

Guidelines and Criteria for CIRRAU and ECONAU servers at Statistics Denmark

Allan Timmermann^{1,2}, Ane Marie Closter^{1,2}, Marie Kempf Frydendahl^{1,2}, Susanne Jensen³, Jonas Mikkelsen³, Maria Knoth Humlum³, 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
- 2: The servers
- 3: Data security
- 4: Microdata
- 5: Transferring files
- 6: Best practice in data management
- 7: Sanctions against non-compliant users
- 8: Statistics Denmark Terms of Agreement

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 criteria and 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 data resources. 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 access to a server located at Statistics Denmark, which contains several specific research projects in which the data needed for each project is accessible. Researchers can access relevant research data 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. 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 controller responsible for the research project. 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 can grant a personal two factor 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 are pseudo-anonymized, 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 use their own 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 up by the authorized institution to which the project belongs.

CIRRAU and **ECONAU** have several hosted servers at Statistics Denmark. 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.

For details, please see the CIRRAU/ECONAU Processing Details here:

CIRRAU: <http://cirrau.au.dk/>

ECONAU: <https://econ.medarbejdere.au.dk/research/project-databases/econau/>

3. Data security

Security rules

Only persons employed at or affiliated with (by legal agreement) 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 Aarhus University's server at Statistics Denmark is allowed from Denmark or through an encrypted gateway to Aarhus University followed by personal two factor VPN access to Statistics Denmark. Aarhus University provides these encrypted gateways for all users.

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

It is not allowed to access a 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).

Adhering to CIRRAU's standard operational procedures and Aarhus University's general and specific policies, internal and external researchers may access CIRRAU/ECONAU servers at Statistics Denmark from Denmark and other GDPR countries.

Access to CIRRAU/ECONAU servers at Statistics Denmark from 3rd countries demand as an additional technical safeguard that potential download rights was disabled before traveling (contact the datamanagement team before traveling) and use of AU fully managed equipment or the authorized VDI.

It is unauthorized to access CIRRAU/ECONAU servers at Statistics Denmark from specified 3rd countries such as China, Iran, North Korea, Syria, Pakistan, Russia and India. This policy is a CIRRAU/ECONAU-specific addition to Aarhus University's one size fits all model for data access from 3rd countries.

New users must participate in an online or physical course through AU on rules for accessing and working with microdata.

No attempts must be made to identify individual physical persons or firms.

New users must sign/approve 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 8).

Additionally, users must read and comply with Aarhus University's information security policy (<https://medarbejdere.au.dk/en/informationsecurity/informationsecuritypolicy/>), Guidelines and Criteria for CIRRAU and ECONAU servers at Statistics Denmark (this document) (<http://cirrau.au.dk/data-resources/users-guides/>) and CIRRAU/ECONAU Processing Details (<http://cirrau.au.dk/data-resources/users-guides/>). All three documents must be read annually.

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

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 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 on 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 conducts 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. This also applies for previous exported tables, so by comparing two tables it must not be possible to identify individuals.

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 7 and 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, contact immediately:

- **CIRRAU:** Susanne Vind sbv@dst.dk (with CIRRAUs Allan Timmermann ati@econ.au.dk cc - please find further contact information here: <https://cirrau.au.dk/contact>)
- **ECONAU:** Kasper Buch Tanderup kta@dst.dk and Håkan Bernhardsson hbe@dst.dk (with ECONAU Maria Knoth Humlum mhumlum@econ.au.dk and ECONAU data managers datamanager@econ.au.dk 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.

In your email, please provide a detailed statement of when and how the violation happened and describe the nature of the violation. It is important that you do not include any microdata in the description of the violation. If a file with confidential microdata has been downloaded, it must be deleted immediately from all locations including e.g. local computer drives and e-mail accounts. Inform the relevant persons at

Statistics Denmark along with information about your username (3- or 4-letter ident), project number, name of the file, date and time of transferring the file and whether you have overruled warnings from the system that scans files (explain why).

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 data governance expert or the project controller responsible for the research project. 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>

6. Best practice in data management

For procedures and guidelines that must be followed to work on the server, please see the CIRRAU/ECONAU Processing Details here:

CIRRAU: <http://cirrau.au.dk/>

ECONAU: <https://econ.medarbejdere.au.dk/research/project-databases/econau/>

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.

CIRRAU or ECONAU 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

All researchers accessing research data through Statistics Denmark both need a signed affiliation (Tilknytningsaftale) and user (Brugeraftale) agreement with Statistics Denmark. See details under the section “Øvrige aftaler, dokumenter og vejledninger” in Danish here:

<https://www.dst.dk/da/TilSalg/Forskningsservice/Dataadgang/Autorisering>