ESSnet DARA Work Package 1:
Handbook for members of the ESS, especially Access Facility Staff

This handbook is written by the ESSnet DARA project members
Stephen Bond, Maurice Brandt, Tony Chapple, Anja Crössmann, Eric Debonnel,
Philippe Donnay, Ana Dulce Duarte Pinto, Kamel Gadouche, Anja Hlawatsch, Julia
Höninger, Gyoergy Káplán, Joaquim Machado, Anja Malchin and Zoltán Vereczkei.
# Table of contents

1. **Introduction** .................................................................................................................... 4
   - The Handbook for ESS members .................................................................................... 4
   - List of Abbreviations ........................................................................................................ 8

1. **Accreditation guidelines for Access Facilities** ................................................................. 9
   1.1 Process of accreditation ............................................................................................. 9
   1.2 Organisational requirements .................................................................................... 10
   1.3 Requirements for buildings housing DARA access .................................................. 10
   1.4 Support staff requirements ....................................................................................... 11
   1.5 Requirements for information management environment ......................................... 11

2. **Technical connection to DARA network** ....................................................................... 12
   2.1 Secure remote access system ................................................................................. 12
   2.2 How to set up the system ......................................................................................... 13
   2.3 Secure access ......................................................................................................... 13

3. **Recognition of research entities** ................................................................................... 15
   3.1 Current practice – under Regulation (EU) No 557/2013 ........................................... 15

4. **Process and models for signing confidentiality undertakings** ....................................... 16
   4.1 Current practice – under Regulation (EU) No 557/2013 ........................................... 16
   4.2 Discussion of sanctions for confidentiality breach .................................................... 16

5. **Review process for research proposals** ......................................................................... 18
   5.1 Current practice – under Regulation (EU) No 557/2013 ........................................... 18
   5.2 Application form for research proposals – Proposal for changes .............................. 18

6. **Research management system** ..................................................................................... 20
   6.1 Principles and framework ......................................................................................... 20
   6.2 Suggested workflow and databases ......................................................................... 20
   6.3 Suggested system content ....................................................................................... 21
   6.4 Suggested databases .............................................................................................. 21
   6.5 Suggested variables ................................................................................................ 21
   6.6 Research administration used by GENES for the SD-Box solution .......................... 25

7. **Management of user accounts** ..................................................................................... 26
   7.1 User accounts for researchers ................................................................................. 26
   7.2 Functionalities of researchers’ accounts .................................................................... 27
   7.3 Structure of disk space for researchers ..................................................................... 28
   7.4 Handing out login information .................................................................................. 29
   7.5 User accounts for AF staff and their structure .......................................................... 29

8. **Data management** ......................................................................................................... 31
   8.1 Statistics available for use ........................................................................................ 31
   8.2 Preparing an EU Statistic to be available through Access Facilities ........................ 31
   8.3 Creating data sets for research projects ................................................................... 32

9. **Input and output checking** ............................................................................................ 33
   9.1 Output checking guidelines to be applied ................................................................. 33
   9.2 Workflow of output checking .................................................................................... 34
9.3 Input checking ........................................................................................................... 35
9.4 Responsibilities regarding output and input checking in the long term ................. 36
9.5 Specific output checking guidelines ......................................................................... 36
10 End of research projects ......................................................................................... 38
10.1 Rules regarding research project end .................................................................... 38
10.2 Archiving and removing accounts ....................................................................... 38
Appendix ....................................................................................................................... 39
Appendix 1: Compliance checklist ................................................................................ 39
Appendix 2: Draft application form for Access Facilities ............................................. 42
Appendix 3: Code of connection (CoCo) .................................................................... 52
Appendix 4: Scenarios for the Research Management System .................................... 54
Appendix 5: Principles for research project administration ........................................ 55
Appendix 6: Variables & Databases ............................................................................. 56
Appendix 7: Screenshots from research administration system for SD Box solution ... 57
Appendix 8: Scenarios for Output checking .................................................................... 61
Appendix 9: Form for specific output checking guidelines .......................................... 64
Introduction

by Anja Hlawatsch and Julia Höninger

The Handbook for ESS members

The aim of the ESSnet Project „Decentralised and Remote Access to Confidential Data in the ESS“ (ESSnet DARA) is to implement remote access to confidential microdata from Safe Centres in the National Statistical Authorities (NSAs) of Member States (MS). Through a secure channel, researchers work with microdata stored on a safe server a central node (at Eurostat); data are not transferred to the MS but remain on the server at the central node (Eurostat). This way researcher can use confidential European statistics in their own MS without travelling to Luxembourg. The successful operation of the DARA project depends upon (1) an effective IT solution and (2) efficient workflows so that the necessary decisions can be taken quickly, sensibly and easily.


- access to a scientific-use file in the recognised research entity or
- access to secure-use files from an accredited Access Facility (AF).

This manual describes the data access through an AF.

The ESSnet DARA uses the two terms Access Facility (AF) and Safe Centre. An AF is a general term used to describe the "physical or virtual environment and its organisational setting where access to confidential data for scientific purposes is provided" (definition from Reg. No 557/2013) by the DARA system. "The AF shall be located within National Statistical Authorities (NSA)". A Safe Centre is a secure room where the access to confidential data is guaranteed including the organisational and administrative infrastructure. If a Safe Centre is accredited by Eurostat it becomes an AF and can offer access to European confidential data stored in the remote server in Eurostat. In the ESSnet DARA context, the AF is the Safe Centre located within an NSI or NSA.

The manual is meant for Safe Centres’ staff members working either at Eurostat or in NSAs of the EU that already are or want to be accredited as AF. It describes all steps in the process of providing researchers data access. Data access is therefore broken down into a set of discrete steps, which are presented in chronological order. Staff members can use the manual as a guidebook when setting up an AF.

The steps of the processes are linked to each other and depend on the technical solution. Whenever one part of the process is finished it triggers the start of the next, and this is highlighted in the handbook.
Regarding the technical solution, the manual is generic. Due to unavailability of VIP SICON infrastructure for DARA the ESSnet DARA team opted for the alternative solution and set up a pilot using the basic system of the SD Box with – instead of Eurostat – France (GENES) acting as central node. It was adapted to European needs and the requirements defined in WP2 of the ESSnet DARA. Special details of the tested DARA pilot solution have been added at the end of each chapter in a grey text box.

Picture 1: Sketch of a decentralised remote access system, implementation of DARA pilot

In a system of decentralised access to microdata from Safe Centres located in NSAs, different tasks have to be done by different parties. The first group of tasks is more Eurostat related and focuses on the preconditions that have to be fulfilled before researchers come to the Safe Centre (chapters 1 to 5 of the handbook):

- A Safe Centre is accredited as AF (accreditation criteria described in chapter 1).
- The technical infrastructure is installed (chapter 2).
- The entity applying for access is recognised as a research entity by Eurostat (chapter 3).
- Confidentiality undertaking and terms of use are signed by the duly designated representative of the research entity (chapter 4).
- A research proposal is submitted by the researcher affiliated with the recognised research entity (chapter 5).

The second group of tasks is more Safe Centres oriented. Chapters 6 to 10 describe how the access to the DARA system is managed and organised for researchers, in particular:
how staff members of an AF keep overview of all research projects, researchers and research entities (chapter 6),

how the accounts to the remote access network have to be set up (chapter 7),

how the data for the project has to be provided (chapter 8),

how the input and output is checked (chapter 9),

what to do at the end of the research project (chapter 10).

The ESSnet DARA team established the following roles that people can take in the system. One person sometimes takes two or more roles, sometimes several people share the work of one role.

Table 1: Roles in the remote access system

<table>
<thead>
<tr>
<th>Short name</th>
<th>Long name of role</th>
<th>Description of role</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITU</td>
<td>IT Unit</td>
<td>Responsible for deployment, set up and maintenance of the remote access system.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>See especially chapters 2 and 7.</td>
</tr>
<tr>
<td>SO</td>
<td>Support Officer</td>
<td>Single contact point between researcher and other people in the process. Helps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>researchers with all questions regarding the research project. Works in the AF.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>See all chapters.</td>
</tr>
<tr>
<td>OC</td>
<td>Output Checker</td>
<td>Checks researchers’ output against disclosure of confidential data.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>See especially chapter 9.</td>
</tr>
<tr>
<td>DM</td>
<td>Data Manager</td>
<td>Prepares researchers’ datasets and metadata and transfers them into researchers’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>workspace.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>See especially chapter 8.</td>
</tr>
<tr>
<td>Researcher</td>
<td>Researcher</td>
<td>Uses microdata and different statistical software via remote access within a Safe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Centre of an AF.</td>
</tr>
</tbody>
</table>

The workflow, the relationships and interactions between the different roles, is depicted in the graph below. For the ESSnet DARA project group it is important that the researcher has one main point of contact. Contact between Output Checker (OC) and researcher is required, if there are questions relating to the release of output.
Picture 2: Sketch of workflow in the remote access system
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF</td>
<td>Access Facility</td>
</tr>
<tr>
<td>CIRCA</td>
<td>Communication and Information Resource Centre Administrator</td>
</tr>
<tr>
<td>DARA Box</td>
<td>SD Box especially designed for international microdata access</td>
</tr>
<tr>
<td>DARA pilot</td>
<td>Remote access system with DARA Box and France as central node</td>
</tr>
<tr>
<td>DARA</td>
<td>Decentralised and Remote Access to Confidential Data in the ESS</td>
</tr>
<tr>
<td>DG DIGIT</td>
<td>Directorate-General for Informatics (within European Commission)</td>
</tr>
<tr>
<td>DM</td>
<td>Data Manager</td>
</tr>
<tr>
<td>EC</td>
<td>European Commission</td>
</tr>
<tr>
<td>ECAS</td>
<td>European Commission's user Authentication Service</td>
</tr>
<tr>
<td>ECHP</td>
<td>European Community Household Panel</td>
</tr>
<tr>
<td>ESS</td>
<td>European Statistical System</td>
</tr>
<tr>
<td>ESSC</td>
<td>European Statistical System Committee</td>
</tr>
<tr>
<td>ESSnet</td>
<td>ESS network</td>
</tr>
<tr>
<td>EU-SILC</td>
<td>European Union Statistics on Income and Living Conditions</td>
</tr>
<tr>
<td>FSO</td>
<td>Federal Statistical Office</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technology</td>
</tr>
<tr>
<td>IM</td>
<td>Information Management</td>
</tr>
<tr>
<td>ITU</td>
<td>IT Unit</td>
</tr>
<tr>
<td>LFS</td>
<td>Labour Force Survey</td>
</tr>
<tr>
<td>MS</td>
<td>Member State</td>
</tr>
<tr>
<td>NSA</td>
<td>National Statistical Authority</td>
</tr>
<tr>
<td>NSI</td>
<td>National Statistical Institute</td>
</tr>
<tr>
<td>OC</td>
<td>Output Checker</td>
</tr>
<tr>
<td>RDC</td>
<td>Research Data Centre</td>
</tr>
<tr>
<td>MAMS</td>
<td>Microdata Access Management System</td>
</tr>
<tr>
<td>SDC</td>
<td>Statistical Disclosure Control</td>
</tr>
<tr>
<td>SO</td>
<td>Support Officer</td>
</tr>
<tr>
<td>sTESTA</td>
<td>Secure Trans European Services for Telematics between Administrations</td>
</tr>
<tr>
<td>VIP-SICON</td>
<td>Vision Infrastructure Project on Secure Infrastructure for CONfidential data access</td>
</tr>
<tr>
<td>WG</td>
<td>Working Group</td>
</tr>
<tr>
<td>WGSC</td>
<td>Working Group on Statistical Confidentiality</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
</tbody>
</table>
1 Accreditation guidelines for Access Facilities

by Steve Bond and Tony Chapple

Table 2: Sources that have been used to create chapter 1

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulation (EU) No 557/2013.</td>
<td>European Statistics as regards access to confidential data for scientific purposes.</td>
</tr>
<tr>
<td>RatSWD.</td>
<td>Criteria of the German Data Forum for the establishment of Research Data Infrastructure.</td>
</tr>
<tr>
<td>HMRC Datalab</td>
<td>HMRC Datalab: Engaging with the External Research Community.</td>
</tr>
<tr>
<td>SDS / VML</td>
<td>Various papers describing the VML and SDS infrastructure and accreditation documentation.</td>
</tr>
</tbody>
</table>

1.1 Process of accreditation

Regulation (EU) No 557/2013 establishes the conditions which govern remote access to confidential data for researchers. A key element of the security model underlying the move to remote access for Eurostat microdata lies in the AF which must be accredited. This document builds on Eurostat Guidelines which state that AF will be assessed on criteria referring to the purpose of the AF, its organisational structure and standards for data security and data management.

The following chapter presents a detailed set of requirements along with minimum standards, where appropriate, that a Safe Centre will need to meet in order to be accredited as an AF. These requirements have been developed from accreditation guidelines used in existing Research Data Centres (RDC) run by partners in the ESSnet DARA team. A checklist for all requirements that have to be met can be found in the appendix (see Appendix 1). These requirements were developed under assumption that a system similar to the one tested by the ESSnet team will be developed by Eurostat. This system should ensure secure remote access to the data stored in the Eurostat secure environment.

The following describes the process for NSAs that wish to be accredited as AF and to provide access to confidential data of European statistics for scientific purposes via remote access in a Safe Centre:

1 Eurostat initiates a call for applications to become AF (starting with NSIs).
2 The duly designated representative of the NSA hosting Safe Centre wishing to become AF makes a formal application to Eurostat for accreditation (see Appendix 2), also submitting the forms in Appendix 1 and Appendix 3.
3 The selection of the AF is carried out by means of the assessment procedure regarding the Guidelines of Regulation (EU) No 557/2013 conducted by Eurostat.
4 Eurostat assesses the AF against criteria specified in this document. A site visit may be requested by Eurostat to check details in the application form.

5 If the proposed AF doesn’t have sufficiently trained staff for output checking or screening inputs, other MS will be asked to provide cover and support whilst experience is gained. As in the medium term Eurostat will be responsible for output checking, this criterion may therefore be suspended and introduced at a later point when output checking will be done by MS.

6 The list of selected AF is presented for adoption to the ESS Committee. A MS may take the decision to refuse the use of their data within the proposed AF.

7 A contract between Eurostat and the AF is signed for a defined period of time.

8 The name of the accredited AF is added to the published list of accredited AF at the Eurostat website.

1.2 Organisational requirements

Any Safe Centre wishing to become an AF must be able to demonstrate that it has the organisational structure, management processes and experience with research using confidential data to ensure that remote access from it poses no additional risk to confidentiality compared with existing AF in Eurostat.

An AF must be an organisation with an established reputation for being a body that promotes research and access to data for scientific purposes. Initially this will be limited to National Statistical Institutes (NSIs). According to the Guidelines, the “development of Access Facilities will be implemented in a stepwise manner starting with NSIs. Based on this experience, Access Facilities may be extended to other statistical authorities and then other facilities.” The term NSA in this document refers to NSIs and Other National Authorities responsible in each MS for the development, production and dissemination of European statistics.

The organisation also must have a proven track record of managing access to confidential microdata, either through remote data access, a Safe Centre in a RDC, or through scientific-use files.

The organisation must be able to demonstrate that the necessary legal background and infrastructure is in place to be able to pursue any actions that are needed in the event of a breach or suspected breach of confidentiality.

1.3 Requirements for buildings housing DARA access

A Safe Centre must be located in a secure building on the site of the NSA in order to prevent unauthorised access to the DARA terminals (either when they are in use by authorised researchers, or when they are not in use). The secure building must be controlled so that entry is only allowed following production of identification with a photograph (e.g. passport, ID card, or driving licence) against an authorised booking.

The Safe Centre must be locked when not in use, and configured so that unauthorised removal of confidential information is not possible.
The building must also provide adequate arrangements for the conduct of efficient research. It should be quiet, temperate and with access to comfort facilities. It should also conform to European and National legislation for disability access.

1.4 Support staff requirements

The staff of the Safe Centre who interact with the researchers provide a key role in both data security and researcher effectiveness. The SOs need to have necessary skills and experience with: security in an RDC environment, working with researchers and with the data that are being analysed. It is essential that the support function is carried out consistently across all AF. The support staff ensures that all researchers who gain access to the DARA system do so lawfully, working on a project approved by Eurostat.

It is not a requirement for new AFs that they provide output checking roles, as new AFs may not have sufficient expertise in this specialist subject. However, the operator of a prospective AF must provide evidence that it aims to develop this expertise.

1.5 Requirements for information management environment

Perhaps more than the other areas, the requirements for the information management (IM) environment are driven by the technical details of the remote access solution. However, as part of an application to become an AF, the applicant must demonstrate that no element of its infrastructure which connects (directly or indirectly) to the DARA infrastructure poses an additional threat (see Appendix 3). Such threats include the introduction of malware which interrupts operation of DARA, and penetration which compromises data security.

DARA pilot:
For the DARA pilot, the minimum requirement for the AF is a high bandwidth broadband connection with keyboard, mouse and monitor. The internet connection had to have a fixed IP address, a proxy to establish the connection is permitted. The connection is recommended to be through a separate tunnel, separately from the organisations own IT infrastructure.

Within the context of the ESSnet DARA pilot, data are not physically present at the point of access, but instead held securely within GENES environment and GENES assumed a role of a central node of the network. The access point allowed the researchers controlled access to view and manipulate data using locked down technology.
2 Technical connection to DARA network

by Maurice Brandt and Anja Crössmann

2.1 Secure remote access system

A decentralised remote access depends upon the effective IT architecture and efficient workflows (IT management) so that the necessary decisions can be taken quickly, sensibly and easily.

Major goals are to offer a highly secure remote access platform that data producers can trust:

- Strong user authentication based on certificate (PKI),
- Prevent any data leak: the user cannot get data without control.

Easy access and usage for researchers despite high security:

- Provide a complete environment with scientific and editing software.

The ESSnet DARA project did consider two kinds of technical solutions for a remote access system:

1) VIP-SICON

The VIP-SICON System ("Vision Infrastructure Project on Secure Infrastructure for CONfidential data access") is a framework designed for different purposes. The main goal is to set up a secure infrastructure for sharing and accessing EU confidential data from MS, e.g. Euro Group registers data. Part of VIP-SICON is to build an IT system within DG DIGIT/Eurostat for a remote access for scientific purposes via a thin-client and to set up the working platform and applications for researchers. It had been planned to use a CITRIX, VMware or Linux based system.

2) Alternative solution: DARA pilot

The DARA pilot is an alternative solution implemented due to unavailability of SICON by the end of the year 2012. France (GENES) was playing the role of the central node and allowed the DARA partners to test a highly secure remote access platform, its specific procedures and the IT security. This system was designed for an international microdata access as a proof of the DARA concept, but not applicable for a real production phase. This is not because of IT aspects; it is mainly caused by legal restrictions of the MS. The data sources remained in one unique central node infrastructure (isolated from the production network and internet).

Both solutions focus on various aspects such as IT infrastructure, technological features, security issues and procedural aspects i.e. guidelines, documentation of the infrastructure and the workflow.
2.2 How to set up the system

The general framework for a remote system is based on a central server at the secure central node and a thin client with individual user authentication in a controlled endpoint with a secure and encrypted connection. It is very important to consider, that the microdata at the central server can only be accessed by the researcher within an already secure environment of official statistics. The central server at Eurostat should make sure that:

- the connection is encrypted and secure
- the type of connection and network can be used by all endpoints of connection
- the identification of the researcher
- thin client provided by central node to assure compatibility and support

If the thin client is a hardware device it can be remote controlled and supported by the central node. This means all points of access are harmonised from an IT point of view. This makes it easier to support the access points in the Safe Centres to set up the system and to maintain it.

DARA pilot:
A SD Box from CASD in France especially designed for the project the so called DARA Box is supplied together with a special keyboard with fingerprint reader by GENES to all ESSnet DARA partners. To install the DARA Box, it is connected to electricity, a keyboard with fingerprint reader, a mouse and a screen. A dedicated broadband connection with a fix IP address is necessary. When connecting the DARA Box to the internet, the IP address and the proxy had to be dialled with buttons on the box. Colours signal the current state of play.

2.3 Secure access

The Researchers have access to microdata and different statistical software within a full Windows desktop. The working environment in the DARA network is meant to be user-friendly whilst complying with high level security requirements.

Researchers perform various actions and tasks within the DARA network environment as follows:

- use the functionality of the available/requested software tools to their full potential, including graphs
- browse all the data files and metadata that are available to them
- create, store and edit syntax files
- convert files to other formats
- combine, merge the available datasets
- perform folder management tasks within the working folder that suits researchers’ needs: new folders can be created, renamed, moved, deleted, etc.
- create temporary working files, save the content and resume the work at a later date.

Apart from the basic functionalities listed above, the following additional functionalities and features are available for researchers:

- Additional metadata and/or syntax files can be requested. Also external data files can be requested if they have been approved in the research proposal. These files
are made available for the research project after an input checking procedure (see chapter 9 of this manual).

- By default, the environment also offers a full office suite and software. Researchers have to be able to use software (e.g. SAS, STATA, R, SPSS, Microsoft Office, LATEX...). Researchers might request additional software and/or other versions of the software available.
- Backup service.
- Shared disk space for members of the same research project.

Besides technical, there are also administrative and organisational protection measures for the data. Researchers need to pass five barriers before the results are released in the DARA pilot. Apart from the technical barriers there are three barriers that are secured through the workflow of granting access to confidential data:

1. After a research entity is recognised, it can submit research proposals that identify certain persons as researchers who are permitted to get access.

2. If the research proposal is accepted, the research institution signs a contract; the researchers have to sign an individual confidentiality declaration and can book a DARA terminal in the Safe Centre.

3. There are physical identification checks at the entrance of a NSI or at the Safe Centre.

And of course there is always the output checking procedure at the end which assures that only safe output is released to the researchers.

Picture 3: Steps for microdata access
3 Recognition of research entities

by Ana Dulce Pinto, Anja Hlawatsch and Julia Höninger

3.1 Current practice – under Regulation (EU) No 557/2013

Only recognised research entities can submit research proposals (see chapter 5 on proposal review) and thus have access to confidential data of the ESS. A list of all recognised research entities is published on the Eurostat homepage:


To become a recognised research entity, a research entity has to be assessed by Eurostat. It has to meet the criteria established by the Commission in cooperation with the ESS Committee. The duly designated representative of the research entity has to fill out the application form for research entities. Together with this form, the signed confidentiality undertaking (Annex 12.2) and the terms of use (Annex 12.3) have to be sent to Eurostat Unit B1. For more information, templates and application forms please consult above mentioned Eurostat website.

Eurostat assesses the information provided in the above mentioned documents. If the assessment is positive, the name of the research entity is published on the Eurostat website. Eurostat provides NSAs with information received from applicants via CIRCABC. Recognition of the research entity enables researchers from that entity to submit research proposals.

Eurostat is responsible for re-assessments of recognised research entities at regular intervals. Compliance audits concerning the observance of the terms of use of confidential data may be conducted by Eurostat also on an ad-hoc basis. Entities that do not meet the assessment criteria are removed from the list of recognised research entities.

DARA pilot:

During the DARA pilot phase, no research institutions or confidential EU statistics were directly involved. DARA project members and other staff members in the NSA tested the functionality of the system with a public-use file. Furthermore selected researchers that were personally appointed as test researchers had access.
4 Process and models for signing confidentiality undertakings

by Ana Dulce Pinto, Anja Hlawatsch and Julia Höninger

4.1 Current practice – under Regulation (EU) No 557/2013

Since 8 July 2013 Regulation (EU) No 557/2013 is the legal framework that implements Regulation (EU) No 223/2009 as regards access to confidential data for statistical purposes (repealing Commission Regulation 831/2002). The Guidelines linked to the Regulation contain models of:

- the confidentiality undertaking that has to be signed by the research entity's duly designated representative when requesting the recognition of the research entity (art. 4º, nº2) (see chapter 3 of this handbook)
- and the individual confidentiality declaration that has to be signed by each researcher when submitting a research proposal (art. 5º, nº 2) (see chapter 4 of this handbook).

4.2 Discussion of sanctions for confidentiality breach

Confidentiality breaches are not matters of criminal cooperation at the EU level and considering this the Commission / Eurostat cannot apply criminal sanctions on this subject. If a breach of confidentiality occurs, according to the territorial principle in many countries the national law of the place where the relevant performance occurs is the legal basis for a prosecution (principle lex loci delicti).

However, national laws are very different and the possibilities of prosecution according to national laws vary. In order to have an overview, in January 2013 Eurostat sent a questionnaire to all MS (plus Switzerland and Serbia) to inquire about the existing provisions on sanctions for violation of statistical confidentiality. The results showed different scenarios on administrative, penal and disciplinary sanctions among the 24 MS (as of September 2013) that filled in the questionnaire.

Therefore in a decentralized scenario, there is currently no equal treatment in the case a breach of statistical confidentiality occurs. The treatment will depend on the national provisions in criminal law and local administrative provisions. The contractual sanctions will be applied by Eurostat in a uniform way.

As Article 26º of Regulation 223/2009 states that MS and the Commission must take all the appropriate measures to prevent and sanction violations of statistical confidentiality, the ESSnet DARA team highlights that all MS must converge on a common understanding on the issue. Considering the difficulties of a European law everything should be locally undertaken to allow MS to apply their national sanctions as the subsidiary law but in a harmonised way.

In order to reach a common understanding, MS should take the existing provisions on criminal and administrative sanctions (see Eurostat questionnaire on applied sanctions
against breach of confidentiality in the MS) and check between themselves the principles and boundaries they can accept considering a decentralised scenario in order to define what can be harmonised among them.

As an additional criterion for accreditation of AF, the applying institution shall demonstrate which sanctions there are and how are the sanctions for violation of statistical confidentiality dealt in the national law where the AF is situated geographically (see chapter 1 of this manual). Furthermore, the institution shall describe the ability to prosecute (if applicable) and the ability to manage and pursue any actions that need to be taken in the event of a breach or suspected breach of confidentiality. When an AF is accredited and signs a contract with the Commission, the obligations concerning the protection of confidential data and organisational measures (art. 8, nº 6 of the Regulation 557/2013) shall be determined and the national legal measures allowing prosecution in case of misuse of confidential data should be described.
5 Review process for research proposals

by Maurice Brandt and Anja Crössmann

5.1 Current practice – under Regulation (EU) No 557/2013

The assessment process for research proposals is described in the Regulation (EU) No 557/2013, Art. 5 “Research proposal” and Art. 6 “Position of national statistical authorities” as well as the Guidelines of the Regulation chapter 5 “Research Proposals: Guidelines for assessment”.

5.2 Application form for research proposals – Proposal for changes

Once the possibility is given to access European confidential data in the accredited Safe Centres, it is important to change the application form for research proposals. The following suggestions describe how the changes could be done: Items 3. “Datasets to be used”: 3.1 “Please select the dataset(s) to be used” and 3.2 “Please state the type(s) of confidential data for scientific purposes to be used” should be combined in one question to 3.1 “Please select the dataset(s), types and location of confidential data for scientific purposes to be used”. The datasets should be categorised by the availability of the ways of access – scientific-use files and secure-use files and the locations in which Safe Centre which dataset can be used.

**Scientific-use files**
(Confidential data available in your recognised research entity via e.g. CD-ROM provided by Eurostat)

- European Community Household Panel (ECHP)
- Labour Force Survey (LFS)
- European Union Statistics on Income and Living Conditions (EU-SILC)
- Adult Education Survey (AES)
- Community Innovation Survey (CIS)
- Structure of Earnings Survey (SES)
- Community Statistics on Information Society (CSIS)
- Continuous Vocational Training Survey (CVTS)
- European Health Interview Survey (EHIS)

**Secure-use files**
(Confidential data available in Eurostat or another recognised Safe Centre)

Safe Centre of Eurostat, Luxembourg
- Community Innovation Survey (CIS)
- Structure of Earnings Survey (SES)

(e.g.) Safe Centre of Hungary, Budapest
- (e.g.) Labour Force Survey (SES)

(e.g.) Safe Centre of Germany, Wiesbaden
- (e.g.) Community Innovation Survey (CIS)

etc.
Thus the accessibility of the datasets, types and existing Safe Centres would be clear for the researcher from the first view.
6 Microdata access management system

by Zoltán Vereczkei

6.1 Principles and framework

A microdata access management system (MAMS) that provides the necessary information to administer research projects in the DARA network is necessary. When it became evident that the VIP-SICON system would not be available within the DARA timetable, the DARA partners decided that DARA Box solution was to be used for the DARA pilot, the decision was made that the MAMS should not be developed any further. Instead, the work done should be summarised to serve as a recommendation for a future MAMS built for the real implementation. At the same time, to reflect more on the DARA pilot, the current practices of the SD Box system, regarding the management of research, is briefly summarised in this chapter.

During the earlier stages of development, DARA partners identified 4 scenarios as possible approaches for the MAMS. The DARA partners agreed that the MAMS should correspond to scenario D (please see Appendix 4), therefore the system is to be used for administration issues for all access modes and will store all relevant documentation for all research projects. At this stage, DARA partners were informed that Eurostat would develop a new research management system in 2013. As the whole MAMS is only partly connected to the remote access solution and is out of scope of the ESSnet DARA to handle, this recommendation should be used for the aforementioned Eurostat project. The task of ESSnet DARA, related to this system, is to clarify the concept, the needs and the content as far as possible.

One of the main goals of ESSnet DARA is to define rules and standards for remote access to confidential data in the ESS, including the definition of the workflow and documentation for all users of the remote access system. Research project administration, as part of this workflow, shall follow the several principles (see Appendix 5):

6.2 Suggested workflow and databases

The concept of the workflow and the roles are defined in the document “List of user requirements for a remote access system” (Deliverable in WP2 of ESSnet DARA). Data should be added to the system basically at the following three stages:

1. When a new research proposal is received
2. When the research proposal is approved or rejected
3. At later stages when output is created (where relevant) and at the end when the research project is finished

Ideally, the database should be managed by the person who has all the information required. This person should be the SO of the project coordinator in the AF, who is connected to all important roles of the workflow and provides support during the research project.
Depending on who will receive the research proposals, the following scenarios are envisaged:

1. If Eurostat receives the research proposal, information on the proposal will be recorded at Eurostat.
2. If Eurostat receives the research proposal but the proposal is forwarded to the AF concerned (where research will take place), the information on the proposal will be recorded by the SO at the AF.
3. If the AF receives the research proposal, information on the proposal will be recorded by the SO at the AF.

6.3 Suggested system content

This part delivers a recommendation for the variables of the MAMS. Each variable is associated with one or more status letters. These letters are to be interpreted as follows:

→ A: automatically generated
→ S: standard format
→ M: manual data entry
→ L: content is selected using a standard list or data entry is supported by such lists
→ D: documentation upload

6.4 Suggested databases

Five databases are proposed to be set up and kept up-to-date. These databases function in the background of the MAMS and provide information for the automatic lists for the corresponding variables. The databases contain the names and contact information of all relevant entities (address, telephone number, e-mail address; except the database of microdata (MDB) where the names of available microdata sets are listed):

1. Database on Accredited Access Facilities (AAPDB)
2. Database of Research Entities (REDB)
3. Database of Researchers (RDB)
4. Database of Microdata (MDB)
5. Database of Output Checkers (OCDB)

Following the approach of the three stages where information shall be entered in the system, the proposed variables are grouped and summarized in Appendix 6.

6.5 Suggested variables

1. **Date of the reception of research proposal | M, S**
   
   The day the AF receives the research proposal. The date shall be entered manually according to the standard format of <DD> <MM> <YYYY>.

2. **Unique ID of the research proposal/project | A**
3. **Project coordinator AF** | L

A coordinator AF shall be assigned to each research proposal. This function is particularly useful in situations when researchers work on the same research project at different AF. The project coordinator AF is responsible for the administration of the project. By default value, the project coordinator AF is the one recording information of the research proposal and its details are listed in the AAFDB (coordination can be assigned to another AF).

4. **Type of access** | M

First of all, the type of access should be chosen. Currently 3 types of access to European confidential data are foreseen: microdata release on CDs/DVDs; access in Safe Centre environment at Eurostat; access in remote access environment in a Safe Centre of a MS

5. **Name of research entity** submitting a research proposal | M, L

The research entity can be chosen from the list (content of REDB). In case of a new research entity, the information is to be recorded manually. When the name of a new research entity is entered and saved, the information is automatically registered into the REDB in the background. Next time when the name of a research entity has to be chosen, the new research entity will be on the list offered by the system.

6. **Contact information of research entity** | M, A

If the name of the research entity can be found in REDB, the contact information (address, e-mail and telephone) is automatically listed. If there is a change in contact details, the information can be overwritten (prompt changes made to REDB). If a new research entity is to be recorded, the contact information has to be given manually (address, e-mail and telephone). This information is automatically registered into the REDB in the background.

7. **Names and contact information of researchers** | M, L

The names of all researchers who want access to the data at the AF have to be registered. If the name of researcher can be found in RDB, the contact information (address, e-mail and telephone) is automatically listed. If there is a change in contact details, the information can be overwritten (prompt changes made to RDB). If a new researcher is to be recorded, the contact information has to be given manually (address, e-mail and telephone). This information is automatically registered into the RDB in the background.

8. **Name and contact information of the principal researcher** | M, L

The name of the principal researcher has to be indicated. This researcher can either be one of the researchers accessing the microdata (name can be selected from a list based on the content of RDB) or a researcher not listed among those researchers (the principal researcher might not work with the microdata). If the name and/or contact information of the principal researcher is not found in RDB, manual data entry will be required. Making changes is also possible (overwrite information in RDB). If a new principal researcher is
to be added to the database, contact information has to be registered as well. This information is automatically registered into the RDB in the background.

9. **Name of microdata to be accessed** | M, L

   If a list of available microdata at Eurostat can be compiled, it can automatically be offered for selection (content of MDB). As researchers might want to access multiple microdata files or want to link these datasets with other microdata, manual description may also be needed.

10. **Reference time of microdata to be accessed** | M

    Reference time of the accessed microdata shall be stated. If the researcher would like to access multiple microdata files, reference time should be recorded for each file. As microdata files might refer to years, months or other periods of a year or several years or several periods of given years, this information is to be recorded manually.

11. **Type of the research proposal** | L

    A standard codelist of research project types has to be defined. The selection can be made from this standard codelist.

12. **Short description of the research proposal** | M

    Brief and concise description of the research project (further clarification of research project type, intended utilization of research results, etc.).

13. **Research proposal** | D

    The received research proposal is to be uploaded into the system.

14. **Date of approval or refusal of research proposal** | M, S

    Date of research project acceptance or refusal shall be entered manually according to the standard format of <DD> <MM> <YYYY>. Date of acceptance or refusal is the date when the unit responsible for acceptance of research projects gives its formal approval or refusal.

15. **Reason for refusal** | M

    In case of refusal of a research proposal, the reason of refusal shall be summarized.

16. **IDs and names of the Access Facilities** | L

    IDs and names of the AF where the actual access takes place for the research project have to be selected from AAFDB. Eurostat is responsible to keep the list up-to-date (has to be harmonised with the Accredited AF list published on Eurostat website. All information should practically be extracted from the same source).

17. **Contact details of the Access Facilities** | A

    With the IDs of the AF selected, the contact details of the AF (address, e-mail, telephone, representatives and their contact information) are automatically derived from AAFDB.

18. **Starting date of research project** | M, S

    Information entered manually according to the standard format of <DD> <MM> <YYYY>. Starting date of the project is the date when:
− the account of the researcher is set and access to microdata is granted (remote access)
− the access to microdata is granted (Safe Centre)
− the microdata are released on CD or DVD (release on CD or DVD)

19. Finish date of research project | M, S

Information entered manually according to the standard format of <DD> <MM> <YYYY>. Finish date of the project is the date when:
− the last output is delivered to the researcher (remote access and Safe Centre)
− the finish date of the research project stated in the contract is reached (release on CD or DVD)

20. Duration of the project | A

This field is automatically generated based on the values for “starting date of research project” and “finish date of research project”. Value for this field (in months) is calculated only when both criteria are entered. As long as only the “starting date of research project” field is defined, the status of “Currently ongoing” is shown. If a maximum duration of research projects can be defined, the value for this field will not exceed that threshold.

21. ID of contract | M

Unique ID of the contract has to be entered manually.

22. Contract | D

The signed contract is to be uploaded into the system.

23. Number of outputs checked | M

This information is recorded manually by the SO at the end of the project.

24. Name of output checkers | L, M

Name of the OC should be selected from OCDB. Selection of more than one name is a possibility.

25. Details of output checkers | L, M

Name of the OC linked to the given research can be selected from OCDB. In that case, the previously registered contact details are listed. Making changes is possible (overwrite information in OCDB). If a new OC is added to the database, this information will automatically be registered into the OCDB in the background.

26. Availability of archived research outputs | M

Links to the archived research outputs should be given where access to the research outputs shall be granted to authorised personnel. Links are to be given for each output.

27. Costs of the research project | M

Costs of the research project should be registered.
6.6 Research administration used by GENES for the SD-Box solution

As the DARA Box solution has been chosen for the pilot system for ESSnet DARA which is based on the SD Box system, widely used in France for remote access to statistical information, DARA partners decided that it would be useful to describe the key characteristics of the research administration perspectives of the SD Box system.

General information on user management:

The management system is based on a web application. It is aimed at easing daily tasks of 3 different actors, all staff members of GENES – CASD.

− IT staff can manage SD Box, virtual research environments and users
− Statisticians can manage research entities, data sources, projects, input/output requests and researchers
− Contract manager can generate different contracts and update relative information

Brief description of the process of project management:

Researchers have to submit an application form for their research. When successful, this information is transmitted to the CASD: statisticians register it in the web application and a copy of the form is archived. Some missing information can be gathered later, for instance when the researcher comes to be enrolled and signed an agreement and contracts. As of now, the information is kept for 4 years after the end of the project, notably general user information, access rights and all input/output requests.

Management of user rights:

Upon receiving a project habilitation, statisticians put access rights in the web application. Inside the closed environment, statisticians grant read-only access to the project, shortly after project creation, but this can be completed for instance if data is not yet available. In order to actually work on the data, the researcher has to be enrolled and activated and its project environment created and running.

Only IT can deactivate an account, either on a motivated request from a non-IT colleague through the web application or for IT reasons.

Screenshots taken from the research administration system for the SD Box solution are included in Appendix 7 (only available in French)).
7 Management of user accounts

by Anja Hlawatsch and Julia Höninger

7.1 User accounts for researchers

When a research proposal is accepted by all relevant stakeholders, the SO of the AF, where data shall be accessed, creates a new research project and user accounts for the researchers. There are different options for who sets up the project and the accounts depending on the IT solution for decentralised data access:

- If only the central node administrators can create new accounts, then the microdata access team in Eurostat will provide the central ITU with all necessary information (name, project number, etc.).

- If Support Officers can create as well user accounts, they will set them up themselves on the basis of the information provided by the microdata access team in Eurostat.

Research accounts have a login name. This can be the researcher’s last name\(^1\), the last name and initial letter of first name or all researchers in a project are numbered. To log in, either a password, a smartcard and/or biometrics are required to identify oneself. The person responsible for setting up an account either sets a password or registers a smartcard or biometrics (for handing out the identification see section 1.4 of this chapter). If biometrics are used for identification, a decentralised registration process is preferred to avoid researchers having to travel to the ITU of the central node.

The DARA project team recommends for the beginning of a real implementation that the central node is located in Eurostat and the ITU of Eurostat creates the user accounts for all stakeholders involved in the DARA network.

Should users (a researcher or staff member of the AF) of the DARA network forget their password, they have to inform their SO and request a new password. If another mode of identification is used instead or in addition to a password (i.e. smartcard), this has to be replaced as well when lost.

\(^1\)If the last name is not unique, the project number can be annexed to the last name, e.g. Perreira0008.
7.2 Functionalities of researchers’ accounts

First a research project is created and then researchers’ accounts are attributed to the project. Several researchers can work on the same project. However, if one researcher works on different projects, the accounts have to be separate.

It is suggested that the projects will be named DARA0001, DARA0002 and so on. Some folders in each research project should automatically be created (see the following section).

If the ITU of the central node is responsible for setting up accounts, it also defines the settings for all researcher accounts. A standard profile is recommended. Researchers’ accounts have to be given the right to use different statistical software packages (desirable are the following: SAS, STATA, R, SPSS, Microsoft Office, LATEX). Furthermore, the profile defines when researchers’ accounts will expire (see chapter 10 of this manual), but the ITU of the central node can block accounts temporarily if an ID element is lost or a confidentiality breach is suspected. The settings include how much computation power and how much disk space is available. The DARA project team suggests that every research project should start with 8 GB RAM and 100 GB data storage. The ITU of the central node can change all these technical parameters during the project if a project has exhausted the limit and needs more working space or in case the project is prolonged. In this case the SO notifies the ITU of the central node.

In a real production phase, there might be different types of accounts for the researchers which can be more personalised. Research projects differ in software used, data sets differ greatly in size, and different analyses require different amounts of memory space. Therefore in a real production phase, different virtual desktops can be set up for different types of research projects.

DARA pilot:

First a research project is to set up in the management tool, then several researchers are attributed to the research project. All researchers have a standardised profile, all have access to the same set of software: SAS, STATA, R, StatTransfer, Microsoft Office, LATEX.

The DARA pilot system has 8 GB RAM and 100 GB storage.
7.3 Structure of disk space for researchers

User accounts for researchers need to have the structure as follows:

- The folder “Data” contains the original data. Researchers can read, but not write in this area.
- The folder “Metadata” contains information about the project data as well as guidelines on how syntax has to be written for an efficient check by the OC.
- In the folder “Input”, Safe Centre staff put input from the researcher (such as syntax) that has been checked for confidentiality (see chapter 9 on input checking).
- In the folders “Analysis” and “Output”, researchers have the permission to read and write and create new folders.

Several researchers of the same research project are able to access one shared disk space dedicated to the joint research project. Two options are suggested: Either all researchers share one “Analysis” folder or there are personal analysis folders accessible to single researchers and a shared “common” analysis folder accessible to all researchers of the project.

If a researcher is active in more than one project, different and completely separated accounts are created so transfer of data or other files is prevented.

Picture 4: Folder structure of research projects

```
/DARA0000x (Research project)

Analysis

HERE YOU CAN CREATE MORE FOLDERS AND ORGANIZE YOUR RESEARCH

Data
Input
2013_10_31
Metadata (includes also guidelines on how to write syntax)
Output
2013_10_21
2013_10_12
```

Picture 4 shows one example structure that fulfils the requirements defined by the DARA project group. The folders of the (original) data and metadata are filled by the DM. The researcher only has the right to write within the folders “Analysis” and “Output”. In all other areas access is limited to ‘read’.

The folders “Input” and “Output” are empty at the beginning of a research project. In “Output”, there will be a text message asking researchers to create a folder with the date for every time they request output: “Please create a folder with the date in format YYYY_MM_DD and put..."
your results in this folder if you want results mailed to your email address. Thanks, ESSnet DARA team.”

In a next stage of development, it is desirable to have an automatic email notification when new output is stored in the “Output” folder. The folder “Input” is filled by Safe Centre staff in the same format.

### 7.4 Handing out login information

The user account name and password are to be given to the users of the DARA network through a safe communication channel. As it is desirable that the SO is the single point of contact for the researchers, the SO hands out identification information. At the first visit to the Safe Centre, the SO checks the ID of the researcher. The SO then hands out user name, password, smartcard or token. Alternatively the SO calls the ITU of the central node, hands the phone to the researcher and the researcher gets the password on the phone. Another option is to send the password or smartcard by physical mail to the researcher.

The researcher has to change the password at the first log in to the system.

**DARA pilot:**
In the DARA pilot, the disk space is organised as described above. However, output is requested using an application. Therefore there is no need for a physical “Output” folder where researchers have to organise output with the current date, but these folders were automatically created by the application. OCs were notified of new pending output and input within the DARA network.

### 7.5 User accounts for AF staff and their structure

The central node creates user accounts for new SOs. New accounts for DMs and OCs may either be created by the central node or by the SO of the local AF. New accounts become necessary when an AF hires new staff members or a new AF is added.

User accounts for SO, DM and OC will have the permission to access all disk space of researchers’ projects and DM and OC need to have the right to use the same software as the researchers. Depending on how the output checking is organised, OCs also have to have own disk space and the right to create folders and save files. DM should additionally have access to StatTransfer. The rights of the roles SO, DM and OC are set centrally. Whenever a new staff member is enrolled, the rights are set in concordance to the role.

**DARA pilot:**
The finger print of the researcher had to be presented at the time of enrolment. The SO of the AF could enrol researchers. The smartcard, on which the encrypted finger print is stored, is handed to the researcher after the enrolment.
DARA pilot:
There are standardised profiles for the roles DM, OC and SO. DM and OC accounts can
not be created by the local SO, only by the central node. In the pilot SO, DM and OC have
access to all projects of all countries. In a real implementation this can be set according to
the final decision in the workflow.
OC can use statistical software packages and save documents and data. In the pilot OC do
not have access to the research projects.
8 Data management

by Anja Hlawatsch and Julia Höninger

8.1 Microdata available for use

A decentralised access to secure-use files from AF should start with the two datasets that have been available the longest period of time in the Safe Centre in Luxembourg: Community Innovation Survey (CIS) and Structure of Earnings Survey (SES). These data sets are prepared as secure-use files and there are output checking rules available. Furthermore there is a real value added in making CIS and SES available as secure-use files as with enterprises-based microdata, a lot of information is lost during the anonymisation process to produce scientific-use files.

The datasets available in accredited AF will always be published on the Eurostat website: http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/introduction

DARA pilot:
In the proposal, the ESSnet DARA project team announced that for the pilot, the ECHP would be used. However, with the central node in France and the revision of Regulation (EC) No 831/2002 not in place as of January 1st 2013, the DARA team agreed to use a public use file. A subsample of a public use file is used in the tests with the DARA-Box.

8.2 Preparing an EU microdata to be available through Access Facilities

In the real production phase, other data sets will be made available in due process. Additional datasets will be offered through the decentralised data access or subsequent years of data already available will be added. A process for how European data sets will be created and included in the service will emerge. In order for the data to be prepared by Eurostat, NSAs transmit data to Eurostat. Eurostat produces European statistics for general public and for some datasets prepares microdata for researchers, either in the form of secure-use files and/or scientific-use files.

Additionally, metadata and the specific output checking guidelines (see chapter 9 of this handbook) have to be compiled and agreed upon. Metadata has to be made available to current and potential researchers, so it should be published online. Specific output checking guidelines should only be made available to staff of AFs.

The ESSnet DARA project team discussed two different scenarios for how the data management could be organised in the long term of a real production phase. In a scenario A Eurostat would continue to do the data management. Scenario B, where different countries become experts of and a “centre of excellence” to a statistical topic, is recommended. Germany has organised on the national level the data management for research purposes in such a way that the responsibility for each of the over 100 statistics available in the Research Data Centres (RDC) is clearly mandated to either the Federal Statistical Office of Germany (FSO) or one of the Statistical Offices of the Federal States. However, whether scenario A or
ESSnet on Decentralised and Remote Access
to Confidential Data in the ESS – ESSnet DARA

B might be realised is heavily dependent on developments within the European Statistical System and on new EU regulations. Scenario B can be seen as a long term vision for development. Furthermore it is a question of how the financial burden can be shared (see WP3 cost and benefit analysis).

8.3 Creating data sets for research projects

From the European data set, a file that is specific to a research project will be created. A research proposal may have requested access to the data of only some of the countries. Then only the data that MS agreed upon can be made available. The DM will be notified by the coordinator of the review process of research proposals (see chapter 5 of this handbook) about the data MS gave consent for. They will then prepare the project specific data set and put it on an exchange area on the central server. The DM can either be members of staff at Eurostat or in one of the AF. As soon as the project specific data is available, the DM will move it into the researchers account.

A tailored dataset for each research project with only the variables needed for the research might be seen as more elaborative, associated with less risk, but more burdensome in creating. The DARA team suggests standardised data sets for all research projects.

DARA pilot:
During the DARA pilot, no new European data sets are compiled.

DARA pilot:
A small extract from the PUF is on the data storage inside the DARA network. DMs have access to the data storage “Raw_sources” and can compile specific data sets in this area. Then they placed a newly prepared data set into the folder “To_Be_Distributed\DARA00X” for a specific project. The SO of the respective project DARA00X has in turn access to this folder and moves the data into the project’s “DATA” folder (see chapter 1.7 of this handbook). Communication between DM and SO is done via email outside the DARA network.
9 Input and output checking

by Anja Hlawatsch and Julia Höninger

9.1 Output checking guidelines to be applied

Every output has to be checked before it is to leave the premises of the statistical offices in order to assure the statistical confidentiality of the microdata. Results which have been validated are safe to be published or used further outside the AF.

The guidelines on output checking should generally be followed whenever output is checked, which has been generated by researchers in a Safe Centre. The ESSnet SDC published in 2009 the guidelines for output checking. They are available from:


They have also been published in the book Hundepol et al. (2012)\textsuperscript{2}, \textsuperscript{3}. These general guidelines for output checking were discussed and approved by the Expert Group on SDC in April 2012.

It would be desirable if output is checked according to the principles-based approach. However, a high level of experience in output checking on the side of the OCs and in-depth experience with SDC on the side of researchers is required for a principles-based approach. Therefore, alternatively output can be checked according to the rule-of-thumb model.

In addition to the general guidelines, there should be specific output checking rules for every dataset that is available through AF. In these specific guidelines, the parameters for sensitivity rules, country specifics\textsuperscript{4} and details about the statistical disclosure control methods will be documented. The specific guidelines should be a separate document drafted by the WGSC assisted by the Expert Group on SDC and agreed upon by the sectorial WG and the NSAs as laid out in section 7.3 of the Guidelines to the Regulation (EU) No 557/2013. These specific output checking guidelines will be compiled before a new dataset is released as secure-use file and published on the Eurostat website.

The DARA project team suggests the form in the chapter 9.5 to develop the specific output checking guidelines.


\textsuperscript{3}It would be desirable that these guidelines be updated according to new disclosure risks that have been identified.

\textsuperscript{4}Output checking rules should be as far as possible harmonised. Having different rules for different countries should be avoided.
9.2 Workflow of output checking

The workflow for output checking is as follows:

1. The researcher places the statistical analysis for publication in the folder “Output” for output checking (check chapter 7 on user accounts). The researcher leaves a text note with an agreed email address, usually relating to the research entity.

2. The researcher informs the SO of the AF that the file has been placed for export.

3. The SO informs the OC of the AF that there are new documents to control.

4. The OCs check the relevant output folder of the researcher and control the output. To assure an adequately fast response for the researcher, checking-time should on average not exceed five working days. OCs have the option to look into the data the researcher uses, they have the statistical software packages and an own working directory. The OC can email directly to the researcher a comment on the released output or an explanation for a rejected output.

5. The OC can either release the complete output or rejected it in total, in which case researchers can improve the output if they still wish it to be released. In this case a hash tag is used to demonstrate the integrity of the output. The OCs can also carry out small modifications to the output in order to suppress single parts of the documents. The second option puts more burden on the OC but is more user friendly, as researchers do not have to travel to the AF for minor amendments to the output.

6. After the output is classified as safe, the OC sends it via email to the researcher and sets the SO in copy.

The OCs access the requested output of the researcher on the central server. It is most secure, if the output is still stored on the central server and not transferred to the MS. Every output checker has to have an account on the DARA network. This way syntax and output of research projects can be archived to have documentation and a history of the analysis that has been done and of the output that has been produced and released.

Often there are further enquiries, things are unclear, and researchers have to improve their output. Therefore there has to be a way of communication between the OC and the researcher. Even though the SO is the point of contact for the researcher in a variety of matters, all communication concerning the output should be directly between OC and researcher as fewer misunderstandings occur and output can be cleared faster. Whenever communication between the researcher and the OC takes place, the SO has to be notified (put in copy, for example).

DARA pilot:
A subsample of a public-use file is used by the DARA project to test the feasibility of the output checking procedures and workflow. Even though the microdata are already anonymised, the output checking process is carried out pretending the data were original and confidential. However, there are no specific output checking guidelines compiled for the public use file.
As an improvement in the output checking process, it is desirable if an automatically generated message informs the responsible OC that there are new documents to control in the researchers’ folder named “Output” (as described in chapter 7). But this is not mandatory.

9.3 Input checking

Researchers can import syntax into their DARA network workspace. Researchers email documents that they want to have imported to the functional email address. It is advised that the researchers do this two working days in advance of their pre-booked access to ensure that there is sufficient time for the input to be checked. The functional email address is published together with the list of accredited AF. In the email, researchers indicate whether the input shall be put in their personal input folder or the input folder common to all researchers of the project.

The SO of the Safe Centre receives the documents that shall be imported. The SO places the files received by email in a folder with the date of the day the email was received in the folder “Input”. The OC is informed automatically that there is input to be checked. The OC checks all documents whether the import is admissible. However, only syntax, metadata and technical papers on statistical methods shall be imported. Further data sets apart from the EU statistics applied for within the DARA network can only be imported if they were described in the research proposal and the import has been approved beforehand. The input becomes available to the researcher after the input has been accepted by the OC.

DARA pilot:

Within the DARA pilot the workflow is slightly different than in the recommendation above. If researchers want output, they include all files in a zip-archive. Then they right clicked on the archive and chose “DARA Output Request”. The output then appears in the application for OCs. OCs are informed automatically about new output by a dedicated application when they are logged in to the DARA environment. This output is marked with a unique identifier (hash code), so that researchers can verify with this code that their output is not altered when the output is released.

Using the hash code, outputs can not be altered, i.e. no suppressions can be incorporated. The complete output package can only be accepted or rejected. Unfortunately there is no possibility for the OC to write a comment to the researcher to explain why output is rejected or comment on accepted output. This has to be done via email. Furthermore, OCs can not send output to the researchers. Only the ITU of the central node is able to extract files from the closed system. The OC has to send an email to the central node asking them to email the approved output to the researcher.

DARA pilot:

Within the DARA pilot, only the central IT node can import files into the DARA network. The SO forwards documents that researchers want to import to the central IT node and provides the information for which project the input is and if the input shall be copied to the personal workspace of a single researcher or to the workspace for the whole project. The file is then imported by the central IT node to either the personal input folder or the shared input folder, but invisible until the output checker approves the input.
9.4 Responsibilities regarding output and input checking in the long term

In General, the output checking is performed by Eurostat or by the NSA providing the data as agreed in section 7 of the Guidelines to the Regulation (EU) No 557/2013. For the long term, the ESSnet DARA team discussed different scenarios how work could be distributed between Eurostat and AFs in the MS. They all have advantages and disadvantages. The different scenarios are described in more detail in Appendix 8. They include different potential stakeholders to become OC: either Eurostat, staff members of the local AF where the researcher is accessing or one country is responsible for a statistic and checks all output of that domain. Furthermore OCs from the countries whose data is used might be included as additional output checkers if they request this at the time the research proposal is voted on. There could also be a peer review process to double check the output. Any of the above mentioned groups can be the double checker. There could also be a version of non-mandatory double checks. Only when the first OC wants a second opinion, a call for a double check is sent. This system would then provide for a learning process, when the first OC asks for a double check to learn and confirm how output checking is done best, especially at the beginning of a research project.

If double-checking is implemented, it should also be done within the DARA network. The two OCs have to have a possibility to communicate with each other, either through a ticketing system or by writing text comments in the output folder and communicating via email.

To assure a competent output-checking, training for OCs is needed. In 2014 there will be a course on “output checking” in the European Statistical Training Programme. The ESSnet DARA team recommends these courses to be offered regularly.

In a best case, researchers should also be briefed before accessing the confidential microdata. A possible solution could be different e-learning tools, one focusing on output checking for staff members and another one addressing researchers.

9.5 Specific output checking guidelines

All results to be published or otherwise released, shall be checked to avoid disclosure of confidential data: tables, graphs, classification levels and systems, all different forms of presentation, linear and non-linear estimation, simulation, modelling, different types of developed analysis, particular indices and all (other) kind of econometric methods.

The general guidelines how output that has been generated with microdata should be checked are laid down in the following document:

The guidelines on output checking have also been published as chapter 6 “Data access issues” in the book by Hundepol et al. (2012)\(^5\).

To collect specific output checking information a form can be found in Appendix 9. It has to be filled out for every statistic available.

\(^5\) Hundepool, Anco; Domingo-Ferrer, Josep; Franconi, Luisa; Giessing, Sarah; Schulte Nordholt, Eric; Spicer, Keith; de Wolf, Peter-Paul (2012): Statistical Disclosure Control, Wiley Series in Survey Methodology, John Wiley & Sons Ltd, Chichester.
10 End of research projects

by Anja Hlawatsch and Julia Höninger

10.1 Rules regarding research project end

A research project ends after the elapse of the time period the data are requested for in the
research proposal. If a project is requested for a duration of less than five years, it can be
prolonged to a maximum duration of five years. An extension of project duration is only
possible while the project is still running; it cannot be requested when the project has already
ended.

It is a good service to notify researchers about the expiration date several weeks before a
research project will be ending.

10.2 Archiving and removing accounts

All syntax that has been written and results produced in the remote access system should be
archived for a certain number of years, ten years are suggested. If a breach of confidentiality
is suspected, the archived syntax, results, rejected and released output can be used to
investigate. The duration of time the documents are archived should be mentioned in the
form for research proposals.

On the central server, an internal area shall be devoted to the archiving of projects. Any
duplicates of the data set, including the work files of data sets, have to be deleted. The
archiving of the project disk space will be prepared by the SO. It is recommended that the
SO can move the files of the expired research project to the archive.

The researchers’ user accounts and the research project are to be removed from the user
account data base when a project has ended. Similarly, when members of staff leave the
Safe Centre, the central node has to be informed to set the accounts to inactive or delete
them.

DARA pilot:
During the testing phase of the pilot, no research project expired. Within the DARA pilot,
the local SO can set the accounts of researchers to inactive but he is not supposed to
delete them in the management application. The archiving of the project disk space can
be prepared by the SO.
# Appendix

## Appendix 1: Compliance checklist

<table>
<thead>
<tr>
<th>Item</th>
<th>Requirement</th>
<th>Compliance Statement (AF)</th>
<th>Compliant Y/N (Eurostat)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Organisational</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i)</td>
<td>Duly Designated Representative: the Safe Centre must have a named person who has overall responsibility for all aspects of access.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii)</td>
<td>Security Officer: the Safe Centre must have a named person who has responsibility within the NSA for information security to act as Security Officer for the DARA access.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iii)</td>
<td>The AF must demonstrate established measures to prevent and sanction violations of statistical confidentiality in the country where the AF are located.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iv)</td>
<td>Support Officers: The Safe Centre must name at least two persons who will be responsible for supervising, looking after and providing technical support to researchers whilst in the Safe Centre accessing DARA. CVs have to be provided for the named SO.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v)</td>
<td>Output Checker: The Safe Centre must name at least two persons who will be suitably trained to provide output checking. Alternatively a plan must be presented outlining how this expertise will be developed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vi)</td>
<td>All personnel working on the DARA Safe Centre must sign a confidentiality declaration and have necessary security clearance to the level required to work within the NSA.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vii)</td>
<td>All personnel working on the DARA Safe Centre within the AF should receive annual security awareness training.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>viii)</td>
<td>AF should have procedures in place for informing DARA when staff leave post such that their access rights can be removed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Safe Centre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i)</td>
<td>The Safe Centre must be located within the buildings of a NSA.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii)</td>
<td>Access to the buildings containing the Safe Centre must have full access controls.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iii)</td>
<td>Access is to be prohibited unless the researchers have pre-booked and have their name on the day’s access list.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iv)</td>
<td>Access is to be prohibited unless a valid photo ID is presented (passport, identity card, photo driving licence will be acceptable).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v)</td>
<td>The Safe Centre must keep a record of all researchers’ attendance. These must be transferred to Eurostat electronically monthly so that they may be compared to usage of DARA. The Safe Centre must</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>Requirement</td>
<td>Compliance Statement (AF)</td>
<td>Compliant Y/N (Eurostat)</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>---------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td>provide Eurostat with an annual report on usage and research outcomes, along with reports on any incidents as soon as they are discovered and subsequently resolved.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vi)</td>
<td>Buildings must have adequate measures for emergency evacuation and sufficient signage and support for researchers whilst in the Safe Centre.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vii)</td>
<td>Buildings must have adequate facilities for researchers including: breakout area with toilets, smoking area, access for disabled persons. If access to the internet is provided for researchers in the breakout room (through wi-fi or a dedicated terminal), the Safe Centre must describe arrangements in place to ensure that such access does not affect security of the DARA system.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>viii)</td>
<td>The Safe Centre must have an appropriate mortice lock for out of hours, and a punch or key code lock for office hours. Access to the room must be to named support staff from the Safe Centre and site Security personnel only.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ix)</td>
<td>The Safe Centre must have adequate temperature control to provide a safe and productive environment. It must conform to EU and National legislation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x)</td>
<td>The Safe Centre should contain partitions between workstations so that researchers may not casually oversee each others’ work.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Support

i) Researchers should be provided with a tour of the facilities and made aware of any emergency procedures.  

ii) Researchers should have access to a telephone that permits communication with the SO. The SO should be available to respond within 10 minutes throughout the day. Arrangements for lunch times should be agreed with researchers and cover for vacation periods should be assured.  

iii) The SO should check on the researchers at least every hour to ensure that they have all that they need. Ideally the SO will be located close to the Safe Centre, although it is not necessary to be in constant vigil.  

iv) Support Officers should have the necessary skills to support researchers in a Research Data Center. For NSAs applying to be AF who don’t have an RDC, a plan for developing these skills must be presented. Experience of the SO working with scientific-use files should be provided.  

v) The SO should have knowledge of the data being used during the research, and have knowledge of the statistical software that is being used so that basic help can be obtained. The NSA should have available detailed technical support available to the SO for more detailed questions on the statistical software. Evidence of this will need to be provided during the accreditation process – as a minimum, questions should be answered within four hours.  

vi) The SO may or may not be qualified as OC. If they are providing both roles, then this should be made clear during the accreditation process.  


### 4. IM environment

<table>
<thead>
<tr>
<th>Item</th>
<th>Requirement</th>
<th>Compliance Statement (AF)</th>
<th>Compliant Y/N (Eurostat)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i)</td>
<td>The Safe Centre must provide Eurostat with a named person responsible for the IM environment that will be connecting to DARA.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii)</td>
<td>The IM environment must be accredited in accordance with the NSA’s policy and procedures to ensure that there is no access to DARA apart from through the designated terminals in the Safe Centre following agreed authentication. A risk assessment should be provided.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iii)</td>
<td>The Safe Centre must sign the Code of Connection, provided in Appendix 2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iv)</td>
<td>The Safe Centre must protect the DARA system from unauthorised physical access, this includes power and network cables. Consideration should be given to the protection of cables against accidental or deliberate damage, interference or interception.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v)</td>
<td>Equipment connected to DARA must have appropriate maintenance and receive up to date software patches and updates. Confirmation of this must be provided by the AF.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vi)</td>
<td>Any equipment connected to DARA must be locked down to ensure that it has no connection to the internet, local network, or printers, and that it is not possible to connect any devices to the external interfaces, connect a device to the local network, install or remove hardware or software, or boot the access point from floppy, CD-ROM, DVD-ROM or any other media. A test schedule must be provided dealing with: initial set up, regular check and checks following system changes / updates.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vii)</td>
<td>Depending on the DARA solution implemented, the Safe Centre must demonstrate that they have adequate facilities for protecting and storing any cryptographic devices or key generators used during the authentication process.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>viii)</td>
<td>If NSA has policies for home and mobile working, these must not apply to DARA access. This includes wi-fi within the NSA premises.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ix)</td>
<td>The Safe Centre must provide evidence of the physical protection of equipment to ensure that it cannot be removed from the premises. As a minimum, this could be an inspection of hardware by the SO prior to the researchers leaving.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x)</td>
<td>A log of all reported system faults must be maintained at the Safe Centre and shared monthly with Eurostat.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>xi)</td>
<td>In the event of any suspected breach of information from within Eurostat, the named person responsible for the NSA IM environment will be informed and will be required to co-operate with the subsequent investigation from Eurostat.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2: Draft application form for Access Facilities

Annex to the Guidelines of Regulation (EU) No 557/2013
Application form for Access Facilities

This application form is intended for National Statistical Authorities (NSA) wishing to be accredited as Access Facility to provide access to confidential data of European statistics for scientific purposes via remote access in a Safe Centre.

As a first step, please send this form electronically to ESTAT-FACILITIES-ASSESSMENT@ec.europa.eu

The information provided in the application form will be examined by Eurostat, which will elaborate a report on the assessment including a recommendation on the type of confidential data to which access can be provided by the Access Facility. The report will be made available to the NSAs. If the facility has fulfilled all the assessment criteria, Eurostat will ask the ESSC for an opinion. NSAs might accept or refuse access to their data via given Access Facilities. If access is refused, the data of particular country is removed from the datasets. Accredited Access Facilities may provide access only to the data of those countries that do agree.

The following criteria will be taken into account when deciding on the status of the facility:

1) the purpose of the facility;

2) the established record or reputation of the facility as a body promoting research and access to confidential data for scientific purposes;

3) the internal organisational arrangements for:
   - competence and experience of staff providing information about the survey and statistical disclosure control methods; commitment of staff to respect statistical confidentiality;
   - full compliance with and follow-up of legal procedures governing access to confidential data for scientific purposes;
   - established provisions for reporting, accountability and auditing;
   - established measures to prevent and sanction violations of statistical confidentiality in the country where the facility is located;

4) the safeguards in place to ensure security of the data.
The application form should be accompanied by the annex A1 – A7 (see page 10), initialled on each page and signed on the last page by duly designated representative of the Safe Centre within the facility. All the documents should be sent to:

For the attention of Ms Dominique Reuter-Wagner
European Commission
Eurostat — Unit B1
Jean Monnet Building
L-2920 Luxembourg

Applicants will be notified by email of the outcome of the assessment. If ‘Access Facility’ status is granted, a contract between Eurostat and the Access Facility will be signed and the name of the Access Facility will be published on the Eurostat website.

The duly designated representative must immediately inform Eurostat of any changes to the information provided in this application form. A site visit may be requested by Eurostat to check the details in the application form.

The application form has to be re-submitted at Eurostat’s request.

Processing of personal data is protected in accordance with Regulation (EC) 45/2001. All information collected will be processed by Eurostat for the sole purpose of verifying the applicant’s compliance with the Regulation. All relevant questions must be answered. Failure to answer all relevant questions will result in refusal of the application. Applicants have the right of access to, and the right to rectify, the data concerned. Applicants have the right to have recourse at any time to the European Data Protection Supervisor.
1. Identification and roles within the facility

1.1 General information
Official full name of the facility:
Short name – acronym:
English name:
Postal address:
Web address:
Country:

1.2 Legal status within the European Statistical System
(in accordance to Regulation (EC) No 223/2009, Article 5(2) and the published list on the Eurostat website):

☐ National Statistical Institute (NSI)
☐ Other National Statistical Authority (NSA)

1.3 Duly designated representative of the Safe Centre within the facility:
Name:
Position:
Telephone:
Email:

1.4 Prime duly designated Support Officer\(^6\) of the Safe Centre within the facility:
Name:
Position:
Telephone:
Email:

1.5 Second duly designated Support Officer of the Safe Centre within the facility:
Name:
Position:
Telephone:
Email:

\(^6\)A Support Officer of the Safe Centre is the main contact person for researchers and all other people involved in the process.
1.6 Prime duly designated Output Checker of the Safe Centre within the facility:

Same person as prime Support Officer  □ yes  □ no, details:
Name: 
Position: 
Telephone: 
Email: 

1.7 Second duly designated Output Checker of the Safe Centre within the facility:

Same person as second Support Officer  □ yes  □ no, details:
Name: 
Position: 
Telephone: 
Email: 

1.8 How many additional Support Officer / Output Checker are available in the Safe Centre within the facility:

.................. other Support Officer.

.................. other Output Checker.

1.9 Duly designated IT Manager within the facility, who will be responsible for the connection to the remote access system:
Name: 
Position: 
Telephone: 
Email: 

1.10 Duly designated Security Officer within the facility, who is responsible for information security:
Name: 
Position: 
Telephone: 
Email: 
2. Purpose of and internal organisational arrangements in the facility

2.1 Main purpose and activity of the facility:

Please describe the established provisions for reporting, accountability and auditing in the facility:

2.3 Please describe how research is promoted and how access to data for scientific purposes is organised in the facility.

3. Existing of national provisions on sanctions for violation of statistical confidentiality

3.1 Please describe, in a comprehensive manner, the frame of national provisions existing in the MS where the facility is located (type of sanctions, kind of penalties, if there are differences between data concerning physical and legal persons).

3.2 Please describe, in case a violation of statistical confidentiality occurs in the MS where the facility is located, the sanctions applicable to a researcher who has been given access to confidential data in accordance to Regulation (EC) No 223/2009 and Regulation (EU) No 557/2013?
3.3 Please describe who, in case of illicit use or disclosure of the data, is responsible for monitoring and prosecute a researcher who has been given access to confidential data in accordance to Regulation (EC) No 223/2009 and Regulation (EU) No 557/2013?

4. Premises of the Safe Centre

4.1 Please describe the location of the Safe Centre within the premises of the facility.

4.2 Please describe how the technical and organisational requirements for accessing the Safe Centre are implemented within the premises of the facility.

4.3 Please describe the breakout facilities of the premises enclosing the Safe Centre, including the arrangements in place to ensure that those do not affect the security of the Safe Centre. Consider e.g. access to the internet, telecommunication, toilets, smoking area, access for disabled persons.

4.4 Please describe the emergency support facilities of the premises enclosing the Safe Centre.
5. Safeguards in place within the Safe Centre

5.1 Please describe how the Safe Centre is physical and virtual protected from removing data or other unauthorised access?

5.2 Please describe the Safe Centre's computer system. How is the computer network isolated from the rest of the organisation and the outside world?

5.3 Please describe the safeguards in place for safekeeping of data in the Safe Centre's computer systems:

6. Support Officer

6.1 Please describe the above named Support Officer's competence and experience in managing access to confidential data for scientific purposes:

6.2 Please describe the above named Support Officer's knowledge about the relevant European surveys and the statistical disclosure control methods provided in the Safe Centre:
6.3 Please describe the above named Support Officer’s knowledge of the relevant statistical software packages:

7. Output Checker

7.1 Please describe the above named Output Checker’s knowledge about the relevant European surveys and the statistical disclosure control methods provided in the Safe Centre:

7.2 Please describe the above named Output Checker’s knowledge of the relevant statistical software packages:

8. Additional information

8.1 Additional comments — free text
I, the duly designated representative of the Safe Centre within the facility applying as Access Facility, hereby certify that the information given in this questionnaire is complete, accurate and correct and that any change(s) will be reported immediately to Eurostat. I understand that Eurostat is authorised to check the accuracy of the information given in this questionnaire at any time. I understand that Eurostat may require more information, if necessary.

I confirm that my organisation submits this request in order to be accredited by Eurostat as an Access Facility. That accreditation will allow my organisation to provide access to confidential data for scientific purposes. I am aware that in case my organisation is accredited as Access Facility, the name of my organisation will be published on Eurostat website.

Furthermore, I commit myself to taking and maintaining all necessary measures in compliance with the above requirements.

At: (please state location)       Date: □□□□20□□
Name:

Position:

Signature:

Eurostat use only:
Date received: ........................................
Identification number: ..............................
Application accepted □ Signature ........................ Date: ..............................
Application refused □ Signature ........................ Date: ..............................
Comments:
Annexes

A1. Legal act creating the facility and/or Articles of incorporation

A2. Organisation chart

A3. Curriculum vitae of each Support Officer

A4. Individual confidentiality declarations to be signed by any personnel working in the Safe Centre of the facility

A5. Compliance check list

A6. Code of connection

A7. Terms of use to provide confidential data for scientific purposes (initialled on each page)
## Accredited Network Configuration Statement

Applicant’s IT System / Network Administrator to complete and sign.

<table>
<thead>
<tr>
<th>Please provide your network accreditation details:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accrediting body:</td>
</tr>
</tbody>
</table>

Please provide your Gateway IP Address:

*Note that servers must have static IP addresses, even if DHCP is used.*

<table>
<thead>
<tr>
<th>Gateway IP Address:</th>
</tr>
</thead>
</table>

Please confirm you are able to upload user certificates within browsers and are able to install CA certificates supplied by Eurostat.

Y / N

I confirm that:

- in addition to the controls stated within this application the host organisation will continue to comply with existing code of connection (CoCo) requirements for its accredited network.

- all connecting network, system and application and infrastructure is located in secure accommodation.

- where appropriate, an IT Health Check / audit is performed every 12 months by an independent service provider

- hosts are scanned for the presence of vulnerabilities at least once every 3 months and a patch management policy is employed.

- appropriate measures are in place to identify and isolate malicious software, including - as a minimum - viruses, macros, dangerous file types, mobile code and spyware.

- content analysis of all incoming and outgoing data is performed at the gateway.

- each user of the accredited network is allocated a unique user ID and password and authorisation procedures are used to control access to network.

- all local and remote attached devices and network infrastructure supporting devices are security hardened.

- restricted information stored on a remote device, removable media and data in
transit, is protected using encryption to an approved standard.

- personal firewalls will be installed, enabled and subject to configuration management for all remote working devices.

- I understand that Eurostat reserves the right to run host-checking tools on connections made to DARA.

<table>
<thead>
<tr>
<th>Signed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name and date</td>
<td></td>
</tr>
</tbody>
</table>

**Applicant to forward completed form to Eurostat**

**Eurostat Official Use Only**  *Accept / Reject with corrective actions*
## Appendix 4: Scenarios for the Research Management System

<table>
<thead>
<tr>
<th></th>
<th>Option A.</th>
<th>Option B.</th>
<th>Option C.</th>
<th>Option D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong></td>
<td>RA only with no documents attached</td>
<td>RA only with full documentation uploaded</td>
<td>All modes with no documents attached</td>
<td>All modes with full documentation uploaded</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>System is used only for remote access research projects; documentation (i.e. contracts, confidentiality undertakings) is not stored in the system, only supplementary information</td>
<td>System is used only for remote access research projects; full documentation is stored in the system and can be extracted from it</td>
<td>System is used not only for remote access but all access modes and their research projects; documentation is not stored in the system, only supplementary information</td>
<td>System is used not only for remote access but all access modes and their research projects; full documentation is stored in the system and can be extracted from it</td>
</tr>
<tr>
<td><strong>Pros</strong></td>
<td>- Support for all remote access research projects</td>
<td>- Full support for all remote access research projects with access to all documentation</td>
<td>- Support for all research projects</td>
<td>- Full support for research projects with access to all documentation</td>
</tr>
<tr>
<td></td>
<td>- Probably easiest option to implement</td>
<td>- Standard documentation schemata can be generated by the system</td>
<td></td>
<td>- Standard documentation schemata can be generated by the system</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Provides basis for transparent and standard management for all research projects related to access to ESS confidential data</td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td>- Documents cannot be accessed at one place</td>
<td>- Unknown relationship with administration of research projects of other access modes</td>
<td>- Documents cannot be accessed at one place</td>
<td>- Close coordination with administrative processes for other modes is a necessity</td>
</tr>
<tr>
<td></td>
<td>- Unknown relationship with administration of research projects of other access modes</td>
<td></td>
<td>- Close coordination with administrative processes for other modes is a necessity</td>
<td>- Partly outside of scope of DARA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Partly outside of scope of DARA</td>
<td>- Probably the most difficult option to implement</td>
</tr>
</tbody>
</table>
Appendix 5: Principles for research project administration

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The system should be stored centrally</td>
</tr>
<tr>
<td>2</td>
<td>A common interface has to be created where information on all projects (rejected, finished or ongoing) can be collected</td>
</tr>
<tr>
<td>3</td>
<td>Authorised personnel at all AF shall use the same infrastructure to record research project information, using the same interface and standards</td>
</tr>
<tr>
<td>4</td>
<td>Availability of the system and access to it from authorised personnel should be guaranteed any time</td>
</tr>
<tr>
<td>5</td>
<td>In order to assure that staff of different AF can use the system at the same time, instant system update has to be guaranteed</td>
</tr>
<tr>
<td>6</td>
<td>The system should impose as low an administrative burden on users as possible, i.e. researchers could fill out their applications online, so that all relevant data are stored in the system and can export a pdf-file for submitting their signature</td>
</tr>
<tr>
<td>7</td>
<td>Important and relevant criteria are to be recorded in the system, unimportant or irrelevant content should be omitted</td>
</tr>
<tr>
<td>8</td>
<td>The system should be as user-friendly as possible (access conditions, usability, layout, etc.)</td>
</tr>
<tr>
<td>9</td>
<td>The users shall be supported by automation, pre-defined lists wherever possible (drop-down menus, nomenclatures, standard code lists, etc.)</td>
</tr>
<tr>
<td>10</td>
<td>Pre-defined and custom reports can be generated from the system</td>
</tr>
<tr>
<td>11</td>
<td>Archiving project information is to be assured as well</td>
</tr>
<tr>
<td>12</td>
<td>Containing information both on approved and rejected research proposals</td>
</tr>
<tr>
<td>13</td>
<td>Providing access to all ESS members. All members should have access to content that is in connection with them (filter on user profile)</td>
</tr>
<tr>
<td>14</td>
<td>Being able to assign coordinators (AF) to projects. With appointed coordinators, the issue of &quot;more researchers working on the same project but at different Access Facilities&quot; problem is addressed as such coordination will be the task of the appointed AF</td>
</tr>
<tr>
<td>15</td>
<td>Storing contact details of all researchers, not just the principal researcher of each research project</td>
</tr>
<tr>
<td>16</td>
<td>Having functionality for history tracking. That means that every activity in the system (modifications, registration of new information, etc.) is logged and if information in the system needs to be overwritten, the previous state will also be stored</td>
</tr>
<tr>
<td>17</td>
<td>Being able to link research projects. If a previous research project is somehow linked to another ongoing project or received research proposal, a link between them should be made, indicating the relationship between the two</td>
</tr>
</tbody>
</table>
Appendix 6: Variables & Databases

**RESEARCH PROPOSAL**
- Reception of research proposal
  - Date of reception of research proposal
  - Unique ID of the research proposal/project
  - Project coordinator RDC [AAPDE]
  - Type of access
  - Name of research entity [REDE]
  - Contact information of research entity [REDE]
  - Names and contact information of researchers [RDE]
  - Name and contact information of the principal researcher [RDE]
  - Name of microdata to be accessed [MLDE]
  - Reference time of microdata to be accessed
  - Type of the research proposal
  - Short description of the research proposal
  - >> Upload: research proposal

**APPROVAL OR REFUSAL**
- Approval
  - Date of approval of research proposal
  - ID of contract
  - >> Upload: contract

- Refusal
  - Date of refusal of research proposal
  - Reason for refusal

**RESEARCH PROJECT**
- Remote Access
  - IDs and names of access points [AAPDE]
  - Contact details of the access points [AAPDE]
  - Starting date of the research project
  - Finish date of the research project
  - Duration of the project
  - Name of output checkers [OCDE]
  - Details of output checkers [OCDE]
  - Number of outputs checked
  - Availability of archived research outputs
  - Cost of the research project

- Access in Safe Centre
  - IDs and names of access points [AAPDE]
  - Contact details of the access points [AAPDE]
  - Starting date of the research project
  - Finish date of the research project
  - Duration of the project

- Release on CD or DVD
  - Starting date of the research project
  - Finish date of the research project
  - Duration of the project
Appendix 7: Screenshots from research administration system for SD Box solution

Screenshot 1 – Boxes and virtual servers
Screenshot 2 – Contracts
Screenshot 3 – Data sources
Screenshot 4 – Projects
### Appendix 8: Scenarios for Output checking

<table>
<thead>
<tr>
<th>No.</th>
<th>Possible scenarios</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Local AF where researcher is accessing checks the output</td>
<td>- Workload is spread over several actors.</td>
<td>- Misalignment of incentives, AF do not only provide support and access to researchers, but additionally check output. → Maybe MSs do not want to participate in creating an AF with remote access to European statistics.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Responsibilities are clear.</td>
<td>- Maybe MSs do not want the output including their own data to be checked by another AF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- If access to ESSnet DARA is costly, AF that earns money has to provide the</td>
<td>- Every AF staff has to have the knowledge of the different disclosure control methods in different statistics.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>manpower and check the output.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Short channels of communication between AF and researcher (problems could</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>be addressed easier)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Eurostat checks the output</td>
<td>- Responsibilities are clear.</td>
<td>- All workload is on Eurostat.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Increases incentives for MS to provide an access facility.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Eurostat already has the expertise.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Maybe the MSs trust more in the output checking of Eurostat.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>One country is responsible for a statistic and checks all output of that domain</td>
<td>- Work is distributed over several agents.</td>
<td>- Responsibility for a statistic is a burden, maybe MSs do not volunteer for that.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- The country will be the expert; differences between outputs can be taken into account.</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Possible scenarios</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3a</td>
<td>Every country is on two days of the month responsible for all outputs that are</td>
<td>- Workload is distributed</td>
<td>- Distribution of work is not along any specialization.</td>
</tr>
<tr>
<td></td>
<td>requested for release</td>
<td></td>
<td>- Harder to manage, every MS has to have access to all accounts, no overview of all outputs within one research project.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Too much administration with little competence for the output checking.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Unrealistic scenario.</td>
</tr>
<tr>
<td>4</td>
<td>Every country whose data is concerned checks the respective part of the output</td>
<td>- Maybe MS will request this scenario to retain control.</td>
<td>- Organizing the process is more difficult as many countries can be involved.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Researchers have to wait longer for their output.</td>
</tr>
<tr>
<td>5</td>
<td>There is a peer review process to double check the output. Any of the above</td>
<td>- Increase trust among MSs. Output checkers learn by seeing how others do it</td>
<td>- More difficult and time-intensive for MSs and researchers.</td>
</tr>
<tr>
<td></td>
<td>mentioned groups can be the double checker.</td>
<td>Output checking should not drift apart.</td>
<td>- Then describe how the output with the first output check and remarks is transferred among all in the check involved. How is the</td>
</tr>
<tr>
<td>No.</td>
<td>Possible scenarios</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>6</td>
<td>Countries whose data is concerned want to check output only for single research projects that they think especially critical.</td>
<td>- At least not a general veto, for some research projects there is a delegation process documented? In what order are the checks executed, simultaneously or does the second person only start when the first is finished.</td>
<td>- It takes longer, the process is more complicated. Otherwise want could ask them to be the output checker for all countries or exclude the data of that country from the analysis (if that is acceptable for the researcher).</td>
</tr>
<tr>
<td>7</td>
<td>Specific Guidelines for every statistic if there are specific rules that go beyond the principles based approach. Example: One statistic, i.e. agriculture statistics, could have a different minimum cell count.</td>
<td>- This could make objections from single countries obsolete.</td>
<td>- It is another document to be produced and updated. It would be clearly the job of the data manager to compile the document together with metadata.</td>
</tr>
</tbody>
</table>
Appendix 9: Form for specific output checking guidelines

Specific Guidelines on Output Checking for

Name of statistic: _________________________________

1. Specific characteristics of this statistic

Definitions and terms used in the statistic can be found on this website:
Please insert a LINK to a document or a webpage where metadata can be found

- 1.1. Which of the variables will be made available in the Access Facility (AF)?

Please include a link to a metadata document that contains a list of variables. You can enumerate variables that are not accessible in the AF. Otherwise please include a document as annex to this specific guideline or justify why not.

- 1.2. Which reference years will be made available through the AF?

- 1.3. In this statistic, are the respondents asked to report on a yearly, quarterly or monthly basis?

☐ yearly ☐ quarterly ☐ monthly

1.4. Are there different levels of units?

For example in the university statistics one can look at students or universities, in household surveys persons, households or families can be analysed, in business statistics local units and enterprises can be analysed.

☐ no ☐ yes, which ones? .........................................................

1.5. What is the most detailed regional level that shall be included in the data set? Which regional levels shall be available?

- 1.6. Are further steps of data preparation necessary before making the data available?

i.e. exclude certain groups, duplicates, because they are always excluded in reports.
2. **SDC rules to be applied**

2.1. Which disclosure risks are to be avoided for this statistic?

Results that would permit

- [ ] attribute disclosure and/or
- [ ] the identification of a microdata unit

shall not be disseminated.

2.2. Is there a threshold rule / a minimum frequency?

- [ ] yes (with \( n \geq \square \))
- [ ] no

Shall it be applied to

- magnitude tables? [ ] yes [ ] no
- frequency tables? [ ] yes [ ] no

Specific to individual Member States

BE:

DE:

2.3. Shall group disclosure be prevented?

- [ ] yes [ ] no

Specific to individual Member States

BE:

DE:

2.4. Is it necessary to protect dominant units from disclosure? If yes, how are dominant units detected?

- [ ] yes
- [ ] no
- [ ] only categorical variables, dominance rule unapplicable

If yes, which rule and which parameters apply?

- [ ] \((1,k)\)-rule with \( k = \square \)
- [ ] \((2,k)\)-rule with \( k = \square \)
- [ ] \(p\%\)-rule with \( p\% = \square \% \)
- [ ] other: ............................................................................................................................
  .................................................................................................................................

Specific to individual Member States

BE:

DE:

---

7 If a dominance rule applies for this statistic, researchers shall produce the sum, the largest and the second largest contributor to every result. These values will be used to calculate whether there is a dominant unit, they will not be released.
2.5. Which method shall be applied?

☐ cell suppression

If yes, how are cells for secondary suppression chosen? ........................................
...................................................................................................................
...................................................................................................................
...................................................................................................................
...................................................................................................................

☐ perturbative methods

pretabular: ☐ yes ☐ no
if yes, how: ............................................................................................................

posttabular: ☐ yes ☐ no
if yes, how: ............................................................................................................

☐ traditional pretabular method, details:
..............................................................................................................................

(Cell suppression is by far the most frequently used method, however for some statistics
perturbative methods may be used.)

2.6. Are there certain tables or variables that are treated differently?

.......................................................................................................................................
.......................................................................................................................................
.......................................................................................................................................
.......................................................................................................................................
.......................................................................................................................................
.....................................................................................................................................

3. Exceptions/ Application to different types of analysis

Are some types of analysis treated differently than described in the „Guidelines on Output
Checking“? Types of analysis are called “classes” in the guidelines on output checking. For
example are there additional risks when regressions are estimated besides the ones
described in the guidelines?

☐ no exceptions to the general guidelines on output checking
☐ the following exceptions apply:
....................................................................................................................................
....................................................................................................................................
....................................................................................................................................
....................................................................................................................................
....................................................................................................................................
....................................................................................................................................