Guidelines and Criteria for CIRRAU and ECONAU servers at Statistics Denmark

Allan Timmermann[1,2], Malene Thygesen[1,2], Ane Marie Closter[1,2], Susanne Jensen[3], Jonas Mikkelsen[3], Maria Knoth Humlum[3], Carsten Bøcker Pedersen[1,2]

[1] Centre for Integrated Register-based Research (CIRRAU), Aarhus University, Aarhus, Denmark

[2] National Centre for Register-Based Research, Aarhus University. Fuglesangs Allé 4, 8210 Aarhus V, Denmark

[3] Department of Economics and Business Economics, Aarhus University. Fuglesangs Allé 4, 8210 Aarhus V, Denmark

CONTENT

1. Introduction

This document describes how to access and use the CIRRAU and ECONAU servers at Statistics Denmark (DST) and provides guidelines to best practice for responsible use of these common computer resources as well as guidelines for working with microdata and data security.

The Danish population-based registers is an important research tool, enabling Danish researchers to carry out representative population-based studies on, for example, the potential clustering of disease in families, and effect of parental income dynamics on health outcomes, and many more. These data constitute an important and rare asset. Similar data are only currently available in Sweden, Finland and Taiwan, with varying differences in coverage periods and the extent of other available data that could be linked to the population data. The Danish population-based registers will in connection with specific cohorts, intervention studies, and biobanks continue to provide the basis for significant knowledge relevant to the aetiological understanding and possible prevention of human diseases.

To access these unique data-sources including personal information on the entire Danish population, we must respect legislation and confidentiality while also ensuring flexible access to these valuable resourced. Our focus is to ensure that personal data are used in compliance with GDPR and henceforth ISO27001, and within these constraints ensure a flexible access to these valuable data-sources.

The solution offered by CIRRAU and ECONAU consists of VPN access to a server located at Statistics Denmark, which contains a number of specific research projects in which the data needed for each project is accessible. Researchers can access projects on the server through a personal computer. Downloading data is neither possible, nor permitted, and thus all data processing must be performed on the server. A challenge when applying such a set-up is that all researchers share computer resources, including CPU, virtual memory and disk storage. Given the large size of many population-based registers, a single researcher with an (possibly inadvertently) unfortunate practice can affect negatively on the possibility to perform research for all other researchers. Therefore, please read and comply with these guidelines for the benefit of all users. Users that repeatedly do not comply with guidelines will have their access to data terminated.

Most datasets on the server contain microdata, i.e. individual level data on persons (personal data) and data on single companies, firms or institutions available from national registers or other resources. When conducting research on microdata, you must ensure that no microdata information is transferred to any unauthorized persons. Researchers may access data for the approved research only and must never reveal any microdata information to anyone outside the project. This is by far the most important criterion for any register-based study of individuals, companies, firms or institutions.

A CIRRAU or ECONAU research project is an independent unit where access to data can be granted upon approval from the relevant authorities. Data must be used only within the boundaries described in the project description. Data from one project cannot be used in another project without approval from the relevant authorities. These typically include the Danish Data Protection Agency, the Danish Health Data Authority, and Statistics Denmark.

Researchers employed at Aarhus University, i.e. CVR-no: 31119103, or PhD students enrolled at Aarhus University, may gain access to data for which Aarhus University is Data Controller. Researchers with dual

employments for which one of them is at Aarhus University may access personal data only during time-periods at which they work for Aarhus University. To document employment at Aarhus University, CIRRAU and ECONAU may request and store a copy of the letter of employment from Aarhus University. This letter should include name and time period of employment, but must be blinded regarding CPR-number and salary. Researchers not employed at Aarhus University need a collaborative agreement and either a data processor agreement or likewise.

The CIRRAU and ECONAU project databases were established as a long-term data resource from which researchers may access interdisciplinary data capabilities. All data that are already included in the CIRRAU or ECONAU project database can be provided free of charge to associated researchers in specific research projects, whereas additional data must be paid for by the researchers' own funds. When data are needed from other sources, Statistics Denmark charges a fee for adding the data to the project on the CIRRAU or ECONAU server. This fee is also paid by the researcher.

To get access to an existing CIRRAU or ECONAU project, please contact the project's leading researcher. To set up a new project please see:
CIRRAU: http://cirrau.au.dk/
ECONAU: https://econ.medarbejdere.au.dk/research/project-databases/econau/

All researchers with access to a project are required to notify CIRRAU or ECONAU of any changes in workplace, job position, email and mobile phone number; please find contact information at the above webpages.

For research papers based on CIRRAU or ECONAU data, we require that at least one author states the relevant affiliation. The authors decide among themselves, who and how many authors should state affiliation. Please use one of the following affiliations:
For CIRRAU: Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark
For ECONAU: Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark

2. The servers

CIRRAU/ECONAU is able to grant a personal VPN access to a server located at Statistics Denmark for each researcher. This server contains the relevant personally identifiable data. The server is accessed through the researcher's own computer. Downloading and copying personally identifiable information to the researcher's own computer is neither possible nor permitted. The server is managed by Statistics Denmark on behalf of CIRRAU/ECONAU, using strict security measures which, among others, preclude users from downloading information, modifying security settings, installing and modifying system and software. All personal data have been deidentified, but data are, according to Danish legislation, still considered as personal data.

Each research project works independently of other projects on the CIRRAU or ECONAU servers. One, and only one, research institution is responsible for the project, and each research institution uses a unique Statistics Denmark authorization.

On **CIRRAU** servers, the authorized institution can be:

- National Centre for Register-based Research, Aarhus University - authorization no. 9
- Department of Political Science, Aarhus University - authorization no. 24
- Department of Economics & Business Economics, Aarhus University - authorization no. 43
- Department of Public Health, Aarhus University - authorization no. 50
- Department of Psychology, Aarhus University - authorization no. 105

Projects on the **ECONAU** server is authorized by:

- Department of Economics & Business Economics, Aarhus University - authorization no. 43.

Please note that a user is always subject to the specific rules set by the authorized institution to which the project belongs.

CIRRAU has several hosted servers at Statistics Denmark, for example: Srvfsencrr6, Srvfsencrr8, Srvfsencrr11, etc.
ECONAU has two hosted servers at Statistics Denmark, named: Srvfseeconau2 and Srvfseeconau3.

Each server contains specific research projects, which are named using six digits e.g. 703373, and data are placed in a project according to the permissions of the project.

Access to the servers:

Please see the document "Forbindelse til forskerservere I Danmarks Statistik" at https://www.dst.dk/da/TilSalg/Forskningsservice/Vejledninger for a guide on how to access the remote desktop at Statistics Denmark once you have received your id and password.

Software and configuration:

The three statistical software applications: R, SAS and STATA are installed on all servers, as well as standard applications such as Word, Excel etc. Additional specialized software is available on some servers. The configuration can vary slightly between servers. However, data files are located on either the D:\ , E:\ , G:\, X:\ or Y:\ drive, which are common disk areas containing the folders: Rawdata and Workdata as well as the personal folder: Kode on some servers.

The Rawdata folder:

Example path: D:\Data\Rawdata\<project number>
Read only, rawdata files are stored here. Users cannot make any changes to data or write new files. Your syntax/code should read data from this location.

The Workdata folder:

Example path: D:\Data\Workdata\<project number>
The Workdata folder is the common workspace on the project. All researchers with access to the project can read and write files. Save your syntax/code, derived datasets and results here. When you get access to a project, the first thing you should do is to create a new folder in the workdata folder for the purpose above. Give the folder a name that corresponds to your researcher ID. Example: your researcher ID is ABCD and

you just got access to the project 703301, you create the folder “ABCD” with the following file path: G:\Workdata\703301\ABCD. On the ECONAU servers the folder is automatically generated and named with DST ID and last 4 digits of the project number.

The Kode folder

Example path: D:\Data\Kode\<your DST ID + last four digits of the project number>
Available on CIRRAU servers only. For example D:\Data\Kode\ABCD3373
This is a personal folder where you can save syntax/code that only you can access. Only code and short text documents are allowed to be saved in the Kode folder. Any other types of files, including data files, are not allowed to be saved in the Kode folder. For this purpose, use the Workdata folder instead (see above).

C:\ and F:\ drive

The C:\ drive is a system drive that contains the operating system and installed software. The F:\ drive contains a disk swap area for SAS and STATA temporary workfiles. It is important that you DO NOT under any circumstances save any files to the C:\ or F:\ drive. There is no back-up of files at C:\ and F:\ and, Statistics Denmark occasionally, deletes files from these drives.

General information about the servers

CIRRAU and ECONAU servers are restarted on the first of each month between 00.00 and 08.00. Please, refrain from running jobs at this time; otherwise, your work will probably be incomplete or lost.

Information specifically regarding CIRRAU server 6 and 8

The servers Srvfsencrr6 and srvfsencrr8 use the same disk area, but have separate CPU and memory. Projects located on srvfsencrr6 can login to srvfsencrr8. This setup enables a larger number of users to run analyses at the same time. To optimize usage, CIRRAU requires researchers that primarily use R or STATA to login to srvfsencrr8 and SAS users to login to srvfsencrr6. Server srvfsencrr8 has the most recent updated version of STATA and R (including all packages from CRAN, updated approximately once a year).

Information specifically regarding ECONAU server 2 and 3

The servers Srvfseeconau2 and Srvfseeconau3 have separate CPU and memory. Projects located on srvfseeconau2 can login to srvfseeconau3 and vice versa. This setup enables a larger number of users to run analyses at the same time. MATLAB is only installed on srvfseeconau2 while srvfseeconau3 has the most recent updated version of STATA. ECONAU recommends users to work primarily at srvfseeconau3 so that srvfseeconau2 may be used to run larger jobs (please contact ECONAU data managers if you need a slot to run bigger jobs).

Statistics Denmark formats

Statistics Denmark provides several SAS formats, which allow different categorizations of many variables, e.g. different levels of education and the groupings used in Statbank Denmark.
As an example, to access the formats in SAS you need to write the following syntax in your SAS editor:

```
libname fmt ‘\\srvfsenas3\formater\SAS formater I Danmarks Statistik\FORMATKATALOG’ access=read only;
Options fmtsearch=(fmt.times_personstatistik fmt.times_bbr fmt.times_erhvervsstatistik fmt.brancher
fmt.uddannelser fmt.geokoder fmt.disco fmt.statistikbank fmt.disced fmt.sundhed);
```

Statistics Denmark has set up guidelines on how to use the formats, see shortcut "DST-Formater" (looks like a star) on the desktop. Here you can also find datasets with most formats in STATA and .TXT format. Researchers are of course welcome to modify these formats to use in R, STATA etc.

CIRRAU formats for SAS

CIRRAU also provides formats in SAS (NCRR formats), which allow various categorizations of data e.g. country of birth. The location of formats varies slightly on different servers.
As an example, for srvfsencrr6, to access the formats in SAS you need to write the following syntax in your SAS editor:

```
libname format "C:\NCRR6_DS\SAS\SAS9.4\NCRRformat\sas9_3_2013" access=read only;
Options fmtsearch=(format.cpr format.demograf format.demografi format.grupper format.geo);
```

Batch submit SAS

To batch submit SAS, Locate the directory containing your SAS-program in MS-Dos prompt. Enter "sas94 <my sasfile>" or simply right click and choose batch submit.

Signing out of servers:

Click "start", click "your username" and click "sign out". Ensure that you sign out properly as soon as your jobs have finished. Otherwise, data stored in memory will use resources, impeding the response times of other users. If you have large jobs running that may require several hours to finish, you can disconnect from the server by closing the remote desktop or click "start", click "power options", click "disconnect". Make sure to login and sign out properly as soon as possible after your jobs have finished.

3. Data security

Security rules

Only persons employed at Aarhus University or at a relevant data processor and assigned to the project may gain access to data.

Data may not be sought extracted from the server at Statistics Denmark in any way using whatever media. This also includes screen dumps, photographs, manual transcript of the screen, video, facetime, Skype, or any other method.

When connected to the server at Statistics Denmark, the content of the screen may not be shown to persons who are not themselves granted access to the project.

When connected to the server at Statistics Denmark, the computer shall not be passed on to unauthorized persons.

The password for accessing the server at Statistics Denmark is strictly personal.

All descriptive analyses and regression modelling must be covered by the project description for the "Project" and must be necessary for the project.

Access to data at the server at Statistics Denmark is allowed only from Denmark or through a personal two factor VPN access to Aarhus University followed by personal two factor VPN access to Statistics Denmark. CIRRAU and ECONAU provides these VPNs for all users.

Access to data at the server at Statistics Denmark is allowed using a computer with fully operational and updated antivirus software. Aarhus University or the data processor must have implemented procedures ensuring all PC's have updated antivirus protection.

It is not allowed to access the server at Statistics Denmark from locations where there is any risk that any other third party may unintentionally see the content of the screen (e.g. public areas).

New users must participate in a course locally at CIRRAU or ECONAU on rules for accessing and working with microdata.

No attempts must be made to identify individual physical persons.

New users must sign a contract on data access with Statistics Denmark.

Violating data security is a very serious breach of the agreement between the researcher and Statistics Denmark. Non-compliance with the terms may exclude a researcher from access to data at Statistics Denmark temporary or permanently. In worst cases, Statistics Denmark excludes the entire research environment from the research servers for a period of a few to several months. For that reason, **make sure that you understand and comply with the Statistics Denmark "Terms of Agreement"** (see section 7).

Users must read and comply with Aarhus University's information security policy (https://medarbejdere.au.dk/en/informationsecurity/informationsecuritypolicy/) and CIRRAU and ECONAU guidelines (http://cirrau.au.dk/data-resources/users-guides/). Both documents must be read annually.

Users must inform CIRRAU or ECONAU immediately regarding changes in employment for associated users.

Users must inform CIRRAU or ECONAU immediately in case of the Data Processor's breach, anticipated breach, or any suspected or actual unauthorized use of sensitive data.

Data Governance Expert

New users working under the authorization of NCRR and ECON (DST authorization no. 9 and 43) may be granted access after supervised training and approval by a Data Governance Expert. This person is appointed by CIRRAU or ECONAU and supervises all issues related to accessing personally identifiable information on CIRRAU or ECONAU servers at Statistics Denmark. The Data Governance Expert adheres to the following criteria: a) Person with extensive experience in handling personal identifiable information at Statistics Denmark, and b) Employed at NCRR or ECON, Aarhus University, and has been granted access to transfer files from Statistics Denmark for a period of not less than 3 years, and c) The user and the Data Governance Expert must collectively ensure that all analyses are performed in compliance with the permissions for the project and that they are necessary for the project. The appointed Data Governance Expert ensures authentication of their associated new users.

CIRRAU and ECONAU impose the rule that all new users initially cannot transfer any files from the servers. Extraction of results from the system is possible by contacting the appointed Data Governance Expert. Results are information intended for publication, which do not contain microdata. Upon each extraction of data, the researcher must confirm:

a. "I ask permission to export the following result file(s) from the CIRRAU or ECONAU secure partition at Statistics Denmark at the convenience of CIRRAU or ECONAU. I am fully informed of the rules governing export of data from Statistics Denmark (https://www.dst.dk/da/TilSalg/Forskningsservice/Vejledninger), CIRRAU and ECONAU guidelines and criteria (http://cirrau.au.dk/data-resources/users-guides/) and the Danish law on personal data "Persondataforordningen", and I confirm that the results do neither contain microdata nor individual level data.

b. The location of results to be exported: e.g. SrvfsencrrX-D:\Data\Workdata\<file location>

c. I agree and accept that in case any of the requested files do not comply with the guidelines, my possibility to export files from Statistics Denmark will be closed for a period of three months. Subsequent non-compliance will terminate my access to data for a period of not less than 3 months"

4. Microdata

Most data in CIRRAU and ECONAU projects consists of microdata, which is data concerning individuals, single companies, firms or institutions. **All microdata must be treated as confidential information and must remain on the secure servers at Statistics Denmark**. Even though all identifiers such as for example CPR or CVR numbers have been de-identified (replaced by scrambled identifiers), data is still microdata and may not be transferred out from the server. Even if you delete identifying variables such as the de-identified CPR number it is still microdata and may not be transferred. **If the file you want to transfer contains individual observations NO MATTER what the variables contain, it is NOT allowed.**

Researchers on CIRRAU and ECONAU projects are obligated to treat all data as confidential information in accordance with the terms and conditions of the Danish Act on Processing of Personal Data. Confidential information is defined according to GDPR (Persondataforordningen), the Danish Health Data Authority and Statistics Denmark's combined criteria. It applies to any information that relates to less than 5 identifiable physical persons, companies, firms, institutions or other units with an identification number (e.g. households or families). This means that **tables must contain at least 5 units per cell, and that all statistics must be based on groups of at least 5 cases.** For business statistics, an additional confidentiality rule is applied known as the **dominance criterion**. This implies that if the largest or the two largest enterprises in a table cell showing an economic variable amount to a dominant share, i.e. more than 85 percent of total revenue, the dominance criterion will subsequently apply, and information is considered confidential. For employment data the dominance criteria apply to statistics measuring a volume. Here the dominant share is based more than 85 percent of fulltime employed.

5. Transferring files

Getting results out from the servers

Only aggregated results may be transferred from the secure servers at Statistics Denmark. Files containing aggregated results from analyses can be transferred to the email of a researcher. It is of great importance that the researcher has made sure that the files do not contain microdata information. Statistics Denmark saves all transferred files for six months and randomly conduct inspection of files to make sure that users comply with the rules. If rules are violated, the penalty ranges from a personal warning to a permanent lock-out of all users on all the institution's projects at Statistics Denmark. Therefore, make sure that no microdata are transferred. It is only possible to send files smaller than 3,000 KB and compressed files cannot be transferred. For more details on how to send out results, please consult the guidelines at Statistics Denmark's homepage https://www.dst.dk/da/TilSalg/Forskningsservice/hjemtagelse-af-analyseresultater

What CAN be transferred from Statistics Denmark?

- Results of analyses
- Aggregated tables that obey the rules
- Aggregated figures that obey the rules

Only aggregated tables, aggregated figures or aggregated statistics intended for publication in which it is not possible to identify e.g. individuals, households, families, firms or other units with an identification number are allowed to be transferred.

Tables should have at least five observations in a cell. However, all transfers should be evaluated on an individual basis. Having five observations in a cell does not automatically justify that the information is not personally identifiable. In business statistics you also need to take into consideration if the two largest companies in a cell account for more than 85% of the total revenue in the cell. If this is the case you need to have more observations in this particular cell.

When transferring multiple tables, make sure that it is NOT possible to identify e.g. individuals by combining two or more of the tables that you are transferring.

Exact medians, minimums, maximums or percentiles can identify individuals. Make sure to truncate values so they represent five (5) or more individuals. Also, be aware that it is possible to identify individuals from outliers in a figure. Therefore, be very careful not to have single data points in figures.

All output must be manually checked before transferred out. Transferring uncontrolled output is not allowed and considered a violation of the security rules. Users must know exactly what they are transferring. Statistics Denmark randomly inspects transferred files. If security rules are violated, the penalty ranges from a personal warning to a permanent lockout of the whole research environment (https://www.dst.dk/ext/3477468153/0/forskning/Guidelines-for-transferring-aggregated-results-from-Statistics-Denmark--pdf). See also chapter 8 describing sanctions.

To avoid unintended breaches of confidentiality, transferring of log-files, syntax/code and other types of output are not allowed. The reason for this decision is that the majority of earlier breaches of the security rules were unintended and caused by transferring e.g. log files with direct or in-direct personally

identifiable data. If publication of syntax/code is required by a journal, please contact a CIRRAU/ECONAU data manager about how to proceed.

What CANNOT be transferred from Statistics Denmark?

- LOG FILES - since log files may contain error messages or listings with identifiable information
- TABLES with less than five (5) observations in a cell.
- TABLES that, in certain combinations, make it possible to identify individuals etc.
- FIGURES with outliers or single data points.
- Listings with de-identified variables, such as the personal identification number (PNR). Even if the numbers are de-identified and the file does not contain anything apart from the de-identified number itself, it is still not allowed.
- Files with microdata information, even if the de-identified variables are removed.
- SYNTAX, since it can contain identifiable information. Transferring syntax between projects can in some cases be permitted.

If you discover that the rules set up by Statistics Denmark have been broken unintendedly please contact immediately:

- **CIRRAU:** Susanne Vind sbv@dst.dk (with Carsten Bøcker Pedersen cbp@econ.au.dk and Allan Timmermann ati@econ.au.dk as cc)
- **ECONAU:** Solveig Vibe-Pedersen svp@dst.dk and Oskar Enghoff oen@dst.dk (with Maria Knoth Humlum mhumlum@econ.au.dk and ECONAU data managers datamanager@econ.au.dk as cc)

The immediate contact is important since it can be regarded as mitigating circumstances if Statistics Denmark is informed about unintended mistakes as soon as the researcher is aware of the breach.

Remember that the same rules also apply if you have sent your own microdata to Statistics Denmark and later wish to work with the data outside the environment at Statistics Denmark. Once the data is located at Statistics Denmark it cannot leave this environment.

If you are in doubt about the rules of transferring specific information from the server, then you should aggregate the output further or contact your project leader/Data governance expert. An unintended violation of the rules can have very serious consequences for you and the entire research environment.

Please also consult Statistics Denmark's paper for more information on Data Security https://www.dst.dk/da/TilSalg/Forskningsservice/Dataadgang and Statistics Denmark's guidelines regarding transfer of files from the servers https://www.dst.dk/da/TilSalg/Forskningsservice/hjemtagelse-af-analyseresultater

Sending files to the servers

Files with non-personally identifiable information

To place information on the server that do not contain microdata, e.g. syntax/code and documentation, please send the file by email to:

- **CIRRAU:** Susanne Vind sbv@dst.dk, (cc Allan Timmermann ati@econ.au.dk)
- **ECONAU:** Oskar Enghoff oen@dst.dk (cc datamanager@econ.au.dk).

The email should have information about server and project number, where to place the file (complete path, typically in the workdata folder), and a statement stating that the file does not contain any personal identifiable information. Please, send only syntax or documentation when it is necessary and too large to practically type in manually.

Files with microdata

There are two ways to send data containing microdata to Statistics Denmark, e.g. data from sources outside CIRRAU or ECONAU (external data). Only external data that are needed for a project and described and approved in the DST project description can be sent to a project. Always include information about server name and project number as well as a description of the data and which variables that needs to be de-identified. If at all possible, data should be SAS files, which speeds up the handling time and lowers costs.

1) Use recommended mail to send encrypted files to Statistics Denmark on for example a USB-stick. Address the letter to "Service desk" and "ATTENTION: Susanne Vind" for CIRRAU and "ATTENTION: Berit Taul" for ECONAU.
2) Up-load files via Statistics Denmark FSE-UPLOAD. All users have access to this service in f5 dynamic webtop with project login. Guidelines for FSE-upload can be found here: https://www.dst.dk/da/TilSalg/Forskningsservice/brug-af-forskermaskiner

The latter method is preferred.

6. Best practice in data management

Keeping track of the research progress is the most important key to avoid unnecessary permanent datasets, and thereby avoiding unnecessary disk space use. Typically, register-based research utilizes large datasets, and several copies of these datasets will take up disk space, affecting negatively on the disk space for all users.

Keep track of your research progress & use of version control

When you prepare a dataset to perform analyses, e.g. on the relation between income and survival:

1) Save a copy of the SAS/STATA/R program/syntax performing your data management and analyzing tasks, and name this program e.g. inc_surv01.sas (or .r/.do). All file references in the program/syntax should be directly back to rawdata, thereby documenting all changes needed to prepare data for analyses. The perfect program/syntax includes a description of the reason for each of the tasks performed.

2) Save a copy of the workdataset when it is ready to be analyzed, and name this copy inc_surv01.sas7bdat (or .rdata/.dta). By naming your workdataset the same as your program/syntax, you will indirectly keep track of the program/syntax that generated your dataset for analyses.

3) Perform the analyses by writing code within the existing program/syntax or in an additional program/syntax that contain only the analytical code. This additional program/syntax should initially load the workdataset.

4) If changes are needed, (for example after an update of rawdata) generate a new program/syntax for this task by making an edited version of the former program/syntax. Name the new file inc_surv02.sas. As before, in the new version of the program/syntax all file references should be directly back to rawdata, and never to the former version of these data (e.g. inc_surv01.sas7bdat).

5) Refrain from using version control such as final, very final, second final, definite final, final2, etc. There will never be a final version (except when the paper has been published and you no longer work on the project).

6) You are advised to create a text file called 'readme.txt' in each directory where you shortly describe the contents of files in this directory as well as planned future steps, planned changes, along with dates. This will also help you keep track of your research process.

If you follow these guidelines, you will have a clear track of the research progress, and it will be possible at any stage to implement a new update to your data. In addition, it is possible to delete all intermediate datasets, and to regenerate data from program/syntax if needed.

Unfortunately, based on our experience, many researchers fail to keep track of the research process, making it difficult to later implement changes and update data in the project. Typically, when implementing obvious changes to the analysis the researcher do not find it necessary to document the changes. However, when the publisher returns the reviewed paper 9-12 months later, many researchers have forgotten the underlying reason for the implemented changes.

Ensuring the best performance of our common computer resources

To avoid unnecessary disk space use **never keep more than one copy of a dataset**. Many new users make a new permanent copy of a dataset for each minor change. Worst case scenario is that you get access to data1, and based on this you make a new variable called var2 and save this dataset as data2. Then you decide to make another variable called var3 based on the modified dataset data2, and save the new data as data3. If you end up making 10 new variables, you will have 10 nearly identical permanent copies of the same dataset. After a few years, you may request an update to data1; now you will have to repeat all these 10 steps to implement the update, while also saving additional 10 new permanent copies of the same datasets. This is worst-case scenario for data handling regarding permanent copies of datasets. Please feel free to make as many temporary datasets as needed. 'Temporary datasets' means datasets that are only stored on the server during the ongoing login session.

It is important that users:

1) **Keep only one workdataset for each project/paper.** Do not make copies for each minor change in the dataset. Instead, the user should keep the program/syntax that updates the one working dataset with changes. This will also facilitate documentation and overview in the end.

2) **Keep track of your own disk use**. A single user should not generate more than **50 GB per project.** It is easy to check disk-use by placing the mouse on your personal folder in "WORKDATA", make a right click, scroll down and click on "properties" and look at "size on disk". If the size of the data exceeds 50GB, then you are using too much disk space and you need to delete some of your workdatasets.

3) **Limit the use of virtual memory and CPU.** For example, users are encouraged NOT to run several processes at the same time and if possible, run heavy jobs at night or during the weekends. Users are restricted to use a maximum of **25GB memory** without a preceding agreement with a data manager. Please contact CIRRAU or ECONAU if you plan heavy jobs.

4) **Remember to "sign out" properly through "START"** so that temporary files are deleted. Accumulation of temporary files causes problems for disk space and memory.

5) **Remember to empty the "Recycle Bin" If you delete files.** Found on the desktop in the upper left corner. Simply right click on the recycle bin and choose empty recycle bin.

6) **Do not save files at disk C or F –** only save files in the WORKDATA area or in your personal folder "KODE".

To ensure the performance of our common server resources, CIRRAU and ECONAU management can exclude researchers that repeatedly fail to comply with these guidelines and delete their data.

7. Sanctions

According to ISO27001 (A.7.2.3) there shall be a formal and commutable disciplinary process in place to take action against users who have committed an information security event. It is the responsibility of each individual user to ensure compliance with criteria and guidelines. It is the responsibility of each data governance expert to ensure training, awareness and compliance with guidelines for all their associated users.

We have imposed sanctions parallel to those communicated by Statistics Denmark (https://www.dst.dk/ext/3477468153/0/forskning/Guidelines-for-transferring-aggregated-results-from-Statistics-Denmark--pdf).

Also, the first time a Data Governance Experts or one of its associated researchers fail to comply with guidelines resulting in a lockout of the whole institution there is no additional sanction imposed on the Data Governance Expert.

The second time within five years a Data Governance Experts or one of its associated researchers fail to comply with guidelines resulting in a lockout of the whole institution, the Data Governance Expert loses the right to download results from Statistics Denmark and lose the right to practice as a DGE for a period of two years.

Data managers who generally help with many downloads of results and who have the highest degree of training and awareness, but also highest risk of committing a mistake, are not sanctioned for mistakes regarding download of results.

8. Statistics Denmark Terms of Agreement

STATISTICS DENMARK TERMS OF AGREEMENT

1. The data sets to which access is given shall be treated as confidential information in accordance with Section 27, subsection 3 of the Danish Administration Act and Section 152 of the Danish Penal Code.
2. Processing of the basic data may only be conducted from the research environment for which the authorisation been granted, or access can also via the authorized research/analysis environment be switched to linked-up home computers in accordance with the guidelines determined by Statistics Denmark.
3. A computer linked up to Statistics Denmark may not be placed at the disposal of other persons, and the connection shall be completely turned off or disconnected, when the computer is not used, i.e. protected against unauthorized use.
4. Passwords, which are supplied by Statistics Denmark for the project are strictly personal and shall not be passed on to any third party.
5. Basic data as well as derived data sets shall not, neither directly nor indirectly, be downloaded.
6. All transfers of output (tables, analytical results), etc. for printing or for further statistical processing shall only take place in accordance with the guidelines and methods determined by Statistics Denmark.
7. A logging of these transfers is conducted by Statistics Denmark.
8. Confidential data shall not be printed, including data at the level of individuals or firms, and all output shall be aggregated in such a manner that it is impossible to identify individual persons or individual firms directly or indirectly. Attempts at identifying individual persons or firms are not permissible.
9. Access to the data is given for the period: ***2 years***
10. No information from the project in which it is possible to identify an individual person or individual firm may be published.
11. Published information from the project shall be submitted to Statistics Denmark for scrutiny.
12. The “token”, which has been provided for the project, shall be returned to Statistics Denmark when the agreement expires.

On attached researchers in particular.

13. The responsible person signing the agreement of authorisation for the authorised Danish institution, shall approve and assume the responsibility that all existing rules governing access to micro data are observed by the associated researcher.
14. It shall be the responsibility of the authorised Danish institution to inform the attached researcher of the rules governing the use of micro data, including the rules of confidentiality in force as well as the rules governing downloading of data

15. The associated researcher's access to micro data shall pass through the authorised Danish institution and can also be switched to linked-up home computers in accordance with the rules governing work from home
16. The authorised Danish institution appoints a contact person undertaking the responsibility for all contact with the attached researcher and Statistics Denmark.
17. All invoices concerning the attached researcher are forwarded to and paid by the authorised Danish institution in question in accordance with the terms of invoicing applicable to the institution.

A breach of the provisions of this agreement will imply that access to the data is immediately denied. Furthermore, the person who has signed this agreement will in future be excluded from using any of Statistics Denmark's research schemes. In the case of minor breaches, the person will be excluded from Statistics Denmark's research schemes temporarily for a period of not less than three years.

This agreement may be terminated by either party at 3 months' notice. If the authorisation of the research/analysis environment expires or is changed, this agreement is simultaneously cancelled.