



Project N°: 262608



ACRONYM: **Data without Boundaries**

DELIVERABLE D5.1

*Report on concept for and components of
European Service Centre for official statistics*

WORK PACKAGE 5

Servicing European researchers in the use of OS microdata

REPORTING PERIOD:	From: Month 1	To: Month 18
PROJECT START DATE:	1st May 2011	DURATION: 48 Months
DATE OF ISSUE OF DELIVERABLE:	April 30th 2012	
DOCUMENT PREPARED BY:	Partner Numbers 5, 25,1	Partners Names GESIS, CED, CNRS-RQ

Combination of CP & CSA project funded by the European Community
Under the programme "FP7 - SP4 Capacities"
Priority 1.1.3: European Social Science Data Archives and remote access to Official
Statistics

Abstract¹

This report develops a concept for a European Service Centre for official statistics (ESC-OS) which will function as a research infrastructure institution for European official statistics microdata. This institution will be established on the basis of the existing CESSDA network of European data archives and cooperate tightly with the European Statistical System (ESS), which is led by Eurostat and includes the National Statistical Institutes, the National Statistical Administrations and Central Banks. The ESC-OS primary objective shall be to promote the scientific use of European official statistics microdata by working towards an improved data access infrastructure and by providing comprehensive metadata on integrated European statistics as well as national statistics from Europe.

This report will be structured as follows: In Chapter 1 the objectives of the ESC-OS will be laid out, subsequently the underlying rationale and the scope of its services will be detailed. In Chapter 3 we will elaborate on the tasks the ESC-OS shall fulfil, this includes establishing an online platform hosting comprehensive metadata and other services for European OS microdata, as well as promoting the use of OS microdata and assisting the ESS in coordinating data access. After the description of the ESC-OS' functionality we will go on to discuss possible implementations of an ESC-OS both with regards to its functionality and its institutional integration within the CESSDA network. Finally we will lay out how the envisioned ESC-OS should seek to work together with other stakeholders in order to improve data access and documentation.

¹ This report was prepared by Alexander Mack, Christof Wolf, Albert Esteve and Roxane Silberman. The authors gratefully acknowledge feedback from Markus Quandt and Matt Sobek. Preliminary versions of this report have been discussed at the First European Data Access Forum which took place in Luxembourg on March 27-28, 2012 and at the meeting of the CESSDA general assembly which took place in Gothenburg March 20th 2012.

Table of contents

1. Objectives
2. Rationale and scope
3. Tasks
 - 3.1 Online platform
 - 3.1.1 Metadata database
 - 3.1.2 Online library
 - 3.1.3 Routines
 - 3.1.4 Online community
 - 3.1.5 Dissemination of integrated European census data
 - 3.2 Promote use of OS data
 - 3.2.1 Workshops
 - 3.2.2 User Conferences
 - 3.2.3 Incentivizing research
 - 3.3 Coordinate data access
 - 3.3.1 European Data Access Forum
 - 3.3.2 Supporting cross border access
4. Implementation
 - 4.1 Institutional framework
 - 4.2 Online platform for European official statistics
 - 4.3 Service unit for official statistics within the CESSDA-ERIC
 - 4.4 Recommended initial implementation
5. Cooperation
 - 5.1 Eurostat
 - 5.2 NSIs and other national data providers
 - 5.3 Collaboration with open data initiative projects
 - 5.4 Global cooperation
6. Conclusions

1. Objectives

The overarching goal of the European Service Centre for official statistics (ESC-OS) will be to remove barriers to access to official statistics data² in Europe and stimulate research in order to make better use of the wealth of high quality data available. To this aim the ESC-OS will provide researchers with comprehensive services for European official statistics microdata. This will entail extensive data documentation and tools for data preparation as well as training in the use of OS microdata. Moreover the ESC-OS shall seek to facilitate exchange among researchers, data producers and data archives throughout Europe and work towards reducing boundaries to data access.

The primary task of the ESC-OS will be to establish an online platform which offers comprehensive documentation for OS microdata from Europe. This metadata database will inform interested researchers on available OS data sources and how these can be accessed and utilized for social research. Documentation will cover integrated European statistics as provided by Eurostat as well as national OS data from EU-countries. Another major aspect of this online platform will be to incorporate the existing Integrated European Census Microdata (IECM) database which was established by the Centre d'Estudis Demogràfics (CED) in Barcelona in close collaboration with the Integrated Public Use of Microdata Series-International project (IPUMS) lead by the Minnesota Population Center. The IPUMS/IECM disseminates integrated and anonymised European census microdata free of charge to registered users and provides extensive metadata for these censuses both at the study and variable level.

Another major objective of the ESC-OS will be to improve the European dialogue between data producers, distributors and users. The ESC-OS online platform will attempt to contribute to this goal by building a community of contributors and serving as an important forum for advice, discussion and instruction on all matters pertaining to European OS data. Beyond this user community a constant dialogue between producers and users shall be institutionalized by establishing regular conferences, which will bring together face to face representatives from NSIs, Eurostat, national data archives and the research community. Such a dialogue is invaluable for all sides as it can help to improve official statistics by calling attention to current or regional social issues, improving documentation and encouraging the involvement of the researchers from the different countries in the harmonization process.

A major barrier which must be overcome to ease access to OS data within Europe is the issue of cross border access to highly confidential national microdata. Today the major problem in this respect is that access to this kind of data does not follow common rules throughout Europe. While some countries are fairly open with their OS microdata others tend to be overprotective of their data. A commonality throughout most of Europe is that highly confidential OS microdata can only be accessed within the country itself. While researchers can oftentimes visit data providers in order to access microdata in a country, such a solution is costly and not always feasible when conducting comparative research. Thus the ESC-OS should aim to find ways to ease cross border access to OS data and continue the work which is currently conducted within the Data without Boundaries (DwB) project on remote access throughout Europe. Beyond assisting the partners of the European Statistical System in implementing and maintaining programs granting cross border access to OS data via safe data centres or remote access the ESC-OS should also seek to work towards the institutionalization of a common framework for researcher accreditation throughout Europe.

This report develops an idea of what a European Service Centre for official statistic could look like and what role it could play within the European Statistical System. However, we are aware that it is rather unrealistic that all the functions we propose here can be implemented immediately. Section 4 will thus outline two possible implementations of an ESC-OS. The first type is what we would consider the 'bare minimum' in regards to functionality and institutional integration of an ESC-OS

² We use the term official statistics to refer to all statistical data produced by government agencies.

and is essentially limited to an online platform with decentralized maintenance and limited user support. The second type represents our vision of what the ESC-OS should look like in the future and develops the idea of a service unit for official statistics within the CESSDA framework. Additionally we propose a possible set up for the initial implementation of the ESC-OS from which it could be developed further.

2. Rationale and scope

The vast majority of research conducted in the social sciences today is based on secondary data. While many research endeavours could benefit from the application of OS data they are currently vastly underutilized. The most heavily used types of data are taken from scientific surveys. While this data is tailored to researcher’s needs by providing extensive documentation and ready-to-use datasets, it is oftentimes inferior to official statistics data. For one sample sizes from official statistics tend to be much larger than in scientific surveys. Furthermore integrated European Datasets such as the Labour Force Surveys (EU-LFS) or the Statistics on Income and Living Conditions (EU-SILC) offer unmatched possibilities for comparison of EU member states both cross-sectionally and over time. Thus the primary agenda of the ESC-OS will be to inform about microdata from official statistics and make these data more accessible to the research community.

The services offered by the ESC-OS will cover three different types of European microdata: national OS microdata; integrated European microdata as provided by Eurostat and integrated European Census microdata as provided by the IECM. Table 1 details the services which will be offered for the different data types.

Table 1: Overview of services offered by the ESC-OS

Data Type	Data Documentation	Routines	Data Dissemination
National Microdata	study level documentation	Routines for integrating national datasets Routines for operationalizing common social scientific concepts	No ^a (assist ESS in cross border access)
Integrated Eurostat Microdata	study and variable level documentation	Routine generator for custom datasets Routines for operationalizing common social scientific concepts	No ^b
Integrated Census Microdata	study and variable level documentation	IECM system generates custom datasets Routines for operationalizing common social scientific concepts	Yes

a In some countries dissemination of national OS data is carried out by CESSDA partners.
b There is the option that ESC-OS will support ESS/Eurostat in disseminating OS data if this is desired.

Microdata are generally classified as data referring to units of a population. This includes individuals, households, families or businesses. Thus the ESC-OS services’ shall apply to data from such diverse sources as household surveys, administrative data from registers, educational surveys and establishment data as well as fiscal microdata held by the central banks. These include microdata provided by the National Statistical Institutes (NSIs) as well as those coming from National Statistical Administrations (NSAs). One major issue for the future will be to cover the administrative data bases which will be increasingly utilized. Depending on the organization of the statistical system in each country these administrative microdata are more or less coordinated by the National Statistical Institute and may in some cases be held by government agencies with mixed statutes (for instance

administrative microdata on social insurance, unemployment, retirement etc.). While it is envisioned that the integrated European Statistics will be expanded and cover the major national statistical surveys, this is not the case for many administrative microdata which provide rich resources for comparative research in Europe. In addition, in some cases, national statistical surveys may include more information and may be more detailed than the Integrated European microdata. Therefore, the ESC-OS should provide accurate information about the differences between integrated European microdata provided by Eurostat and these national microdata.

One of the most pressing concerns currently is the access to integrated European data such as the EU-LFS, EU-SILC or the Structure of Earnings Survey (SES). Currently Eurostat disseminates scientific use files with anonymised microdata from these programs to research institutions and provides basic documentation. These services are an essential first step, but given the complexity of the data, they still leave the individual researcher with a large amount of work before the actual data analysis can begin. The ESC-OS aims to take this second step for researchers by servicing the research community with more extensive documentation, further and more flexible tools to read data into common statistical packages, and by offering training and establishing a platform for interaction with other users and data producers. Not only do these services generate added value by centralizing elementary tasks of every research endeavour and pooling existing knowledge, they also aim at promoting the use of official statistics data by informing and disburdening researchers interested in using official statistics data.

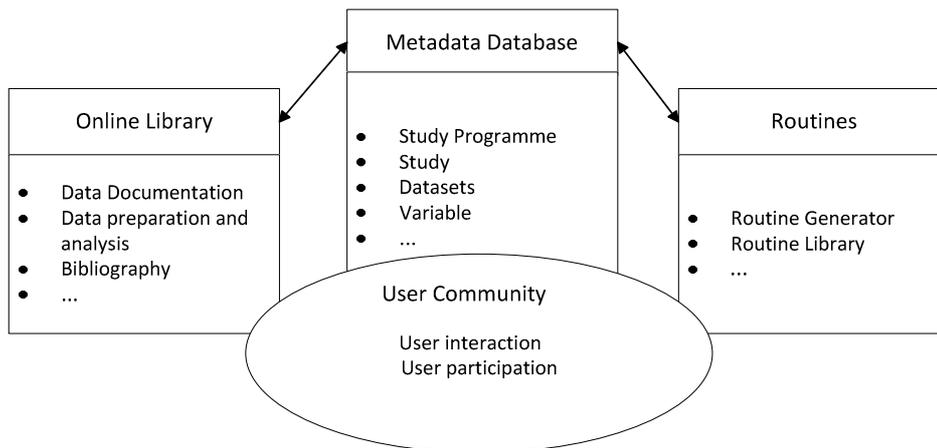
Authorities invest heavily to ensure data from official statistics can reliably inform politics and the public about the economy, demographics and societal trends. While this is the primary objective of data derived from official statistics, the full potential of the data is rarely realized. An increased application of official statistics in scholarly research can help to advance researchers understanding of social and economic processes and this knowledge in turn can help to inform decision makers, thus amplifying the utility of official statistics.

3. Tasks

3.1 Online platform

The most important service offered by the ESC-OS will be to host a webpage which will offer extensive metadata, information on access conditions, links to access providers as well as tools for data preparation and analysis for OS microdata. Additionally it will also handle the dissemination of the integrated census microdata from the IECM. This web service shall serve as a one stop shop for all researcher needs related to OS microdata from Europe. It will aid researchers in data exploration, preparation and analysis as well as providing them with a platform for feedback, communal support and possibilities for popularizing their own work. By assisting researchers in a multitude of ways such as an integrated online service centre cannot only inform on availability and applicability of OS data but also save researchers time and resources by assisting them in data preparation and analysis. By pooling knowledge and resources the ESC-OS online platform will provide true added value to researchers throughout Europe and serve as a central component to increasing the scientific application of OS microdata from Europe. Another major benefit of the ESC-OS services to researchers will be that it will increase the standardization of data analysis and will lead to more comparable research results and enable cumulative research. This section will describe the general structure of this online platform and detail the individual services and how they will be linked to each other to enhance functionality.

Figure 1: Structure of the ESC-OS web platform



The ESC-OS online platform shall be comprised of four elements which will be interlinked to complement each other (see Figure 1 for an overview). Firstly a metadata database will offer extensive documentation on OS microdata. This shall cover both an extensive description of studies as well as detailed documentation of variables. Secondly an online library will collect literature and documents pertaining to European OS microdata. This library will contain data documentation, quality reports as well as research and methodological papers written on the basis of OS microdata. The third element of the online platform will be services which offer routines for data preparation and analysis. This shall entail both a dynamic routine generator which can create routines for custom datasets as well as a database of routines with which researchers can operationalize common scientific concepts and measures. In addition to the services provided an important aspect of the online platform will be to create an active user community which not only utilizes the offered services but also provides input on how the online platform can be continually improved and contributes both to the routines and library by submitting user written content. The utility of the online platform will be enhanced by linking these services closely with each other. When searching the metadata database users will find links to both routines and literature or original documentation which refers to a given study. Likewise users can contextually search the library and when they find an interesting article on their topic of study can directly research information on the dataset used from the metadata database.

3.1.1 Metadata database

Providing comprehensive data documentation for OS microdata from Europe will be the core task of the envisioned ESC-OS online platform. Establishing a database where researchers can find structured metadata on all OS microdata from Europe in English language will be instrumental in the proliferation of OS data for scientific use especially across borders. One of the greatest challenges associated with establishing such a database will be the task of translating existing metadata. Offering a comprehensive and structured metadata database is of utmost importance as it provides researchers with information which allows them to assess whether their research questions can be answered with a given dataset. This aspect becomes even more vital when the access to data entails considerable monetary costs.

To further assist researchers in data exploration the ESC-OS online platform will provide extensive information on access conditions. This service can build on the work conducted in DwB WP3. In fact since there will often exist different versions of a single study with different degrees of anonymisation (e.g. a scientific use file and a public use file), the access conditions for each such dataset will be documented separately. Additionally links to order forms and contact information to the institution responsible for the distribution of a dataset will be provided. Thus researchers

interested in using certain data will immediately be informed on their availability and cost, allowing them to choose the most convenient route to access.

The metadata from the ESC-OS online platform will cover both national and integrated European OS microdata. An important requirement of the underlying metadata scheme will be that it can adequately describe a wide range of different kinds of data such as survey data, register data or business data. Additionally the metadata scheme will have to be able to describe collections of studies and a variety of versions of individual studies as the access conditions will need to be laid out for each type. Thus data documentation within the ESC-OS will follow a hierarchical structure which differentiates the following levels:

1. Study Programme: Study programmes are collections of studies which are carried out in regular intervals. Examples include the European Labour Force Survey or the German Micro Census. This category contains general information and topical classification which apply to all studies within the study programme.
2. Study: Refers to a single instance of a study programme, usually an individual year. At this level information about coverage, sampling design and methodology is relayed. Also additional documents such as questionnaires, quality reports or codebooks will be linked to from here.
3. Dataset: This category is a further differentiation of the study level and refers to versions i.e. different datasets which are produced on the basis of a study. The most common distinction here will be between different types of confidentiality (e.g. public use files and scientific use files) and terms of access for different version will be detailed here.
4. Variables: It covers the variables assessed within a given dataset and will relay the exact item wording in all languages, information on filtering, on continuity of item wording over time, as well as basic descriptive statistics and frequency counts. Furthermore this information will be linked to the appropriate sections in PDF versions of questionnaires and codebooks which can be found in the 'online library'.

Documentation for national statistics will initially be limited to the first three levels while documentation for integrated OS data will also cover variables. A further feature of the database shall be to link documentation of related study programmes such as the integrated LFS provided by Eurostat and its national equivalent. It will be of great importance to detail how different datasets differ in regards to content and anonymisation. Individual study programs and studies will also be assigned keywords to make the database machine readable. This will allow users to search for studies covering a specific topical area in a specific country and receive in depth information not only on the contents of a study but also on the terms of access. Beyond the internal search functions of the online platform an important function for its visibility will be that the metadata can be read by search engines thus allowing users who are conducting a web search for a specific study to be linked to the ESC-OS online platform. The metadata in this database will also be fed into the CESSDA catalogue enhancing its utility and increasing the pathways through which potential users can become aware of the ESC-OS online platform and its services.

For national statistics the aim of the ESC-OS will be to provide an overview of all available data with basic descriptions of all datasets from official statistics in Europe. This database shall inform on topics, methodology and access conditions of studies and will provide links to survey instruments, codebooks and documentation provided by NSIs and Data Archives. Such a comprehensive database is currently not available for Europe as a whole and is also missing in numerous member states. It will help researchers interested in cross border and comparative research to find data suitable to their research needs and to inform them on the conditions of access. While documentation for integrated European OS data will follow the same general structure, the aim is to offer more detailed

information and extend documentation down to the level of variables. Since data collection for integrated European statistics is the responsibility of the respective NSIs and is handled differently throughout Europe the ESC-OS will provide information on sample design, data collection and methodology in each country, document original item wordings and link these to the original survey materials.

3.1.2 Online library

The second pillar of the ESC-OS online platform will be an online library. It should contain not only original dataset documentation, quality reports and other adjoining information on studies but also a bibliography on topical and methodological research conducted on the basis of official statistics from Europe. Where possible the bibliography will provide access or link to PDF documents or abstracts. This library serves a dual function. For one the information contained within will supplement the metadata provided on European OS microdata by linking related publications to the data documentation. Alternatively users can browse the online library by topics and find studies from their field of interest and at the same time find out which data sources are being used in this field of research. Via the interlinkages of the library and the metadata users can immediately access information on this data and its availability.

The most important type of documents hosted in the library will be questionnaires, codebooks and other adjoining materials published by the data producer since these directly complement the information from the metadata database. These documents will be linked to in the study description as they can provide interested researchers with more detailed information on topics such as sample design, implementation or weighting procedures. Ideally the original questionnaires and quality reports shall be linked to the variable level. These links will be placed in such a way that the researcher immediately opens the adjoining document at the page which covers the variable in question. Thus researchers will have all information concerning a variable easily available in one place and can assess comparability between countries or over time.

Similarly technical reports produced by Eurostat, NSIs or other institutions tasked with data preparation of OS microdata such as CESSDA members can give interested researches in depth information on a specific study or study programme. These types of reports cover issues such as sampling, variance estimation or harmonization of educational measures or household typologies. While these publications will not be of great interest to researchers during data exploration, they can be very valuable during data preparation and analysis. Specifically they will also be of greatest interest to 'advanced users' of a dataset. Being able to offer these researchers a good reason to frequent the site could also be beneficial to their participation within the online community.

Additionally the library should establish a bibliography which documents substantive and methodological research conducted on the basis of European OS microdata. These books and articles should then again be linked to the corresponding study. The idea behind this is that reference to earlier work conducted on the basis of a specific study can provide a researcher with important information on the possibilities and limitations of a dataset. Especially research conducted on similar topics can be very conducive to evaluating a specific research design.

Finally it is very important to emphasize that the online library should at all times be open to user input ideally by providing an easy to use online submission form from which they can upload information on publications which relate to European official statistics.

3.1.3 Routines

Routines are portions of code that perform specific tasks within a larger program. All common statistical packages (i.e. SPSS, SAS, Stata and R) have their own set of code for reading, processing and analysing data. By providing routines for data extraction and analysis, the ESC-OS seeks to assist

researchers in the use of OS microdata. Two different types of services are envisioned here: A routine generator which will assist users in data preparation for integrated European statistics as well as a routines library which will offer a multitude of routines for data analysis and which will be open for user created input.

The data extraction routine generator for use with integrated European OS microdata will be an invaluable service for researchers since these datasets are rather complex and require a large initial time investment before one can begin with data analysis. Such a routine generator consists of software that creates routines automatically based on the demands of users regarding a specific dataset (e.g. select countries, years, and variables, impose filters, customize sample sizes, etc.). The advantages of developing a routine generator for European official statistics microdata are manifold. A brief description on how the EU-LFS and EU-SILC are currently distributed is necessary to illustrate the benefits of implementing a more dynamic tool for exploring and selecting the data. Microdata for both datasets are distributed by Eurostat in .csv format. These files consist of a number of records. Each record consists of fields or variables. Once the data file is opened, researchers need to consult the codebook in order to understand the meaning of variables and their labels. The codebook is provided in PDF format and it is left to the researcher to create routines which will label variables and values. A routine generator can handle this tedious task of labelling data and would also help researchers deal with the inherently complex structure of these integrated datasets.

The EU-SILC for example consists of four files for every year, two household level files and two person level files. These files need to be combined by researchers for every year they wish to analyse. For instance, if a researcher wants to compare Italy and Spain at two points in time using SILC data, he or she will have to select from the four .csv files for a given year the variables of interest, then merge these files, from this dataset select Italy and Spain and repeat this process for any other years s/he is interested in before merging the yearly files together. For complex research projects the labelling of variables and values, the selection of variables and merging of years can take several days whereas a routine generator could handle this in a few minutes. As seen by this example, the static way in which Eurostat grants access to Official Statistics imposes a time consuming burden on researchers that in some instances may even discourage the use of the data.

In addition to the routine generator the ESC-OS will seek to host a library of routines to help with data analysis. This will include routines which operationalize common social scientific concepts, scales or indicators in integrated European OS data or routines with which national data can be harmonized for comparative research. In order to be able to keep up to date with the most recent trends in research this library will be open for user generated input.

The routines provided by the routine generator and within the routine library will complement the extensive metadata and documentation provided by the ESC-OS. Together these services will be a valuable tool to explore the contents of the data and to access metadata in a dynamic way. This will guarantee that users obtain a better sense of the contents and quality of data. A further benefit of the routine generator and library is that they will attract both novice and experienced researchers to the ESC-OS website and therefore will make users aware of the other services provided by the ESC-OS.

3.1.4 Online community

In addition to providing extensive documentation on OS microdata the ESC-OS online platform should also provide a platform for user interaction and participation. While the primary objective is to establish a mutual support network among users, a lively community will also be able to provide feedback on the site and generate content which will undoubtedly help to improve the online platform. However, establishing an active user community will be a considerable challenge and require that much effort is put into the online community.

In order to stimulate user involvement two different types of elements need to be considered. Firstly users need to be able to interact with each other. The classic solutions would be a message board or a mailing list. While a mailing list has the advantage of always reaching subscribers' mailboxes it is better suited to the purpose of users asking questions than actual discussions. Since discussions via mailing lists will mean that users will receive mails upon mails, and this could likely lead many users to quickly unsubscribe from such a list. While a mailing list does not offer the same possibilities of archiving information as would a message board it would be important to make logs from the mailing list available via the online platform. An example where such a mailing list serves as an important FAQ for researchers that can be found via search engines is the stalist (<http://www.stata.com/stalist/>).

A message board on the other hand is well suited to maintain lengthy discussions on a certain topic. Also users can easily search or navigate through a message board reviewing older posts when seeking advice on certain topics. The major disadvantage of a message board is that it requires people to actively follow it. This might be unproblematic if there are constant and lively discussions going on which might catch people's attention, but especially in the early stages of the online platform this could pose a real problem. This issue would be alleviated by implementing a RSS-feed. Additionally the ESC-OS staff could work to answer questions posed on the message board and forward them to experts knowledgeable on a given subject. This would require that the staff of the ESC-OS invest considerable time in the administration of the message board. Both a mailing list and message board would benefit from the input of experts and advanced users thus steps should be taken to motivate experts working at NSIs, Eurostat and the various CESSDA partners to contribute to these services.

Furthermore possibilities for user participation should also be implemented. One basic form of participation will be to implement easy ways for users to provide feedback on the site and point out problems. However, going beyond mere feedback there should also exist a space where users can make own contributions to the online platform. One such possibility will be contributing routines or by providing links to relevant (own) publications. Other possibilities for user created input would be to offer a space such as a blog or a Wiki where users can contribute their thoughts on a specific topical issue in a more organized form. Such contributions need to be recognizable and could be incentivized by integrating them into technical reports or similar publications after a review by the ESC-OS staff. Such tools should only be implemented once an active user community has been established.

Additionally the ESC-OS could also host a page on social networks such as Facebook. While these would not necessarily offer the ideal functionality for feedback and exchange, they appeal to younger users, increase the ESC-OS' visibility and can be used as a platform to inform about the ESC-OS' activities and can be useful for networking among the ESC-OS users.

A major challenge in establishing a working online community will be to stimulate user involvement and active networking among the user community. To strengthen such ties the ESC-OS should seek to bring users together face to face and strongly interlink its on- and offline activities for example by hearing the community when planning for events and feeding media from conferences and workshops back into the online community.

3.1.5 Dissemination of integrated European Census Data

A further feature of the ESC-OS online platform will be a data dissemination system for harmonized EU census microdata. This system will build on the Integrated European Census Microdata (IECM) and the International Integrated Public Use Microdata Series (IPUMS) projects. These data are stored at the Minnesota Population Center and will be accessible via the ESC-OS. The IPUMS is the world's largest collection of publicly accessible population microdata. The data can be downloaded at no cost by qualified researchers. IPUMS/IECM is designed to facilitate cross-national research; it seeks to put

microdata into the hands of researchers where they can make the best use of it: on their own desktop using the analytical tools with which they are most familiar. The IECM database currently contains information on roughly 60 million persons from 49 censuses taken in 15 European countries since 1960. The data are harmonized so that common variables are consistently labelled and coded across all samples in the data series. Harmonized documentation highlights comparability issues for each variable across time and between countries. All identifying information has been removed from the harmonized samples, and additional confidentiality measures are taken to ensure the privacy of respondents.

The IPUMS/IECM data are available to researchers through a web dissemination system which will be integrated in the ESC-OS. Users must apply for access, describing a scientific need and agreeing not to redistribute the data or use it for commercial purposes. Once approved, they have access to the entire database. The dissemination system allows researchers to browse the variables while applying filters to display only their selected samples of interest. As they browse the system, users select variables to add to their data cart. When they are satisfied with their selections, they go to the data cart and proceed to make a data extract. This data extract can then be further customized by drawing a subsample, adding information from other household members to person level records or selecting only respondents meeting certain criteria. Data is available as ASCII, SPSS, Stata or SAS files.

The IPUMS/IECM variables can be browsed through a drop-down menu that organizes them into broad subject categories for household and individual characteristics, such as migration, education, or utilities. Variables can also be displayed alphabetically or can be searched for via terms appearing in their variable labels, value labels or description text. Each variable has a tabbed display for accessing the various types of documentation associated with it. Also, each variable has a description and general comparability text that highlights major issues that can affect research using multiple samples across countries. Where appropriate, separate comparability discussions describe change over time in the variable within each country, the data categories, etc. The variable codes page displays a column for every sample that contains the variable, indicating which categories exist in each sample, with an option to display the unweighted case count for each category.

Possibly the most valuable part of the variable descriptions is the feature to deliver questionnaire and instruction text. The way the original census question was asked is the most fundamental information about it. To make this accessible to researchers, IPUMS/IECM inserts tags into the census questionnaires and instructions that associate specific blocks of text with each variable in the microdata. When a user selects the tab to display a variable's questionnaire text, the dissemination system compiles on one screen all of the text relating to every sample that is being filtered. Thus with one click a researcher can see, for example, how the marital status question was asked for all censuses taken over the last decades.

Although data harmonization is the hallmark of IPUMS/IECM, the dissemination system also provides access to the underlying source variables that are specific to each individual sample. This feature allows researchers to deconstruct the harmonized IPUMS/IECM variables or to create their own versions.

The IECM data will also have to be linked with the documentation on non-integrated census files so researchers have the possibilities to inform themselves about the original data and how to access confidential versions of census datasets. Integrating the IECM data dissemination tool into the ESC-OS online platform will provide another valuable resource for researchers and will help popularizing the online platform by drawing the widespread user base of the IECM to the ESC-OS website. For the time being data dissemination is only planned for the existing IECM census data; however, it would be possible to integrate other microdata into this system in the future.

3.2 Promote use of OS data

While the online platforms' main purpose is providing information on European OS microdata the ESC-OS should not limit itself to merely providing data documentation but also seek to actively promote the use of official statistics in scientific research and subsequently increase the visibility of European OS microdata. In this section we will outline a number of measures we consider fruitful for widening the research communities' awareness for the utility of European OS microdata both in the short and long term.

3.2.1 Training courses

Encouraging the scientific use of OS microdata requires that researchers are well informed about available data sources, their conditions of access and how they can work with the data. While the ESC-OS online platform can inform on available data and access conditions and ease the application of datasets by providing routines it cannot teach researchers how to use the data. Thus the ESC-OS should seek to implement a periodical schedule of training courses which is tailored to the needs of the research community.

Currently there is a strong demand for training on the use of integrated European datasets such as the EU-LFS or EU-SILC. These datasets are inherently complex and pose a considerable challenge for researchers without extensive experience in quantitative data analysis. Training courses on these programmes cover issues of data access and accreditation and provide an in depth look at the data with hands on data analysis exercises. Other courses could also focus on integrated census files, study programmes at the national level, methods for integrating national datasets or specific topical or methodological issues. Ideally the ESC-OS should identify areas where there exists demand within the research community and organize such training courses. Through the ESC-OS website and the feedback from the online community it would be easy to discern what issues are currently of interest to researchers and devise programmes accordingly. As an additional service the ESC-OS online platform could also make materials and videos from past training courses available as e-learning tools thereby interlinking on- and offline services further.

Providing training on the use of data would not only have the immediate effect that those attending training courses will be more likely to produce high quality research on the basis of OS microdata, this in turn would increase the visibility of official statistics in scientific publications increasing their appeal as a basis for research.

3.2.2 User Conferences

Bringing together those conducting research on the basis of OS microdata fosters knowledge exchange and networking among the research community. Thus the ESC-OS in cooperation with Eurostat will seek to continue to organize regular user conferences for researchers working with OS microdata from Europe. While today these conferences focus primarily on EU-SILC and EU-LFS they should in the future be expanded to other integrated European OS data sources.

These conferences should bring together researchers from throughout Europe as well as data producers. The primary aim of the User Conferences is to offer a platform for discussion of research findings and research needs. Research papers presented at the Users Conferences can stimulate discussion and exchange of research findings across Europe. They will also serve to highlight methodological challenges and present solutions developed by researchers which might benefit others in the future.

These conferences are not only useful to stimulate exchange within the European research community but can also promote dialogue between researchers and data producers. This dialogue gives researchers the chance to point to new needs for data and metadata and highlight existing shortcomings. Such feedback will enable data producers and the ESC-OS to improve the data and its

documentation. Additionally the scientific discourse can also provide data producers with valuable information on new directions and methods of social research as well as which questions are currently central to scientific enquiry in a specific country or region. The benefits of such an exchange are twofold, as researchers can provide external input and a form of quality control for data producers. In turn having a way to give such feedback will also benefit researchers by increasing the utility of OS data for scientific enquiry.

3.2.3 Incentivizing research

As has been argued earlier the best way to increase the scientific use of European OS microdata is to increase the visibility of such data in scientific publications. Thus the ESC-OS should also consider taking steps toward incentivizing research with European OS microdata.

One such possibility would be to implement an annual award for research conducted on the basis of European OS microdata which could be presented at the envisioned user conferences. Such an award should be offered for both scholarly and methodological contributions. Also a young researcher's award for master theses or dissertations could be implemented.

Since one major barrier to using official OS microdata for research purposes are the financial costs associated with data access particularly for access to highly detailed microdata through the research data centres, the ESC-OS could consider establishing a grant in conjunction with data providers which funds data access for selected research projects. A continuation of the Transnational Access activities under DwB (WP9) would be especially interesting for student researchers.

The ESC-OS should also work towards publishing a series of working papers where researchers and experts working with Eurostat, NSIs or CESSDA can publish substantial or methodological work on official statistics data. This would also provide an opportunity to publish user contributions to the ESC-OS online platform and make them more recognizable in the research community. Perhaps this working paper series could even develop into a peer reviewed online journal.

3.3 Coordinate and improve data access

As the ESC-OS shall take on an important mediating role between data producers and researchers it is also in a premium position to work towards improving access to sensitive microdata throughout Europe. Especially since cross border access to national OS data is a complicated issue in most countries of the EU the ESC-OS will seek to not only inform about conditions of access but also remove barriers to access. First steps in this direction are currently being undertaken within the Data without Boundaries project and the ESC-OS should work to build on these efforts. Within the Scope of DwB partners from the European Statistical System (ESS) and CESSDA are currently planning a Remote Access Network (WP4, WP10) which will allow researchers to establish a secure connection from a national safe centre or their home institution via which they can access confidential microdata from participating partner countries. Additionally a common framework for researcher accreditation is being developed (WP3). Thus an infrastructure for accessing national OS data across borders within Europe will already have been established and the ESC-OS can build on this and support the European Statistical System in expanding this cooperation throughout Europe.

3.3.1 European Data Access Forum

The question of access to official statistics microdata is not only a question of practical implementation but also of politics. Not all countries in Europe are willing to allow access to sensitive microdata outside their national borders. However, first steps are being taken to harmonize regulations within the EU. In order to advance this process and come to common standards and wide spread cooperation a Europe wide dialogue is necessary. The European Data Access Forum will stimulate this dialogue and foster the relations between social science data archives and data

producers. This dialogue shall focus not only on data access but also on the harmonization of data production and must be able to adjust to new boundary conditions such as the ever expanding borders of the EU or the increasing importance of register data in official statistics.

This objective shall be met by continuing to organize regular conferences at which NSIs, other providers of official and administrative data and social science data archives come together with representatives from Eurostat and selected senior researchers. Beyond aiming at expanding existing cooperation and cross border access to all corners of Europe an important goal needs to be the establishment of common standards for researcher accreditation, data documentation and anonymisation.

The European Data Access Forum should be hosted in regular intervals in order to monitor progress and to integrate new member countries of the ever expanding European Union. Additionally regional workshops could also be organized to foster relations between Data Archives, researchers and NSIs in regions where access to official microdata is more problematic. Experiences from the first European data Access Forum held in Luxembourg on March 27-28, 2012 are very encouraging.

3.3.2 Supporting cross border data access

Most providers of confidential microdata throughout Europe already provide possibilities for guest researchers to use workspaces on their premises where sensitive microdata can be accessed. Access to such data is generally granted on secluded workspaces which contain standard office tools and statistical packages, but lack access to the internet or any external drives thereby preventing data theft. Users are only allowed to take home outputs which are checked by the staff beforehand.

Such a service is available both to researchers from the respective country and abroad. But travel and accommodation can quickly become costly and is oftentimes not possible within the financial limitations of a research endeavour. Yet especially in cases where data are only available for onsite access or data is highly complex and assistance from local staff is required such safe data centres are often the only possibility to access specific data. From the perspective of the ESC-OS the implementation of such safe data centres in all countries would be highly welcome as without them data would not be accessible to researchers at all and these safe data centres are also an elementary requirement for establishing a remote access network.

In order to enable easy cross border access to national OS microdata it will be necessary to establish a remote access network between safe data centres throughout Europe. Especially for scientists conducting comparative research having to travel to every country for data access will often not be feasible. Thus investments in such an infrastructure will save individual research institutions considerable amounts of money and enable projects that otherwise might not be feasible.

In order to establish such a network safe centres will have to be connected via a remote computing infrastructure. This can be achieved via a remote desktop protocol which allows the users from one safe data centre to control a computer in another safe data centre. Data security will be guaranteed by the implementation of rigid security standards.

This will enable researchers visiting an accredited safe centre in their country to access data stored at other accredited centres throughout Europe. Since granting access to highly sensitive data will always be the responsibility of respective national agencies the maintenance of such a remote access network will fall to the European Statistical System. The core task of the ESC-OS within such a network would be to support the researchers in the accreditation process and to expand the network of safe data centres throughout Europe.

4. Implementation

This section outlines how the ESC-OS could be implemented within the existing institutional framework. Section 4.1 will detail this framework and the position and responsibilities of the different stakeholders in the realm of European official statistics. Following this section, possible forms of implementing the ESC-OS will be discussed. This discussion will cover both the degree of institutional integration into the planned CESSDA-ERIC and the actual functions an ESC-OS should fulfil. These aspects are of course interrelated and some tasks will only be feasibly implemented by a fully institutionalized ESC-OS. We will present here two forms of what a service centre for official statistics could look like. The first form represents the sparsest version both in regards to its institutionalization as well as its functionality. The second form in turn represents what we believe would be an ideal implementation. Based on these we will also propose an implementation which would be feasible today and ensure maximum efficiency of the ESC-OS as well as enabling it to develop further in the future.

4.1 Institutional framework

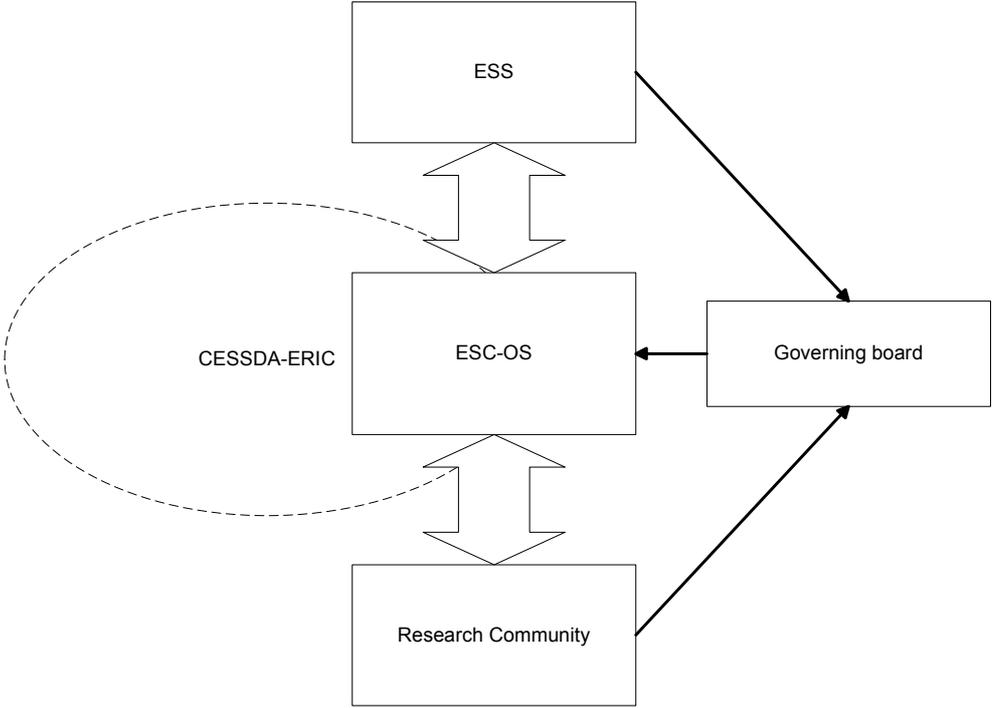
An important requirement for a smooth implementation of the ESC-OS will be that it cooperates with the relevant stakeholders in the realm of official statistics and that this cooperation is clearly defined and beneficial to all sides.

The most important stakeholders in the European context are on the one hand the data producers, i.e. Eurostat, NSIs and other national governmental institutions who produce official statistics. These institutions cooperate under the banner of the European Statistical System (ESS). On the other hand there are the data archives which are united under the helm of CESSDA. Some archives have agreements with their respective NSI and NSAs to provide access to official microdata, some only provide metadata, and others have no role at all in official statistics. CESSDA is currently working towards founding an independent institution, the CESSDA-ERIC. We envision that the ESC-OS should become a subunit of the CESSDA-ERIC responsible for OS microdata. Additionally the research community must be considered a major stakeholder as the services offered by the ESC-OS, and CESSDA as a whole for that matter, primarily aim at supplying them with high quality data and metadata and as such they should be able to articulate their demands regarding data and documentation both to data producers and distributors.

Figure 2 outlines this institutional arrangement. The ESS has two main functions in this regard: data production and providing data access either directly or via other organisations such as data archives. The latter involves not only dissemination of anonymised data but also providing researchers access to highly confidential microdata through safe centres and remote access facilities. The ESC-OS in turn will be primarily responsible for providing user friendly data documentation, provide complementary services for researchers to incentivize the scientific use of European OS data and establish an on-going dialogue between all involved stakeholders.

As the activities of the ESC-OS are primarily aimed at mediating between these two communities it is important that these stakeholders are informed about its activities and have a say in the governance of the ESC-OS. Thus we propose that a governing board be implemented where representatives from the European Statistical System and the research community come together annually and discuss the activities and progress of the ESC-OS, and give feedback from their respective communities on how to improve and better coordinate its services.

Figure 2: Institutional framework of the ESC-OS



4.2 Online platform for European official statistics

In its most basic form the ESC-OS should consist of an online platform maintained and updated by a cooperation of CESSDA partners and thus would not require any standing staff. This online platform’s core task would be to offer extensive documentation for integrated European and national OS microdata and provide information on data access. It could also host further services for researchers such as a routine generator and routines for data analysis. Additionally it could function as a platform to inform the research community about all activities relating to official statistics in Europe such as conferences and workshops. Some of the proposed functions of the online platform which would require regular maintenance or moderation such as an online library or a discussion forum might be difficult to implement without any standing staff to fulfil such responsibilities.

This implementation of the ESC-OS would require that the online platform be maintained and updated by a tight cooperation of CESSDA partners. CESSDA could build on the work from the DwB project, as it will provide an online platform which will host structured metadata. However, CESSDA partners would have to come to an agreement on how to maintain and update this service. Individual CESSDA partners would be tied in through loose commitments, some partners would be responsible for development and administration, while others would only be responsible for updating the database with documentation on studies from their respective countries. Since the CESSDA network currently does not cover all countries in Europe updating information for countries which are not yet part of CESSDA would prove a major challenge. But the existence of an ESC-OS could provide a further incentive for countries not yet involved in CESSDA to join the Network.

4.3 Service unit for official statistics within the CESSDA-ERIC

As has been argued above the degree of institutionalization and the tasks the ESC-OS can handle are to a certain degree interrelated. Especially those services which require constant maintenance or updating are much more difficult to offer without any standing staff. Thus we envision the ESC-OS as a service unit within the CESSDA-ERIC. The ESC-OS would have its own budget and would require a standing staff, which should consist of experts from around Europe. This aspect is of great importance as the ESC-OS could benefit from knowledge about the situation of official statistics in different regions of Europe as well as a wide range of language skills.

The ESC-OS should be responsible for establishing and maintaining an online platform on official statistics, this should include a metadata database as well as an online library and an active community and user support. The updating of the online library and the metadata database would be handled by the ESC-OS staff with CESSDA partners providing them with metadata from their respective country, this integrated metadata would in turn be fed into the CESSDA catalogue as it is essential that services are interlinked in order to ease data discovery. The staff would also be responsible for constantly updating routines, certifying user submitted routines and updating them to the site. Also updating information on countries which are not part of CESSDA would have to be handled by the ESC-OS staff.

As the ESC-OS should become the core institution responsible for the proliferation of European OS microdata it should also be responsible for organizing conferences of European OS data users and data producers and moderating the Europe wide dialogue on scientific use of official statistics between all involved stakeholders.

Another task which would be instrumental in fostering the use of OS data among researchers would be to offer them support in the form of training courses. The ESC-OS would be in a great position to assess the demand for training through feedback and discussion from the online community.

Likewise the ESC-OS should also play a central role in coordinating efforts to grant cross national access to national microdata within Europe. It would have a direct link to the researchers interested in using such services via the online platform and could inform about these services online. Since the ESC-OS will be in constant exchange both with data archives and data producers it would be best set up to coordinate such programs.

A further effort that the ESC-OS should undertake is to establish an award for research with OS data from Europe. Again the ESC-OS could easily recruit a jury through its connections to the scientific community, data archives and data providers. Furthermore the online platform could also be used as a platform to suggest nominees for this award.

Beyond its responsibilities for establishing and maintaining a dialogue between data producers, data distributors and data users in Europe the ESC-OS should also seek to link up with other research infrastructure institutions around the world in order to be able to provide European researchers with better information on data and from abroad and ways to access such data.

4.4 Recommended initial implementation

Our recommendation for an initial implementation of the ESC-OS would be a service team responsible for managing the online platform and fulfilling a number of smaller coordination and service activities. This would require a fixed staff responsible for updating and maintenance of services, which could also organize regular conferences and training courses. We believe that institutionalizing the ESC-OS as a service unit with its own staff would be far more efficient than having CESSDA partners share these responsibilities as a number of the tasks will require constant maintenance, which is better handled by personnel responsible and specialized for these tasks.

If these tasks were to be handled by a cooperation of CESSDA members this would require constant coordination of responsibilities with over twenty partners from CESSDA and would be highly inefficient as it would consume a considerable amount of time for management.

As far as institutional integration is concerned we would recommend that this unit should be integrated into the CESSDA-ERIC and that individual CESSDA partners would be obliged to provide it with metadata from their country on a yearly basis so that the database can be kept up to date.

While the services offered by this implementation are more limited they do cover all the key tasks to improve access to OS microdata in Europe. Once all these services have been implemented and are running smoothly such a unit would be well suited to assessing areas where additional services are required and which services should be expanded in order to further the proliferation of OS microdata for social research.

Finally, if so desired the European Service Centre for Official Statistics would be ready to discuss with ESS and Eurostat ways in which it can assist in data dissemination.

5. Cooperation

Beyond its tight institutional connection with CESSDA and its respective partners the ESC-OS will have to work in close cooperation with a number of stakeholders, first and foremost the data producers, the national statistical institutes and Eurostat. The cooperation with these partners should seek to build on the existing ties CESSDA maintains to these institutions.

5.1 Eurostat

The cooperation with Eurostat will be of utmost importance to the ESC-OS but can be very beneficial to both parties. While CESSDA as a whole and many individual CESSDA partners have built up ties with Eurostat over the years the implementation of a European Service Centre for Official Statistics will necessitate that this entity also work as the communication channel with Eurostat. While this will make communication with CESSDA easier for Eurostat it will also mean that the ESC-OS and the CESSDA partners more strongly coordinate their cooperation with Eurostat.

The areas where cooperation with Eurostat will be of the essence are documentation and access to integrated European OS microdata. While Eurostat produces a wealth of information on its integrated study programmes in quality and country reports which can be downloaded from its homepage structured metadata is currently not available. Thus the ESC-OS' envisioned online platform with its extensive data documentation will complement Eurostat's existing services. Care has to be taken to ensure a smooth working relationship by ensuring constant communication and tying Eurostat into the ESC-OS public activities such as conferences and workshops. Here the ESC-OS should position itself as a communication channel between Eurostat, and data producers in general, on the one hand and the scientific community on the other through which feedback can flow. Building on this knowledge the ESC-OS can then help to increase utility of data for scientific research.

At the political level Eurostat will be an important partner in the ESC-OS mission to improve data access. In the past Eurostat has been very cooperative; however, it must abide to the legal framework which respects both EU and national legislation.

5.2 NSIs and other national data providers

While in most countries in Europe NSIs and their respective RDCs are responsible for dissemination of OS microdata this is not always the case and in some countries this task is also handled by data archives. Thus the ESC-OS will have to be conscious of the specific situation in each country and build on existing relationships. Most CESSDA partners maintain close ties with their national statistical institutes as they offer access and metadata for some official statistics microdata or sometimes even the data itself. The core objective of the ESC-OS must be to integrate and expand existing services and translate documentation if it is not yet available in English language. In contrast to Eurostat many national data providers – in some instances social science data archives - currently already offer structured metadata via their RDCs. The ESC-OS must take care not to duplicate any existing services and instead link to metadata provided by national data providers within the ESC-OS data documentation. This is one reason why the metadata for national statistics shall not be as expansive in scope as that for integrated OS microdata, as the demand must be properly assessed first.

Cooperation with NSIs will be most productive when it is mutually beneficial to both partners and it will be important to communicate to NSIs that establishing a portal with extensive English language metadata will be beneficial. If NSIs come to perceive the ESC-OS activities as free advertisement of their products and services instead of competition they will likely be much more inclined to cooperate closely.

A good working relationship with NSIs will also help the position of the ESC-OS when lobbying for improved access to sensitive microdata for researchers. Establishing a secure and smoothly functioning remote access network could be a first step towards establishing trust here and could lead reluctant NSIs to join such a network and establish closer cooperation with their respective Data Archive.

5.3 Collaboration with open data initiative projects

Beyond the social sciences there are a great number of efforts to improve data access infrastructures for researchers. Projects such as EUDAT or ENGAGE which much like Data without Boundaries are funded by the European Commission are working towards providing integrated data services for scientific research. While these databases might not be contextually related it would be beneficial to seek an exchange with such projects on topics such as documentation standards, data harvesting and software solutions from which all partners can profit.

Furthermore integrating the services of the ESC-OS into larger and cross disciplinary databases should be sought in order to increase its visibility and ultimately improve researchers' level of information on all types of data available.

5.4 Global Cooperation

As a an institution servicing researchers the ESC-OS should not limit itself only to the European Union as comparative social research is often interested in a global perspective. Thus the ESC-OS should work to expand its network beyond the European Union and seek cooperation with similar infrastructure institutions and data providers throughout the world.

Important producers of official statistics who operate at a global scale are among others UNSTAT, the OECD, the World Bank as well as the ILO. The ESC-OS must seek close cooperation with such institutions by interlinking with their online platforms, inviting them to join the governing board and ultimately offering documentation and services for microdata they provide.

Furthermore the ESC-OS will have to seek close cooperation with data archives from abroad by linking services and databases. Here the ESC-OS should attempt to build on existing ties established

via the International Federation of Data Organizations for the Social Sciences (IFDO) as well as other key players such as the ICPSR. Such cooperation will be greatly beneficial to European and foreign researchers interested in comparative research. Building on such connections the ESC-OS could more easily assist European researchers seeking data from abroad by coordinating contacts to the relevant institutions. Likewise researchers from abroad could be forwarded to the ESC-OS to help them with any needs regarding European OS microdata.

6. Conclusions

In this report we have developed a concept for a service centre for European official statistics. Ideally, this service centre should be implemented as a unit of the planned CESSDA-ERIC. It should be responsible for the following tasks:

- Provide extensive data documentation, inform on availability and offer routines for data preparation and analysis for European official statistics via a one stop online service centre
- Handle the dissemination of publicly accessible OS microdata such as the IECM census data
- Establish a community of OS microdata users and encourage Europe wide networking among these researchers
- Lobby for improved access to OS microdata, harmonization of regulations on data access and anonymisation throughout Europe
- Ensure constant communication between all relevant stakeholders in Europe to ensure that OS microdata can remain up to date
- Incentivize research with European OS microdata by informing on its potential for social enquiry, increasing its scientific application and visibility within the scientific community
- Improve access to sensitive OS microdata across borders by supporting the European Statistical System's efforts to establish and expand a network of safe data centres throughout Europe

The success of such a service centre however is dependent on close cooperation with all relevant stakeholders, its acceptance and recognition within the research community and a rigid maintenance of its services. Thus it is of utmost importance that:

- A constant dialogue between data producers, data distributors and data users is implemented and maintained and that this dialogue incorporates all parts of the constantly growing European Union
- Services provided online are well maintained, and constantly updated for enhanced functionality and actuality
- An active user community is established whose feedback and questions are heard and integrated to further enhance services; this entails that services provided on- and offline are integrated to improve networking among users of OS microdata throughout Europe

We have outlined in this report how such a service centre could be implemented and are confident that our recommended initial implementation would be well suited to fulfilling these requirements and can thus help to ensure that official statistics microdata becomes a corner stone of European social research in the 21st century. While this would require some considerable initial and long term investment we are of the conviction that centralizing the described tasks and bundling the necessary know-how will prove more efficient than doing so disparately throughout Europe. We also believe that with its extensive expertise and far spanning network CESSDA is the ideal home for such an integrated European service unit.