all research projects. At this meeting staff from both the CBS Microdata Service Unit as well as selected subject units will participate. Sometimes up to two or three intake meeting are needed in order to get full clarification of the project aim, data needed and conditions under which data access will be granted. At an intake meeting the following issues will be discussed:

- *What is the purpose of the research project?*
- *Which datasets are needed?*
- *Does the project contain external datasets?*
- *What are the do's and don'ts regarding CBS files in relation to the specific project? Usually CBS will invite statistical experts to take part in the interview. The CBS advice will be restricted to the use of the datasets only. Set-up and execution of the research project are entirely the researcher's responsibility.*
- *Which alternatives can CBS offer?*
- *Which persons will participate in the project?*
- *Will the project be executed using the on-Site facilities or the Remote Access solution?*
- *Is the availability of certain software packages essential for the execution of the project? If so, which software packages are needed?*

In order to assist in determining which datasets that are needed a [data catalogue](#) and [documentation](#) of CBS data files can be found on the internet although only in Dutch.

2.1.4 Confidentiality statement

In both countries, confidentiality statements have to be filled out by all researchers that are granted access to microdata. Basically the agreement states that all work on microdata must take place on the research server and no attempts to identify individuals or enterprises must take place.

Denmark

Data confidentiality is a central precondition for the existence of DST's Research Services. Thus at DST it is important to ensure that access to microdata is provided in such a way that both data security and data

confidentiality are guaranteed. The individual citizen and business enterprise have a claim that confidential data should be treated in the strictest confidence. The general rules governing the treatment of confidential data are laid down partly by the Danish Act on Processing of Personal Data and partly by [Statistics Denmark's policy concerning data confidentiality](#)

All datasets for which access is granted through Research Services are confidential, cf. subsection 3 of section 27 of the Danish Public Administration Act and section 152 of the Danish Criminal Code. Against this background, researchers are legally obliged – through researcher agreements (Appendix B5 – 12b) in writing with Statistics Denmark – to extract only general results and not data at individual level of any kind from the research server's located at DST. In Denmark it is considered a very serious breach of the agreement to attempt to identify individual persons – or to remove micro data from the secured environment. Any breach of these rules may exclude the researchers from using any research schemes of DST for a period of time or permanently. In worst cases, also the authorization of the researcher's institution will be withdrawn for a period of time. The confidentiality statement has to be signed by the researcher, the responsible manager of the institution as well as the head of the Research Service Unit. The exact wording of the statement can be found in Annex B5 – 12b.

The Netherlands

In the Netherlands the exact content of the confidentiality statement was not presented. However, the confidentiality statement must be signed by the researcher and his/her employer. As in Denmark the researcher will only gain access to the environment once the statement is signed.

In addition the confidentiality statement is also presented and explained at the mandatory initial intake meeting at the CBS. Also an on-line test must also be passed by the researcher before logging on to the secured environment for the first time.

2.1.5 Price calculation and financial contract

In both countries a price is calculated and a contract prepared and signed for all new projects before access is granted.

In both countries there are no cost associated with the data as such since collection and processing of data is already paid for as part of the National Statistical Institutions production program and obligations. However, the researcher has to pay for using services provided such as consulting, preparing datasets/dataviews, cost for servers, Software licenses, output control etc. In both countries all prices are standardized in order to avoid subjectivity in costs estimation as much as possible. In the Netherlands all cost associated with the project is billed directly to the researchers whereas in Denmark part of the costs are covered by a yearly grant from the Ministry of Higher Education and Science.

The most significant contrasts between the model in Denmark and the Netherlands is that in the Netherlands costs are mainly based on standard fees for services whereas in Denmark costs are based on estimated time for the services needed to provide the needed service, using standard equations. Other major differences are, that in the Netherlands the researchers are charged for output control and that no funding for the unit or the services is received from a governmental body.

Denmark

In order to facilitate register based research the Research Service Unit has since 2002 received a yearly grant from the Danish Agency for Science and Higher Education, an agency under the Ministry of Higher Education and Science. The yearly grant covers part of the costs for researcher from the public sector. Currently the grant covers 10 hours of consulting for all new projects from the public sector as well as the remote access system including servers and their maintenance, license cost for software and administration and finally output control.

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

In addition, yearly contracts for development projects with the aim of improving the research services and documentation are financed by grants from the Danish Agency for Science and Higher Education. For researchers from the public sector, the yearly grant covers part of the research costs. Currently the grant covers 10 hours of consulting for all new projects from the public sector as well as use of the remote access system including servers and their maintenance, license costs for software, administrative costs and finally output control.

For the private sector the cost of all consulting and other services provided must be covered by private research institutions themselves. As a consequence different price models are used in Denmark for the public and private sector respectively.

The price model is constructed so that it takes the following parameters into account:

- *Time needed to construct the population*
- *The number of register years*
- *The number of registers*
- *The number of variables.*

In case the need for consulting exceeds 10 hours for the public sector this is added to the price and in the case of large-scale projects using a large amount of disk-space additional cost will be charged on a quarterly basis based on the actual use of disk-space. Sending output from the research environment and output control is free of charge both for the private and public sector.

During the study visit a standardized price model built in Excel format was presented in details and afterward handed over to ICBS. The model is stored on the internal ICBS network. For ICBS staff members please click [here](#) to open the model.

The structure from a researcher's point of view is divided into the following major groups:

- **Services prior and during the project start-up**
All services are based on hourly rates for time needed by Statistics Denmark for providing the service in question. The current hourly rates are: (a) Public research institutions: DKR 1050 per hour (~ € 150), (b) Private research Institutions DKR 1650 (~ € 220).
 - **Consulting and meetings** (For researchers associated with a public institution the first 10 hours of consulting is free of charge due to the yearly grant from the Ministry after that the researchers are charged per hour. Researchers associated with a private institution pay for all consulting associated to the project. The researcher can choose to be charged based on actual use of consulting or by a 'fixed' price' that is estimated to approximately 20% of the time needed for constructing the datasets).
 - **Authorization** of new researcher institutions and researchers. (Free of charge).
 - Cost for **construction of population(s)**. (Based on hours needed. Varies from 1 hour as a minimum up to in rare cases more than 10 hours for very complex population(s)).
 - Costs per **customized dataset**. (Based on the number of datasets and register years. Price for extracting one register, for one year with only few variables is estimated to take 0.05 hour to extract. However, a minimum of 1 hour is always charged. There is a huge variation between projects. Estimated time can vary from the minimum of 1 hour to more than 40 hours in total - in case of huge projects).
 - Importing **external micro-dataset(s)**. (Based on estimated or actual time needed to check and encrypt data - usually between 1 to 2 hours).
 - **Setting up new users, creation of password, administration**. (Free of charge).

- Services during an on-going project
 - **Output checking** (Free of charge)
 - **Updates** and adding datasets to a research project (Based on number of datasets)
 - **Disk-space** (First 5 Gigabyte (GB) free of charge then DKR 15 (€ 2) per GB per quarter).
 - **Use of the remote access system** (Public Institutions – free of charge, Private Institutions DKR 180 (€ 25) per researcher per day (“log-on fee”).
Public as well as private Institutions have the options of buying their own servers hosted by Statistics Denmark as well as their own software. In those cases they do not pay for log-on fee or for storage
 - **Use of software** (Free of charge for public inst. and part of the log-on fee for private Inst.)
 - **Creation of new users, resetting passwords, IT consulting, administration** (Free)

The financial contract is prepared by the Research Service Unit based on a standard template built into Microsoft software CRM - Customer Relationship Management. It's an integrated, data-driven solution that support all interaction with customers since the CRM systems and applications are designed to manage and maintain customer relationships, track engagements such as e-mails, account information, responsible manager, delivery of data etc.

The Netherlands

As mentioned in the introduction to this section the model in the Netherlands is primarily built on standard tariffs for the services. The full tariff structure in the Netherlands can be found on page 9 of the [Service catalogue](#) (and Annex B5 – 27). A 30% of revenue to overhead is built into the price structure. The total target is: 1.4 million EURO

Below the main cost components are listed:

- Services prior to the project start-up
 - *Feasibility study of available data. (Hourly rate: € 100)*
- Services during the project start-up
 - *Basic start fee (€ 1600)*
 - *Costs per dataset topic (€ 160)*
 - *Authorization new researcher (€ 375)*
 - *Call out fee (€ 530)*
 - *Fingerprint card (€ 150)*
 - *Importing own micro-data (Between € 210-1300 depending on complexity)*
 - *Use of software (SAS € 560 per months, other software € 40-60 per months)*
- Services during an on-going research project
 - *Consulting (€ 80 per month per researcher and € 16 per month per dataset, for extensive questions € 100 per hour)*
 - *Output checking (€ 200 per output)*
 - *Adding datasets to a research project (€ 160 per dataset)*
 - *Extra storage capacity (€ 25 per month per 50 GB)*
- Remote Access facility services
 - *Installation Remote Access computer (€ 1500)*
 - *Service contract RA-PC (€165 per month)*

As in Denmark the Netherlands has administrative tools that support all interaction with customers including contracts, billing, e-mails, account information, responsible manager, delivery of data etc.

2.1.6 Secure remote access environment

Both the CBS and DST provide encrypted secure remote access to de-identified microdata placed on servers at the statistical institutions allowing researchers to work on data from their own computer through an encrypted tunnel through the Internet. The Dutch model is based on Citrix Terminal Server technology and the Danish model is based on Microsoft Terminal Server technology. In both cases micro data remains at a secure environment hosted and maintained by Statistical Institution (DST/CBS). An outline of the systems can be found in figure 2. As illustrated in figure 2 the Dutch system relies on virtual servers⁴ whereas the Danish system relies on physical servers.

The underlying principle of the remote access system in both the Dutch and the Danish model is that researchers can analyse the confidential data from their own computer through a secured internet connection to servers hosted by the National Statistical Institution. In the Danish model researchers gain access to customized copies de-identified ‘Secure-use microdata files’⁵ with project specific identification keys for all direct identifiers. In the Netherlands researchers are granted access to data views⁶ meaning that all projects use the same identification keys and that all researchers get access to full registers and to full populations. Attempts to use unique identification keys for each individual research project have failed in the Netherlands.

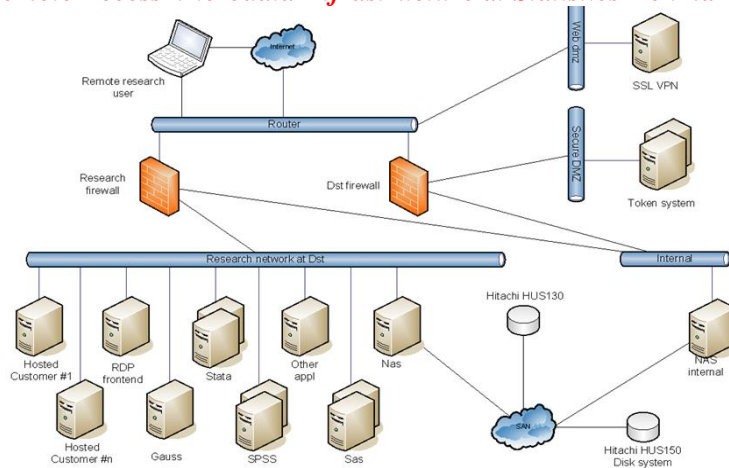
In both cases access is granted from a remote desktop and the analytic work is performed by using software provided by the National Statistical Institution. Further details on national solutions can be found in the sections following figure 2.

⁴ A virtual server is a server that shares hardware and software resources with other operating systems (OS), versus dedicated servers. Virtual servers are a popular way of providing low-cost services. Instead of requiring a separate computer for each server, dozens of virtual servers can co-reside on the same computer. In most cases, performance is not affected and each web site behaves as if it is being served by a dedicated server.

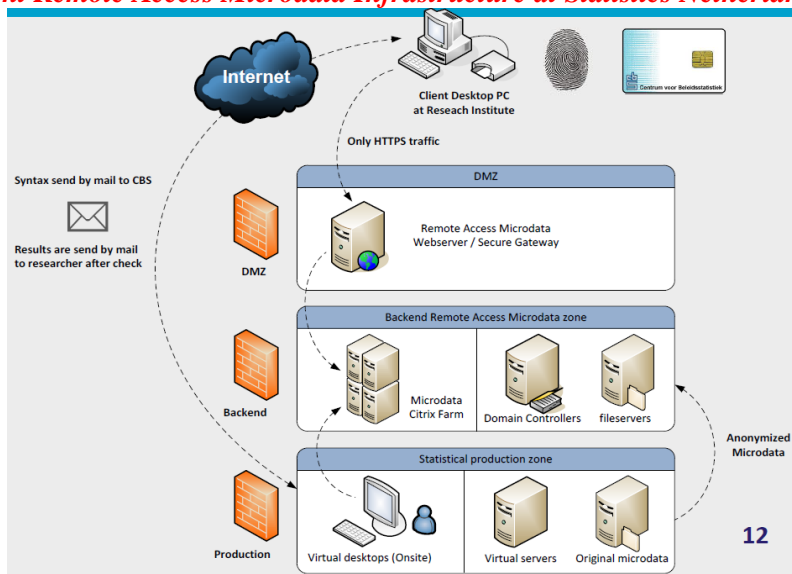
⁵ ‘Secure-use files’ meaning de-identified confidential data to which no further methods of statistical disclosure control (SDC) have been applied

⁶ Virtual View of Data. Unlike data files, data views do not contain actual data. Views minimize data replication and maintenance. After the Mission the RTA discussed with DST their experience of dataviews – an option that is used at DST for research databases. At DST restrictions of SAS dataviews in regards population and variables is easily be managed. DST has no experienced in using unique keys in individual projects for views but like the Netherlands their impression is that it will be extremely difficult/impossibly to maintain. However, experience from DST is that the first thing the researcher is doing after getting granted a view is transferring the views to datasets and place it in the project folder which means that disk-space is saved.

A: Remote Access Microdata Infrastructure at Statistics Denmark (DST)



B: Current Remote Access Microdata Infrastructure at Statistics Netherlands (CBS)



C: Future Remote Access Microdata Infrastructure at Statistics Netherlands (CBS)

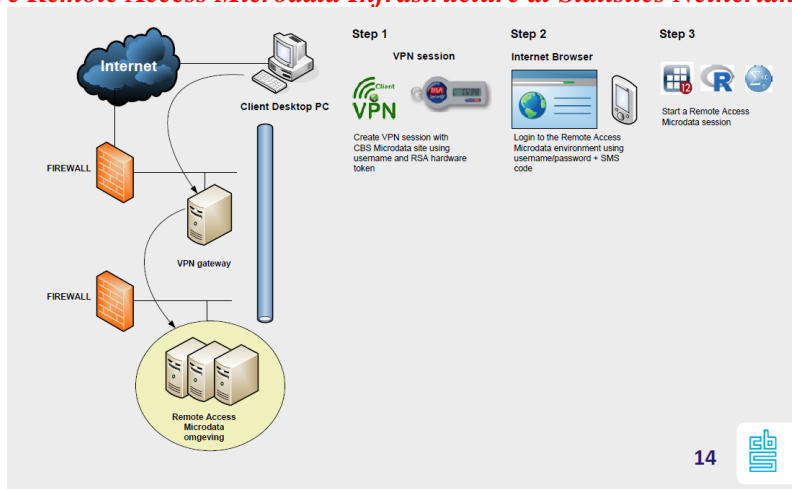


Figure 2: Current and future remote access infrastructure at (A) Statistics Denmark and (B, C) Statistics Netherlands respectively.

Denmark:

Network connections and encryption

The remote access system is based on a number of Microsoft Windows servers. The servers are placed within Statistics Denmark's environment. The remote access system is running on Windows Server 2008 Standard (64 bit). The servers make up a separate network isolated from the Statistics Denmark network area.

The clientless remote access F5 FirePass SSL VPN Solution⁷, based on tokens, is responsible for connection to the system from the Internet. The system's data communication passes along the normal Internet connections of Statistics Denmark and the traffic is encrypted between the user and the system with the HTTPS protocol. The users connect to the system by means of Microsoft Remote Desktop.

User authentication

The user logs on to the remote access system from the user's own workstation, using the Remote Desktop Connection. The user needs to log in to Internet Explorer, from which they are transferred to the TS Web Access service. The user enters their username (three letters) and password (four digit authentication number) after which the system prompts the user for a code (e.g. cWq2Xtt20Cc), which they receives by a text message to a pre-agreed mobile phone. The code can only be used once, and it is received as a flash message and will not be saved on the phone. Alternatively the researcher can read a user code from a personnel entrust token provided by Statistic Denmark – the code is changed every other minute. The first time the researcher log-on to the project the researcher is prompted to enter a new self-chosen password to be used for that specific project. In case the researchers at some point forget their password the password can be reset and re-send by DST.

All log-on to the research server are monitored and logged.

Managing access to data

The application servers share a common file server. Each research project has its own area on the file server, to which the members of the project are included. After the user has logged on to F5 Firepass with a user-specific Username, the user opens a Remote Desktop and logs on to a Frontend server with user- and project-specific credentials as mentioned above.

From the Frontend server the user logs on to one of the application servers. After this, the virtual desktop access opens up for the user and the Windows 2008 Server operating system stripped-down desktop view is displayed. All the necessary programs and folders have been stored as icons on the user's desktop. The user cannot manually edit this view and does not have access to other applications than those displayed on the desktop. File transfers from the user's different projects are blocked.

Each research project has a working directory into which the researchers, who have a permit to handle the project's data, are allowed to write. In the directory, users can set up their own work folders as well as share files with other users within the same project.

All data transfers into the remote access system are handled by personnel at the research services or the IT-support at Statistics Denmark.

⁷ Virtual private network (VPN) extends a private network across a public network, and enables users to work on data across shared or public networks as if their computing devices were directly connected to the private network. ("In the simplest terms, it creates a secure, encrypted connection, which can be thought of as a tunnel, between the researchers computer and a server operated by the VPN service hosted by the Statistical Institution. Applications running across the VPN may therefore benefit from the functionality, security, and management of the private network.

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

When the researcher stops working he can either disconnect from the system or log out of the session. When disconnected, the applications that were open still remain open. Thus, the user can leave long-lasting runs running in the system without having to keep the connection open. In addition, the connection is automatically terminated if the user is not active within a particular time period. When logging off, in turn, the applications that have been opened will close.

Storing files

The update and backup services are part of Statistics Denmark's normal maintenance routines. Backups of the contents of the file server are copied weekly.

Software and tools provided on the servers

| | |
|--|--|
| <i>Statistical package</i> | <i>Micro-soft and word processing programs</i> |
| <ul style="list-style-type: none">• SAS 9.4• STATA/MP 14.2 for Windows• IBM® SPSS® Statistics Version 22• WPS Workbench• Gauss• R• aML, version 2.09• GAMS• LatentGOLD®• Mplus Version 6.12 Demo• SCD/DIGRAM• SPSSToDigram• PLINK! v1.07 | <ul style="list-style-type: none">• Microsoft ®• GNU Emacs, version 24.3.1• gVim - VIM - Vi IMproved, version 7.1• Tinn-R, version 4.0.3.5• MiKTeX TeX• Adobe ® Reader, version 9.1.0 |
| | <i>Other programs</i> |
| | <ul style="list-style-type: none">• Stat/Transfer• WinZip® |
| | <i>Supporting tools provided on the servers</i> |
| | <ul style="list-style-type: none">• Access to documentation and metadata• Codelist (Format library) |

Management of the IT system

Please consult Annex B5 – 4 Comparative table for research services in Denmark and the Netherlands.

The Netherlands

Network connections and encryption

All servers are virtualized using VMware ESX 5.5 with the following features

- *Dedicated Remote Access Active Directory domain used for access control to applications and micro data*
- *Separated from regular statistical production environment*
- *Client workspace is installed by a CBS employee Software and a biometric smart card reader is installed*
- *Biometrics solution is custom build using the Webkey SDK from Bio-key*
- *Secure FTP is used for data transport (micro data and user imports/results) between the statistical production environment and the Remote Access environment.*
- *Remote Access farm is built on Citrix XenApp 6.5 and Windows 2008 R2*
 - *28 standard servers (4 CPU / 16 GB)*
 - *1 dedicated SAS server (4 CPU / 16 GB)*
 - *12 power servers (4 CPU / 64 GB)*

In the new system that CBS plan to release during the course of 2017 Remote Access Microdata environment will only be accessible through VPN with the followed features;

- *VPN uses 2-factor authentication: RSA hardware tokens*
- *For logging into the Microdata environment 2-factor authentication is used: RSA on-demand SMS tokens (smartphone researcher)*
- *No other internet access possible after setting up the VPN*
- *RA client PC does not have to be installed by CBS*
- *Less technical requirements for client PC (before it had to be 32-bit Windows + Internet Explorer)*
- *No more re-authentication*

User authentication

In the Netherlands authentication is currently done as 3-way authentication via fingerprint scan, smartcard with PKI certificates and username/password (Figure 2B). Fingerprints are stored centrally in CBS data centre. In contrast to Denmark that only require authentication each time the system is entered the Netherlands require re-authentication using random intervals (20-30 min)

Software and tools provided on the servers

| <i>Statistical package</i> | <i>Micro-soft and word processing programs</i> |
|--|--|
| <ul style="list-style-type: none"> • SPSS • STATA ? • R ? • Gauss • MLWin • Ox GiveWin • SAS Base | <ul style="list-style-type: none"> • Microsoft ® • ? <p><i>Other programs</i></p> <ul style="list-style-type: none"> • ? <p><i>Supporting tools provided on the servers</i></p> <ul style="list-style-type: none"> • ? |

Management of the IT system

Please consult Annex B5 – 4 Comparative table for research services in Denmark and the Netherlands.

2.1.7 Data extraction/creation of data views

In order to be able to construct customized dataset the research service units in both countries maintain a data library. In Denmark the data library is constructed as a SAS library with direct identifier such as CPR numbers. However, before data are placed on the research server, data are always de-identified by replacing all direct identifiers by project specific random numbers allowing merging of dataset within a specific project but not across projects. At the CBS the Data Service Centre maintain a data library as ASCII files with de-identified keys. Users of STATA and SAS can save converted catalogue files in specific directory than the shared workspace on the remote access network. Access to data is given to the researchers as views to full datasets with the same de-identification key in all projects. As mentioned earlier this means that all projects has same identification keys, with the consequence that a specific person will have the same ID number across all projects.

2.1.8 Analysis

In both the models applied by the CBS and DST a secured working environment is provided for the researcher in which they can store draft versions, syntaxes and outputs. Several software packages are

available on the research server, such as SAS, SPSS, STATA, GAUSS and R. In the Netherlands extra cost are associated with the use of SAS, Ox, Gauss, MLWin due to the high license costs for these programs. The programs are frequently updated with new versions in both countries. Aggregated results from the researcher's analyses can, in both models, be exported to the researchers own computer as long as they keep the rules of confidentiality in each of the countries. In order to ensure that the confidentiality rules are kept output control is done in both countries but in different manner. In Denmark the researcher receive their input instantly whereas in the Netherlands output is checked by CBS before they receive their input, usually within two to three days. For more details please consult section 2.1.9 below.

2.1.9 Output control

Output control is performed in both countries. In Denmark approximately 1000 outputs are created per day and in the Netherlands about 5-10 outputs are created per day. Both countries have limitations of file type that can be mailed home. In Denmark output control it is free of charge for the researcher since the cost is covered by a yearly grant from the Danish Agency for Science and Higher Education. CBS charge the researcher € 200 for a normal output but also have the option of output light that is cheaper and will be controlled faster.

Denmark

In Denmark outputs are not checked prior to export to the researchers own computer, but all outputs sent home by the researchers are logged and stored by DST. A random sample of the exported outputs is checked by the Research Service Unit (about 10%). The legal responsibility of the outputs rests on the authorized institution and its researchers and the obligation of DST is solely to provide guidance on confidentiality. [Guidelines](#) for output can be found on the internet In the event of a data breach, the institutions as well as the researchers are sanctioned. Since the study visit DST has, in order to assist the researchers, implemented an automated warning system for cases where a researcher try to send home non-aggregated data. In such cases the process is automatically stopped before the files are send home by the researcher.

In the event of a data breach, the institutions as well as the researchers are administratively sanctioned by DST. This implies exclusion for at least a month for all researchers associated with the institutions and the research service unit will request a written explanation from the responsible manger of the institution as well as a plan for how to avoid such situation in the future. The written explanation will be presented to the National Statistician and re-opening of access can only occur on his approval. For detailed information and rationale behind the sanctions please consult Annex B5 – 16 Output guidelines.

The Netherlands

In the Dutch model all outputs are checked independently by at least two controllers from the Research Service Unit of the CBS. In case the researchers outputs cannot be accepted, the CBS will contact the researcher in order make sure that they understand the rules for outputs. CBS distinguish between "output light" that is usually approved on the same day and normal outputs that are approved within two working days.

The CBS provide extensive output [Guidelines](#). In addition CBS require that all researchers pass an online test the first time they enter a new research project in order to ensure that they understand the rules for outputs. In addition questions about confidentiality will automatically pop-up at random intervals every time the researcher is logged on to their project (approximately every 30 minutes). Finally confidentiality rules are also discussed and explained at the intake meeting.

Each output need to be accompanied by a description of the output using a template provided by CBS. For standard form please consult the output [Guidelines](#) Appendix 2: Standard form for output verification (page 12).

In brief the researcher needs to fill in the following information:

- *Name of researcher*
- *Name institution or organization*
- *Date on which the output is produced*
- *Project number and name of the project*
- *Name of data files used to produce the output*
- *Description of the purpose of the analysis*
- *Relation with earlier output*
- *Content of the output*

In case a researchers output is not in alignment with the confidentiality rules the researcher is contacted in order to make sure they get a full understanding of the confidentiality rules.

2.1.10 Publications

In the Netherlands it is a requirement for obtaining access to microdata that the results of the project will be published for the public and for each project the principal researcher needs to indicate how and when the results of the project will be published.

Denmark

In Denmark there is no such requirement but researchers are encouraged to send their publications to the Research Service Unit, however, it is only rarely done.

The Netherlands

Within four weeks from the date of publication, CBS should receive the references of the publication. These publications must quote CBS as the source. CBS keeps a [list of publications](#) based on CBS microdata.

2.2 Organizational framework

2.2.1 Internal organisation within the National Statistical Institution

In both the Danish and the Dutch model a specialized Research Service Unit as well as a specialized IT section has been created in order to professionalize and standardize the service provided. The advantage of this central organization is that the individual researcher is fully aware of whom to negotiate with and who that is responsible for the dataset supplied.

Currently DST provides services to ~500 authorized institutions, 2000 researchers and 1200 research projects. The Research service unit consists of the Head of the Unit as well as 20 academics (generalists), 1 IT-specialists, 1 secretary and the IT service desk has 2 employees. At the CBS services are provided to 130 authorized institutions, 530 researchers and 700 projects. The services are provided by 16 staff members of whom 7 have an academic degree. The Data Service Centre has 8 employees. In order to match these ratios CBS will need 3-4 full time staff members to provide services to the research community (ICBS 75 projects and 100 researchers).

In both countries the scheme also requires close cooperation between the Research Services Unit and the producers of data – the subject units and IT. Both countries have service level agreements between in house partners. Concerning the cooperation with subject units a good relationship and clear agreements are essential since they produced the basis for the data to be included in the data repository/ Data service centre/ data library including metadata and documentation used by the researchers.

The internal workflow was only presented in details in Denmark and their internal workflow for production of data as well as consultancy can be seen in figure 3. However, as mentioned earlier the subject unit always take part in the intake meeting at CBS and as at DST the subject unit does also produce the metadata and documentation.

Internal workflows

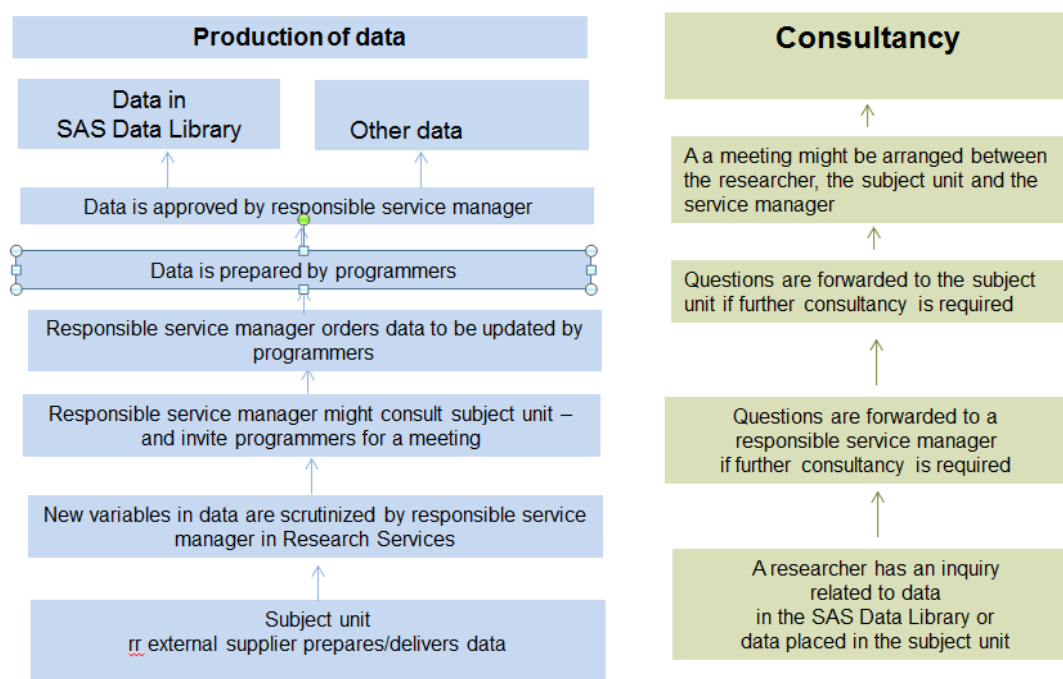


Figure 3: Internal workflow at Statistics Denmark as presented at the study visit.

2.2.2 Organisation of the Research service unit

In both the Netherlands and Denmark there is a division of work in the Research Service Unit. At DST the division of labour is thematic whereas the division at CBS is based on functions. The different approach means that in Denmark the researcher will always only have one contact person throughout the project with special knowledge in the issues of relevance for the customer type whereas in the Netherlands the researcher will have changing contact person through the project with specialist knowledge in a specific task.

Denmark

At DST, there is a clear internal thematic division of work in the Unit of Microdata Services as the work is divided into the following teams (Please be aware that employees can be part of multiple teams):

- Administration (1 employ)
- The Team for services to Ministries (3 employees)
- The Team for services to Public Research Institutions (9 employees)
- The Team for services to Private Research Institutions (4 employees)
- The Team for project with health data (7 employees)
- The Team for project with Geodata (3 employees)
- IT-support – Service desk (2 employees)

The Netherlands

In CBS, there is a clear internal, functional division of work in Microdataservice as the work is divided into four teams:

- Documentation team (about 4 people) - Quality assures project-specific documentation in collaboration with teams that provide the statistical documentation
- Account management team (approximately 6 people) - responsible for setting up projects, meetings with researchers, etc.
- Administration team (about 4 people) - responsible for drawing up contracts, invoices etc.
- IT (1 person) - Responsible for various IT problems, logon, etc.

See also the organizational chart below:

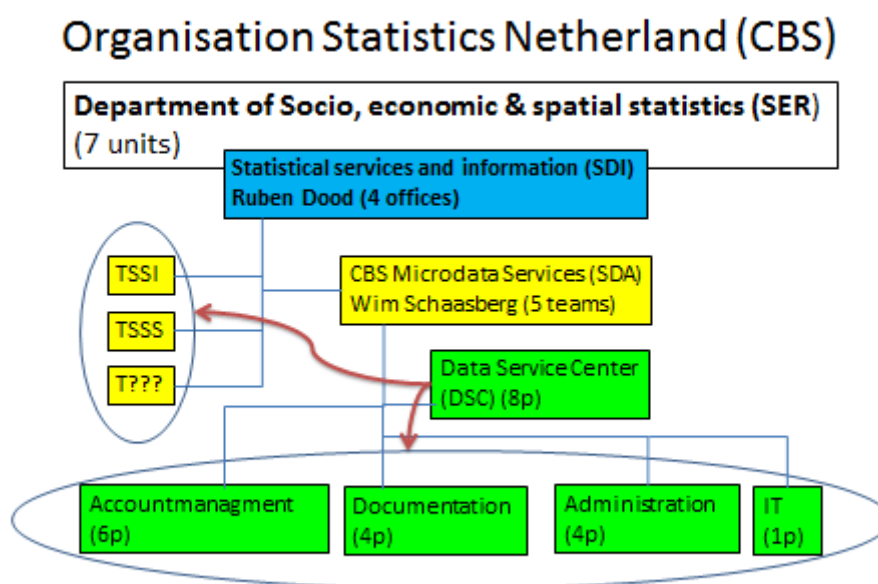


Figure 4: Organizational structure in the Research Service unit at CBS.

2.2.3 Advisory council for research services

Both CBS and DST have regular meetings with a representative group of microdata users: the customer council which serves as a sounding board and a consultation platform. The advisory body discuss the possibilities of and conditions which apply to working with microdata.

In both countries the Council is composed of around ten people who are associated with institutions which use the microdata services. These members represent the interests of the research community in the broad sense. In both countries the council meets at least twice a year.

2.3 Legislation, confidentiality and security

The legislation concerning confidentiality and protection of an individual's integrity is the foundation for how access to data for research is processed by the European National Statistical Institutions. The basic

European legislation currently covering data protection within the EU is the Data Protection Directive 95/46/EC. In general, the laws and regulations of relevance for researchers' access to microdata in the European countries are the implementations of this EU Directive. However, national differences as to how the EU Directive is implemented in national laws and practices exist. In addition, other national legislation, e.g. Statistical Acts and/or acts about public information and secrecy have to be complied with. The Data Protection Directive (European Parliament and Council Directive 95/46/EC constitutes the foundation of data protection rules in the EU and in the individual Member States. The aim of the Directive is to remove the obstacles of the free flows of personal data within the Community, cf. Article 1 §2. However, at the same time it is also essential to protect the fundamental rights and freedoms with regard to the processing of personal data, cf. Article 1 § 1. Another important point in the Directive is that according to article 6: Personal data can only be processed for specified explicit and legitimate purposes and may not be processed further in a way incompatible with those purposes.

According to the European Parliament and Council Directive 95/46/EC Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life (sensitive data).

According to the Directive each Member State must set up a supervisory authority, an independent body that will monitor the data protection level in the Member State, give advice to the government about administrative measures and regulations, and start legal proceedings when the Data Protection Regulation has been violated (art. 28). The data controller must notify the supervisory authority before processing of data is initiated.

The EU Data protection legislation is currently under revision and Directive 95/46/EC) from 1995 will by 25 May 2018 be replaced by the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) by which the European Parliament, the Council of the European Union and the European Commission intend to strengthen and unify data protection for all individuals within the European Union (EU). It also addresses the export of personal data outside the EU. The primary objectives of the GDPR are to give control back to citizens and residents over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU. When the GDPR takes effect, it will replace the data protection directive (officially Directive 95/46/EC) from 1995. The regulation was adopted on 27 April 2016. It applies from 25 May 2018 after a two-year transition period and, unlike a directive, it does not require any enabling legislation to be passed by national governments.

Implementation of the Data Protection Directive at the National level

All EU Member States have national implementations of the Data Protection Directive, which follows the principles of the EU Directive. Consequently, all EU Member States countries have the same level of minimum security regarding protection of personal information.

Denmark

The EU Directive 95/46/EC is nationally implemented in the personal data regulation which contains provisions on the processing of personal data. In summary, the National Personal Data Act defines, among others, the general principles on the processing of personal data in such a way that the legal rights of the individual citizen with regard to the protection and integrity are not violated. Furthermore, general rules for processing of personal data for special purposes, e.g. research, statistics are described, as well as how to handle sensitive personal data, the use of the personal identity number, and the rights of the data subject.

The National Statistics Act is a general law governing the national statistical service of central government authorities. It is applied to the official statistics of central government agencies, institutions and bodies. In Denmark the Statistics Act is, however, of a more general character and does not explicitly specify rules for researchers' access to register data.

The Netherlands

In the Netherlands the following laws apply

- Law on Economic Statistics 1936
- Law on protection of personal data 2001
- Law on Statistics Netherlands 2014. The most important point in relation to providing microdata access for researcher being:
 - Free access to public administrative data
 - Statistical confidentiality as legal obligation
 - Microdata access for statistical and scientific research

2.3.1 Data confidentiality policy and interpretation of the “need to know principle”

In both countries researchers will only be given access to the data which are needed for their research, the so-called “*need to know principle*”. This means that researchers can get access only to the data required for the specified purpose of a given project. Accordingly, the researchers in their applications have to document a reasonable relationship between the requested datasets and the objective of the project. In Denmark the “need to know principle” applies at the level of population, registers and register year as well as at the variable level. In contrast to Denmark, in the Netherlands this only applies to the register and register year since access is always granted to full populations and registers. However, in the Netherlands each register usually contain fewer variables per register compared to Denmark.

2.4 Supporting tools and documentation

2.4.1 Data catalogue

In order to be able to construct customized dataset the research service unit at DST and the Data service centre (DSC) at CBS maintains a data library. The catalogue contains all data produced by the National Statistical Institutions as well as administrative data from other public agencies. In the Netherlands many of these datasets dates back to 1995 and in Denmark multiple datasets dates back to 1980’s, which make them very relevant for e.g. longitudinal research projects. The content of the registers also covers many topics e.g. labour market, sociology, health and business. Several national administrative data sets from other agencies are also included in the catalogue in both countries and the use of these data is ensured either by law or by bilateral agreements between the national agencies and DST and CBS.

Denmark

In summary the [catalogue](#) contains 260 unique files names, 4640 datasets and more than >12.000 unique variables. The catalogue is maintained as an interactive catalogues with hyperlinks to the documentation h at both register and variable level and is updated at a nightly basis.

The Netherlands

In summary the [catalogue](#) contains 441 unique files names and 3416 dataset. The catalogue is maintained as pdf files.

2.4.2 Data library to be used by the Research Service Unit

In order to be able to construct a customized dataset the research service units in both countries maintains a data library. The library catalogue contains all data listed in the catalogue (see above).

Denmark

In Denmark the data library is constructed and maintained by the research service unit as a SAS library with direct identifier such as CPR numbers. However, before data are placed on the research server data are always de-identified by replacing all direct identifiers by project specific random numbers allowing merging of dataset within a specific project but not across projects.

As a general rule data should be included in the data library for research and accessible for the researcher no later than one week from publication by DST for high priority datasets and no later than four weeks from publication or on demand for low priority data. Data will never enter the database before after data are officially released by DST.

The Netherlands

At the CBS the Data Service Centre maintain a data library with ASCII files in a SQL database where all direct identifiers are de-identified. The SQL database makes data available from most other software without any transformation. Access to data is given to the researchers as views to full datasets with the same de-identification key in all projects.

2.4.3 Metadata and documentation

In order for the researcher to understand and use the data correctly metadata is essential both for selection of data to be included in their project and as well for working with the data.

Denmark

In Denmark the production of metadata and documentation is the responsibility of the subject units and it is a policy at DST that no data can be published from DST without documentation of statistics and variables. The documentation is made available for internal users, to the public as well as on the research server on the day of publishing the data.

Documentation of Statistics

All statistical products published by DST are described in [Documentation of statistics](#) that can found on the internet as well as while working on the research servers and for internal DST employees in the documentation system Collectica. The documentation is based on the international standard SIMS (Single Integrated Metadata Structure). Altogether DST's production includes more than 250 statistics products and is documented both in Danish and English.

Documentation of variables

All variables contained in DST's statistics are documented in the documentation system [TIMES](#) (currently in the process of being transferred to Collectica). TIMES consist of a hierarchical constructed common database with associated variables description. TIMES documentation is available in Danish only.

High Quality Documentation

In addition a special documentation format particularly aimed at researcher was initiated in 2008, the so called [High Quality Documentation](#) due to a demand by the research community.

The new element in the High Quality Documentation is the inclusion of the historical dimension in the documentation for each variable. The historical dimension describes possible data rupture and validity periods and the population is specified for each variable. Where applicable a graph and a table are attached in order to illustrate the historical development. The High Quality Documentation is produced in corporation between researchers, subject units and the Research Service unit. What is also unique about this documentation is that it includes a review process by a prominent researcher within the subject area. The cost for the High Quality Documentation has since 2008 been covered by a yearly grant from the Danish Agency for Science and Higher Education.

As for the regular documentation the TIMES system is used for the High quality documentation, however, a few additional fields applied in the template in order to include the extra elements.

Codelist/format library

Codelists and format library are available on the research servers. Guidelines are provided for the researchers.

Netherlands

At the CBS the Data Service Centre maintains the [documentation](#) for data available to researchers. The documentation is available as pdf files and can be found both on the internet and is also provided for each project on the server.

3. Conclusions and recommendations

In this section the specific recommendations of the ICBS participant is presented together with lessons learned during the study visit and these do not necessarily correspond to the views of the EU, DST or the CBS. In order to give a better overview of the recommendations they are grouped into subsections.

3.1 Recommended basic steps for simplified procedures and responsibility

Before data access

1. Institutional authorization - Approved by the national Statistician
2. Project requests –Approved by the Director of the Research Service Unit (RS)
3. Intake Meeting – Mandatory for RR files and optional for MUC files. RS and Subject Units.
Outcome of the meeting a data extraction protocol. RS
 - i) Project extraction protocol approval – Approved by the Confidentiality Committee
 - ii) Price estimated and preparation of contract - RS
4. Signing contracts (Financial Contract and Confidentiality statement) – RS
5. Contract with researchers as sworn employ for RR – The legal Department
 - iii) Creating the research dataset (and related documentation) - RS
 - iv) Project environment created on the server - IT
 - v) Data and documentation transferred to the research server - RS

Working with data

6. Working with microdata on the server - RS and IT Consulting
7. Sending output for approval
 - vi) Output control – RS
8. Receiving output

Closing the project

- vii) Closing the project – IT
- viii) Publications mailed to the RS

3.2 Recommendation for organization

Establish a central Research Service Unit

- At DST and CBS each staff members at the central unit provide service to approximately 90 and 33 research projects and 50 and 40 researchers. In order to match these ratios CBS will need 3-4 full time staff members (ICBS 75 projects and 100 researchers)

Establishment of an IT service desk

Establish a permanent advisory committee for research services

Describe the tasks and responsibility including service level agreements

- Subject matter units (keywords: data, code list, documentation, consulting)
- IT (keywords: Service-desk, management and monitoring)
- The Confidentiality Committee (keywords: Overall policy and rules e.g. on 'need to know', output etc.)
- The Computer Security Office (keywords: Security standards, LogPoint and Log consolidation)
- The Legal Department (keywords: revision of law, input to approval process)
- Central Research Service unit (keywords: Responsible for a Central Research Service unit, responsible for the overall policy, responsible for development projects, responsible for the overall business plan, daily management and coordination with other units and committees, administration, newsletters for the research community, consulting, subtraction of data, update of the webpage, courses for researchers e.g. on output control etc.)
- The Public Council for Statistics (Policy)
- advisory committee for research services (Guidance)

3.3 Recommendation for data and documentation

Data catalogue

- Policy and procedures for what should be included in the data catalogue (Separate catalogues for MUC and RR?)
- Build a data catalogue of available files and variables for researchers, including standardization of names and formats. Overall division by subject area (include ICBS data - Surveys, census, admin data from other national agencies)
- Make catalogue available for the researchers (Internet)

Data Library for RS

- Create a data library with no direct identifier (SAS, SPSS, ASCII??)
- Standardize file and variable names according to the catalogue and the documentation
- Maintained by the Unit for research services
- Service level agreements with SU and IT
- Set up a policy for access

Documentation

- Describe format for documentation – feedback from researchers
- Documentation - step by step based on frequency of use
- Service level agreements with SU
- Make documentation available for the researchers (Internet)

3.4 Recommendation for data for researchers

Need to know

- Should apply to population, registers, variables and year

Data for researchers

- Access to MUC files from work/home via secured internet connection (remote access)
- Access to research room files from secure RR
- Outface MUC files
- Policy and procedures for quality check
- Data delivered as separate datasets
- Project specific keys
- Decide how data should be delivered to delivered as SPSS dataset

Guidelines and supporting tools

- Service catalogue for researchers
- Price structure
- Templates for the approval process
- Booking system for RR
- Status of projects
- Thematic courses

3.5 Recommendation for output control

Output control

- All outputs must be checked before they are mailed to the researchers (1 or 2 controllers)
- Online test for output
- Courses for output (CBS offer five courses per year - Denmark on demand)

Guidelines and supporting tools

- Procedures and processes (Internal)
- Service catalogue for researchers (External)
- Confidentiality policy (External)
- To update and complete guidelines on output (Internal and external)

3.6 Recommendation for IT and administration

Administration

- Building an administrative management system that keeps track of all institutions, researchers and research projects.
- Keep it simple and build it step by step

Technological infrastructures (IT)

- In order to optimize management, data security and flexibility, one technological infrastructures solution for researcher access to microdata should be developed.
- The solution should support both access modes.
- Applications and data will be located on a virtual server at ICBS in Jerusalem.
- Hardware in research rooms should be ‘thin clients’ acting as ‘terminals’ connecting to a research network.
- Decide which program should be available

Business plan

- Development of a more transparent and accurate price structure (External funding, funding by researchers, funding from ICBS)
- Don't forget funding needed for maintenance and development

Legislation

- Initiate steps to change the legal framework by which access to data is provided to researchers (Access to RR files via remote access, access to admin files).

3.7 Next steps

- Finalize the strategic plan for providing researchers with access to micro-data for approval by the management
- Make a draft roadmap for implementing the strategic plan including (SMART)
- Keep up the good work and cooperation between:
 - *Subject matter units*
 - *IT*
 - *The Confidentially Committee*
 - *The Computer Security Office*
 - *The Legal Department*
 - *The International Relations and Statistical Coordination*