



Project N°: 262608



ACRONYM: Data without Boundaries

DELIVERABLE D5.1 (revised and updated version)

Report on 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 46
PROJECT START DATE:	1st May 2011	DURATION: 48 Months
DATE OF ISSUE OF DELIVERABLE:	April 20th 2015	
DOCUMENT PREPARED BY:	Partner Numbers 5, 1, 25	Partners Names GESIS, CNRS-RQ, CED

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

A European Service Centre for Official Statistics

Christof Wolf (GESIS), Alexander Mack (GESIS), Roxane Silberman (CNRS/RQ),
David Schiller (IAB) and Albert Esteve (CED) *

April 2015

Summary

This report develops a concept for a European Service Centre for Official Statistics (ESCOS), a unit providing services for users of European official statistics microdata. ESCOS' primary objective should be to promote the scientific use of European official statistics microdata by providing a wide range of services such as comprehensive metadata and data access infrastructures and by working towards harmonization of data access for scientific purposes throughout Europe. The underlying idea is that services provided by ESCOS should benefit the research community as well as the European Statistical System (ESS). Furthermore such a service centre could take on the role of mediator and communication channel between the research community and the European Statistical System.

This report proposes that such an institution should be established on the basis of the existing CESSDA network of European data archives and should cooperate tightly with the ESS and specifically Eurostat. It needs to be emphasized that the implementation of an ESCOS is an important step towards ensuring the legacy of the Data without Boundaries project. In order to ensure that the developed services and existing cooperation can be maintained a stepwise implementation is proposed. The first and most essential step would be the establishment of an online platform which offers metadata, routines and a platform for exchange. Building on a cooperatively established online platform it should be attempted to institutionalize an ESCOS as a subunit of CESSDA.

* We gratefully acknowledge feedback and input from Paul Jackson, Managing Director of CESSDA-AS, Bergen.

Table of Contents

1. Introduction
2. Objectives
3. Tasks
 - 3.1 Online Platform
 - 3.2 Promote Use of Official Statistics Microdata
 - 3.3 Cross Border Data Access
 - 3.4 Enhance Cooperation: European Data Access Forum
4. Implementation
 - 4.1 Institutional Framework
 - 4.2 Possible Implementations
 - 4.3 Steps toward establishing an ESCOS
5. Conclusion

1 Introduction

This report is a revision of Deliverable 5.1 of the Data without Boundaries project¹. It outlines the concept for a European Service Centre for Official Statistics (ESCOS) which should provide comprehensive services for users of European official statistics microdata. ESCOS should provide comprehensive services to researchers interested in employing official statistics microdata and work as a communication channel between the European Statistical System (ESS) and the research community.

While this report is an output of DwB Work Package 5 it is actually the product of a much larger discussion within the DwB project, the CESSDA community and the ESS. The first version of this report was produced by members of WP5 and published in April 2012. However prior to publication a draft version had been circulated, commented on and revised within the Data without Boundaries project. A second draft was then circulated within the CESSDA community and discussed in the general assembly and also discussed with members of Eurostat. This revised report has been updated to include valuable lessons learned within the project lifetime and to incorporate valuable feedback from Eurostat, NSIs, the CESSDA community as well as Denise Lievesley, who was commissioned by the EC to review the report.

The Data without Boundaries project has developed a wide range of services to aid researchers working with official statistics microdata. Furthermore trans-European cooperation has been intensified and a European dialogue between data producers and the community of data archives has been stimulated. In order to ensure the persistence of these project outputs and to institutionalize the established cooperation a formalization of services is necessary. This report outlines how a European Service Centre for Official Statistics, which would supply these services, could be established and makes suggestions for its stepwise implementation.

The report is structured as follows: Section 2 outlines the goals and objectives to be reached by a European Service Centre for Official Statistics. Section 3 goes on to outline the specific tasks which should be carried out by a European Service Centre for Official Statistics. Section 4 then reports on how such a service centre could be implemented and outlines stages for a stepwise implementation. Section 5 provides a short summary and conclusion.

2 Objectives and Scope

Authorities invest heavily to ensure data from official statistics 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 these data is not realized today because they are underutilized. An increased usage of microdata from official statistics in scholarly research would further advance our understanding of social and economic processes and this knowledge in turn would help politicians to make informed decisions.

¹ The original report can be accessed under the following URL:

http://www.dwbproject.org/export/sites/default/about/public_deliverables/d5_1_european_service_centre_report.pdf

The overarching goal of the European Service Centre for Official Statistics therefore is to remove barriers to access to official statistics microdata² in Europe and stimulate research in order to make better use of the wealth of high quality data available. To this aim ESCOS should provide researchers with comprehensive services for European official statistics microdata. This should entail extensive data documentation and tools for data preparation as well as training in the use of OS microdata. Moreover ESCOS should seek to facilitate exchange among researchers, data producers and data archives throughout Europe and work towards reducing boundaries to data access. It is important to note that ESCOS will not replace services for official statistics at the national level. On the contrary, ESCOS depends on active cooperation between NSIs and data archives at the national level and will build on these.

The vast majority of research conducted in the social sciences today is based on secondary data analysis. While many research endeavours could benefit from the application of OS microdata these are currently vastly underutilized. The most heavily used data are those from academic surveys.³ While these data are tailored to researchers' needs and provide extensive documentation and ready-to-use datasets, they are in some respects inferior to microdata from official statistics. For one, sample sizes from official statistics tend to be much larger than in academic 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 comparisons of EU member states both cross-sectionally and over time.

Thus the primary agenda of the proposed European Service Centre for Official Statistics should be to inform about and provide services for microdata from official statistics in order to make these data more accessible to the research community. These services should go beyond what data producers are able to offer by providing training, tools and thorough documentation which aids researchers both during data exploration and analysis. It has to be emphasized that the services envisioned here go beyond what existing infrastructures can currently provide. Particularly the provision of structured and user friendly metadata is a labour intensive task which cannot be replaced by mechanical harvesting of existing metadata. The services outlined here can aid in data exploration (through provision of metadata and training courses), in accessing data (coordination of remote access, researcher accreditation) and during data analysis (detailed variable level metadata and routines).

One important lesson from the Data without Boundaries project – in particular Task 8 in which in depth interviews were conducted – is that researchers are interested in integrated services that not only cover all steps of a research project but also combine information about data from various sources and a wide range of countries in one place. This is particularly pertinent for cross-national research when researchers are interested in employing microdata from a number of countries.

Thus the primary task of ESCOS should be to establish an online platform which functions as a one stop shop for researchers interested in the use of OS microdata from Europe. This online platform should offer information on all services offered by CESSDA and ESS partner institutions for OS microdata from Europe and provide comprehensive metadata and tools for data handling and

² We use the term official statistics to refer to all statistical data produced by government agencies. This includes both survey and process produced data.

³ A quick search in google scholar nicely illustrates this. For the time period of 2010 to 2014 the leading international social surveys are the ISSP (13,900 hits), the World Values Survey (11,400) and the European Social Survey (9,200). The EU-SILC is mention 5,920 times while the EU-LFS only comes up 738 times.

analysis. The metadata database should inform interested researchers on available OS microdata sources and how these can be accessed and utilized for social research. Documentation should cover both integrated European Statistics as provided by Eurostat, as well as OS microdata provided by national authorities from EU member countries and information on the Integrated European Census Microdata (IECM).⁴

A major barrier which must be overcome to ease access to OS microdata within Europe is the issue of cross border access to highly confidential national microdata. For one access to highly confidential OS microdata throughout Europe does not follow common rules. As legislation for protecting confidential statistical information is set at the national level and the methods of protection to achieve those legislative requirements are also developed at the national level this results in considerable variations in practices between NSIs in Europe. 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 ESCOS should aim to further develop and institutionalize plans for a European Remote Access Network (EuRAN) in order to facilitate the cross border access to confidential OS microdata. Furthermore ESCOS should also work towards the institutionalization of a common framework for researcher accreditation throughout Europe.

The successful and efficient provision of the services sketched in this report relies on trusting relations between service providers and members of the ESS. ESCOS should therefore strive to foster and sustain the European dialogue between data producers, data archives and researchers. The ESCOS online platform should 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 should be institutionalized by establishing regular conferences, which bring together representatives from NSIs and NSAs, 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.

Furthermore the service centre should function as a broker for interests of the research community and could communicate such wishes to the ESS. However as this communication needs to take place at the European level and the services provided aim to service all of Europe this should be handled by a European infrastructure. At the same time the ESS should be involved in defining services offered by ESCOS. Therefore, it seems necessary to involve the ESS in the organisation of the centre, e.g. by having a representative of an NSI and of Eurostat on an advisory or governance board.

⁴ Just as IECM is linked to IPUMS International – a database holding census microdata from around the world – services from ESCOS may in the long run link to similar services outside of Europe.

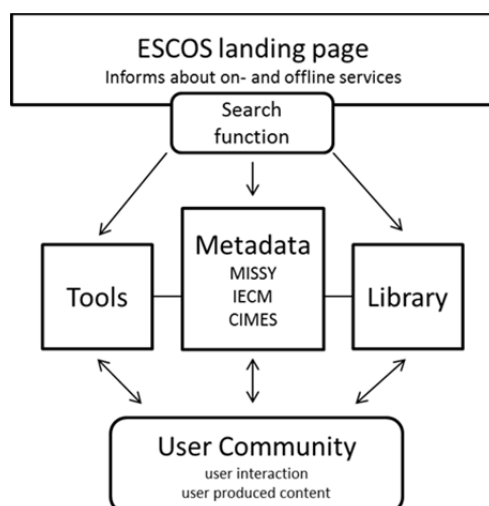
3 Tasks

This section outlines the concrete tasks that ESCOS should handle. This includes first and foremost the establishment of an online portal which should function as a one stop shop for researcher's needs in regards to Official Statistics microdata. Section 3.1 details the elements of such a service. Furthermore ESCOS should aim to promote the use of OS microdata in the scientific realm by organizing training courses and conferences (section 3.2). Additionally, ESCOS could take a leading role in the coordination of cross border data access (section 3.3) and could function as an umbrella institution for cooperation between data providers, archives and the research community (section 3.4). Furthermore, supporting the ESS in data dissemination could be an option for the future, if so desired. However, as currently no concrete plans along these lines exist the topic will not be discussed here.

3.1 Online Platform

The central feature of ESCOS should be an online platform which serves as a Central information point about European Statistics microdata (see Figure 1 for a schematic overview). Such a website should include comprehensive information on all services provided by CESSDA partners and the ESS both on- and offline. The most important aspect of this online platform would be a comprehensive collection of metadata on national and integrated European microdata. Currently these metadata are held in separate systems: MISSY which is hosted by GESIS includes information on integrated European microdata, CIMES which is hosted by CNRS-RQ includes metadata on data collections at the national level and the IECM which is hosted by CED contains information on integrated census microdata. A major advantage of integrating these systems would be that a centralized search function could be implemented for all of them. Furthermore in order to amplify the utility of these metadata and to further aid interested researchers these metadata should be interlinked with two additional databases one which includes tools and routines which aid in data handling and analysis and one which contains a bibliography of official documentation, technical reports and scholarly research.

Figure 1: Structure of ESCOS Online Platform



A user community, most likely in form of a forum and or wiki, can serve as a space for users to get advice from staff and other users and make own contributions to the tools and bibliography. Furthermore users should also be able to download public use census microdata, either directly or via the IECM webpage. In short the idea is to establish an online service which functions a one stop shop for all researcher needs related to OS microdata from Europe. Such an online service would aid researchers in data exploration, preparation and analysis, provide them with a platform for feedback, communal support and possibilities for popularizing their own work. By assisting researchers in a multitude of ways such an integrated online service could not only inform on availability and applicability of OS data but also save researchers valuable time and resources by assisting them in data preparation and analysis. By pooling knowledge and resources the ESCOS online platform could provide added value to researchers throughout Europe and serve as a central component to increasing the scientific application of OS microdata from Europe.

3.1.1 ESCOS Landing Page

The landing page for the ESCOS online platform should fulfil three core functions. On the one hand it should inform about all services and databases made available to users. Furthermore it should provide an integrated search function which can search and browse all elements of the online platform. This search engine should be aided by a common thematic classification and other controlled vocabularies which underlie all databases. Lastly the landing page should serve as an information portal which informs on all current, past and future activities of CESSDA and ESS regarding OS microdata. This would include a schedule of training courses, workshops, possibly also some online classes, as well as information on user conferences and the European Data Access Forum.

3.1.2 Metadata Database

Providing comprehensive data documentation for OS microdata from Europe will be the core task of the envisioned ESCOS 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 microdata for scientific use. 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 or time intensive accreditation procedures.

For national statistics the aim of ESCOS will be to provide an overview of all available microdata with basic descriptions of all datasets from official statistics in Europe. This database shall inform on topics, methodology and access conditions of studies and provide links to survey instruments, codebooks and documentation provided by government agencies and Data Archives. Within the DwB project CNRS has established the Centralising and Integrating Metadata from European Statistics (CIMES) database which includes over 1500 studies from 27 countries. Deliverable 5.2 of the DwB

project⁵ provides additional information on the CIMES system. The database can be accessed at <https://cimes.casd.eu>.

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 ESCOS 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. Two such services are MISSY which is hosted by GESIS and IECM which is hosted by CED. The MISSY system to date includes metadata on 5 different studies provided by Eurostat: the EU-SILC, the EU-LFS, the AES, the CIS and the SES and can be accessed at <http://www.gesis.org/missy/eu/missy-home>. The IECM system disseminates harmonized census microdata from 54 different censuses conducted between 1960 and 2010 in 16 European countries. Additionally the IECM provides users with comprehensive study and variable level metadata. It can be accessed at <http://www.iecm-project.org>. Deliverable 5.4 of the DwB project provides an overview over these databases⁶.

To further assist researchers in data exploration the ESCOS online platform will provide extensive information on access conditions. This service can build on the work conducted in DwB WP3 on accreditation procedures. In fact since there will often exist different versions of a single study with different degrees of anonymization (e.g. a scientific use file and a public use file), the access conditions for each such dataset should be documented separately. Additionally links to order forms and contact information to the institution responsible for the distribution of a dataset should 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. Work on this has already begun and access information has been integrated into the CIMES system.

3.1.3 Online Library

A further database which should be established within the ESCOS online platform is an online library. This database should contain not only the original data documentation, quality reports, technical reports and other adjoining information produced by data providers but also a bibliography on topical and methodological research conducted on the basis of official statistics from Europe. All elements stored within this database should be linked to specific elements within the metadata database. Where possible the bibliography should provide access or link to PDF documents or abstracts. This bibliography would serve a dual function. For one the information contained within supplements 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 learn about the data sources that are commonly employed in this field of research. Due to the interlinkage of the bibliography and the metadata users can immediately access information on this data and its availability.

⁵ http://www.dwbproject.org/export/sites/default/about/public_deliverables/dwb_d5-2_databank-national-survey_report_final2.pdf

⁶ http://www.dwbproject.org/export/sites/default/about/public_deliverables/dwb_d5-4_databank-documenting-integrated-eu-data_report.pdf

Similarly technical reports produced by Eurostat, NSIs or other institutions tasked with data preparation of OS microdata can give interested researchers 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 great interest to advanced users of a dataset.

Additionally the bibliography should include references to 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 conducive to evaluating a specific research design.

Finally it has to be emphasized 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 information on publications can be uploaded.

3.1.4 Microdata Tools

Microdata Tools, i.e. code for statistical analyses, are intended to aid researchers in data preparation and analysis using a statistical package (i.e. SPSS, SAS, Stata or R). Some tools will consist solely of documented code while others will include more elaborate documentation and might take the form of an instruction manual or scholarly paper on methodological issues. Providing routines for data extraction and analysis greatly assists researchers in the use of OS microdata. Two different types of Microdata Tools are envisioned here: setup files which could assist users in data preparation for integrated European statistics as well as a collection of routines and microdata tools which can aid in data analysis and which will be open for user created input. All tools, routines or code will be licenced in such a way that they are accessible free of charge.

Setup files which can be used to import Eurostat microdata into a statistics program of one's choosing provide 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. A brief description on how the EU-LFS and EU-SILC are currently distributed is necessary to illustrate the benefits of such setup-files. Microdata for both datasets are distributed by Eurostat in .csv format. 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 setup file will transform the data into a format usable in a statistical package and will handle the task of labelling variables and values. In addition, Microdata Tools can assist users in creating datasets including data from several countries and/or time points.

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 of interest 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 setup-files could handle this in a few minutes.

ESCOS will also host a library of Microdata Tools to help with data analysis. This will include routines which operationalize common social scientific concepts, scales or indicators in integrated European OS microdata 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 should be open for user generated input.

The setup files and tools complement the metadata and documentation provided by ESCOS. They reduce the initial time investment required of researchers to familiarize themselves with the data, thus reducing hurdles to employing official statistics microdata for scientific enquiry. A further benefit of these services is that they will attract both novice and experienced researchers to the ESCOS website and therefore make users aware of the other services provided by ESCOS.

3.1.5 Online Community

In addition to providing extensive documentation on OS microdata the ESCOS online platform should also provide a forum for user interaction and participation. While the primary objective would be to establish a mutual support network among users, a lively community would also be able to provide feedback on the site and generate content which will undoubtedly help to improve the services provided. 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 available here are message boards or mailing lists. 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.

A message board on the other hand is well suited to maintain lengthy discussions on a specific 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 making it easy to follow the message board by implementing a RSS-feed. Additionally the ESCOS staff could work to answer questions posed on the message board and forward them to experts knowledgeable on a given subject. However this would require that the staff of ESCOS invest considerable time on 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. Especially in the early stages the experience of users that the online platform is actually a place for useful advice and fruitful discussion could aid in its' proliferation.

Further possibilities for user participation should be explored. These should not only include user feedback but also user created content, e.g. routines, technical reports or similar publications. Additionally ESCOS should look to promote its services also via social media thereby increasing ESCOS' visibility.

3.2 Promote use of Official Statistics Microdata

While the online platforms' main purpose would be to provide information on European OS microdata ESCOS 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 ESCOS online platform can inform on available data and access conditions and ease the application of data sets by providing routines it cannot teach researchers how to use the data. Thus ESCOS 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 should 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 focus on integrated census files, business data or administrative data which are increasingly accessible for research at the national level, methods for integrating national datasets or specific topical or methodological issues. Ideally ESCOS should identify areas where there exists demand within the research community and organize such training courses. Through the ESCOS 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 ESCOS 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. Furthermore such training courses can also greatly benefit another important target audience of official statistics microdata: the scientific staff in executive agencies such as the European Commission or national administrations.

3.2.2 User Conferences

Bringing together those conducting research on the basis of OS microdata is invaluable for knowledge exchange and networking among the research community. Thus ESCOS in cooperation with Eurostat should seek to continue to organize regular user conferences for researchers working with OS microdata from Europe.

These conferences should bring together researchers from throughout Europe as well as data producers. The primary aim of the user conferences should be to offer a platform for discussion of research findings and research needs. Research papers presented at the user 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.

However 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 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.

3.3 Cross Border Data Access

As cross border access to national OS microdata is a complicated issue in most countries of the EU ESCOS should seek to inform about conditions of access to national microdata and also attempt to lobby for the removal of barriers to access. First steps in this direction have been undertaken within the Data without Boundaries project and ESCOS should work to build on these efforts. Within the scope of DwB (WP 4, WP 10) partners from ESS and CESSDA are currently planning a European Remote Access Network (EuRAN) 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 institutions throughout Europe. Additionally a common framework for researcher accreditation is being developed (WP3). Thus a pilot for accessing national OS microdata across borders within Europe has been implemented successfully and ESCOS could support the European Statistical System in expanding this cooperation throughout Europe. Additionally ESCOS could support the ESS in managing researcher accreditation if so desired.

3.3.1 Researcher Accreditation

As described in section 3.1.1 an important part of the metadata on national statistics should be detailed information on access conditions and accreditation procedures. This should include not only information on the institutions and entities eligible for access but also details on the procedures for

researcher accreditation which are taken from the accreditation database⁷ prepared as part of WP3 of DwB. ESCOS could build on this work and continually update and expand the information.

As detailed above the most elementary of services would be that all metadata provided by ESCOS includes detailed information on current procedures and links to the most up to date application forms and access conditions which now can already be found in CIMES. Furthermore a properly staffed ESCOS could also provide direct support to researchers seeking access particularly for confidential microdata accreditation relying on the national data archives. A further step would be to work towards and lobby for the harmonisation of access conditions and application forms. ESCOS could function as an umbrella for such discussions⁸. A further possibility in which ESCOS could serve as a coordinator would be by establishing a database of successful researcher accreditations. Previous researchers' accreditation would be registered as a support for speeding up transnational accreditation for projects requiring multiple accreditations. Such a mechanism could function as a driver for possible bilateral or multilateral agreements between national authorities in charge of accreditation towards more harmonized and rapid procedures for transnational access.

As highlighted above having a central reference point and ideally some form of centralized coordination would be a great help to researchers pursuing cross border data access particularly when they are interested in comparative research.

3.3.2 European Remote Access Network

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, so called safe rooms, which contain standard office tools and statistical packages, but lack access to the internet or any external drives or other devices thereby preventing data theft. Users are only allowed to take home outputs which are checked by the staff beforehand. Such services are generally available to researchers from the respective country and abroad. However 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 such safe data centres are often the only possibility to access highly confidential microdata. Over the last few years an increasing number of services offering remote access solutions have been implemented by providers of official statistics microdata. In many cases, access points have to be in the country meaning that researchers from other countries would have to travel, but some are now also offering remote access with access points across borders (e.g. CASD, CBS and IAB). However these systems do not allow researchers to access data located in different research data centres from a single access point, i.e. data located in different Research Data Centres. In order to enable easy cross border access to national OS microdata interlinking such safe centres via a remote computing infrastructure is the most feasible solution and such an approach has been explored within WP4 of Data without Boundaries.

Technically such a system should allow users to connect from an access device via a secure internet connection to a single point of access where users have to log in with their credentials. From here

⁷ http://www.dwbproject.org/service/accreditation_db.html

⁸ The deliverable D3.1 of the DwB project outlines that harmonization of the application form would be possible and many NSIs would be willing to work on this.

they will be forwarded to a virtual research environment (VRE). Via a menu users would be able to access services provided via a Service Hub. These services are coordinated by a rights management system that allows users to only access services they are accredited for. ESCOS could be responsible for handling the rights management system and the VRE could also be customized to integrate other services provided by ESCOS such as metadata. Data providers therefore do not have to build new infrastructures. They can bring in their existing solutions and benefit from the additional tools provided by ESCOS. For users such a virtual research environment works as a graphical user interface allowing for using online services such as data documentation, online applications, messaging services, interfaces to research data, archiving tools, editors, and so on.

3.4 Enhance Cooperation: European Data Access Forum

The question of access to official statistics microdata is not only a question of practical implementation but also of laws, regulations and culture. 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 documentation and services 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 research institutions. 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 European Union. Additionally regional workshops could also be organized to foster relations between data archives, NSIs and researchers in regions where access to official microdata is more restricted. An important tool which could help in establishing working relationships between data archives and NSIs would be to highlight existing best practice models and calculating cost/benefit models which would provide countries with an idea of whether such cooperation could prove beneficial to them.

4 Implementation

This section outlines how a European Service Centre for Official Statistics 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 ESCOS will be discussed. We will present here possible steps towards an implementation of an ESCOS. A starting point would be the establishment of an online platform which is decentrally managed by partners from throughout Europe who provide content and links to their online services. At the other end of this development we see a subunit of

CESSDA with own staff which runs an integrated online platform as described in section 3.1 and coordinates activities of ESS and CESSDA partners in regards to OS microdata.

4.1 Institutional Framework

An important requirement for a smooth implementation of ESCOS 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. We envision that in the long run ESCOS could become a subunit of CESSDA-ERIC. Additionally the research community must be considered a major stakeholder as the services offered by ESCOS, 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 service providers.

Figure 2: Proposed Institutional Framework for ESCOS

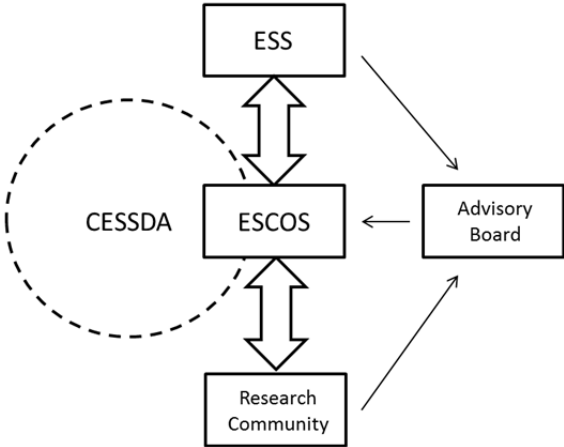


Figure 2 outlines this institutional arrangement. The ESS has two main functions in this regard: data production and provision of data access. The latter involves not only dissemination of anonymized data but also providing researchers access to highly confidential microdata through safe centres and remote access facilities. ESCOS in turn shall be primarily responsible for providing user friendly data documentation, provide complementary services for researchers to incentivize the scientific use of European OS microdata and establish an ongoing dialogue between all involved stakeholders.

As the activities of ESCOS are primarily aimed at mediating between ESS and these two communities it is important that these stakeholders are informed about its' activities and have a say in the

governance of ESCOS. Thus we propose that an advisory board be implemented where representatives from the European Statistical System and the research community come together annually to discuss the activities and progress of ESCOS and give feedback from their respective communities on how to improve and better coordinate its' services.

4.2 Possible Implementations

In its most basic form ESCOS should consist of an online platform maintained and updated by a cooperation of CESSDA partners and thus would not necessarily require an own standing staff as long as partners from different countries share in the responsibilities of updating content and administrating the online service. This online platforms' core task would be to offer extensive documentation for integrated European and national OS microdata and provide information on data access as described in section 3.1. It could also host further services for researchers such as microdata tools for data exploration and 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 training. Some of the proposed functions of the online platform which would require regular maintenance or moderation such as an online bibliography or a discussion forum might be difficult to implement without any standing staff to fulfil such responsibilities. Furthermore as long as the various databases included are not fully integrated a search and browsing functions will be hard to implement.

This implementation of ESCOS would require that the online platform be maintained and updated by a tight cooperation of CESSDA partners. Many of the services outlined above are already provided by members of the CESSDA network or have been established as part of the Data without Boundaries project. Such a solution would most likely require a CESSDA partner to take over the responsibility of hosting the web pages and being responsible for technical administration a task which would undoubtedly require some personnel investment. However to date these services have not been fully integrated. An important lesson learned as part of the DwB project has been that the research community has a heightened interest in a one stop shop solution and a centralized point of reference would provide a first step in that direction. While such an ESCOS 'light' solution could establish a one stop shop online would be dependent on the good will and funding of the participating institutions, a fact that might endanger the long-term sustainability of the project.

As has been argued above the degree of institutionalization and the tasks ESCOS 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 ideally we envision ESCOS as a subunit within the CESSDA-ERIC responsible for official statistics microdata. This would require inclusion of ESCOS in the CESSDA Work Plan and budget as well as a small standing staff. The responsibilities of this service unit would mainly be aimed at coordination and integration of services provided by CESSDA partners who based on principles of subsidiarity would continue to provide services. Such a solution would on the one hand ensure better integration of services but would also be able to ensure that services and programs established today are sustainable.

In relation to national OS microdata the mission of ESCOS would be to inform researchers in the ERA on existing data sources and access conditions and to establish an infrastructure which makes cross border access to such data more feasible. Important services would include provision of English language metadata for all national OS microdata in Europe, comprehensive information on access

conditions and ideally centralized management of cross border data access requests. The work of ESCOS should be that of an information broker and advocate for researcher interests and harmonization of access conditions. Existing cooperation between data archives and national statistical institutes should be maintained at the national level and promoted as examples of good practice at the European level as ultimately such cooperation is mutually beneficial and of great interest to researchers both in the respective countries and abroad. Thus, ESCOS would not replace national services for OS microdata but rather build on these and increase their importance.

At the European level ESCOS could serve as the central contact for both Eurostat and the research community and facilitate the dialogue between these two communities. Activities such as user conferences, the European Data Access Forum or training courses would be central to establishing this dialogue.

4.3 Steps toward establishing an ESCOS

This section describes a number of steps which we see as necessary to establishing an ESCOS within the CESSDA community. The ordering of these steps is not necessarily chronological, nor are they all mandatory.

- Establish an ESCOS online platform: Likely the first step to be undertaken towards establishing an ESCOS. On the one hand an online platform offers researchers great utility and on the other hand many of the services to be offered here have already been established at least partially. An initial implementation of such a site could limit itself to a mere collection of links and a schedule of activities.
- Commit CESSDA partners to participate in ESCOS activities: In order to ensure the actuality of metadata and tools as well as full coverage of European OS microdata much work is required. For example, the CIMES system covers data from Member States to varying degrees. Providing English language metadata for national studies heavily depends on language competencies thus extending the number of involved institutions would be vital for further developing this service. Equally DwB partners have contributed to the metadata in MISSY. The MISSY staff at GESIS will however not be able to keep metadata for all of these studies up to date, the same holds true for routines. Furthermore while the DwB project has stimulated tighter cooperation between CESSDA partners, this cooperation needs to be strengthened to ensure its persistence.
- Establish a regular schedule for trainings and conferences: Training courses on integrated European microdata should be hosted regularly in all parts of Europe. A concept for such courses has been developed and refined as part of the DwB project. Beyond the obvious utility of informing interested researchers on the use of OS microdata, getting in touch with researchers from throughout Europe and informing them on the services provided by ESCOS would be an important step in popularizing its services. Furthermore establishing such cooperation among various CESSDA partners and Eurostat would be beneficial to strengthening and extending ties established as part of DwB. The same holds true for user conferences where researchers and data producers come together to discuss the scientific research with and the methodological progress of integrated European OS microdata. In a first step these joint activities could be based on a memorandum of understanding.

- Establish an ESCOS staff: The amount of dedicated staff required for an ESCOS would be highly dependent on which tasks this staff would need to fulfil. As a general rule of thumb the more decentrally ESCOS is organized the less standing staff it will require. Once an online platform has been established and if all services provided online are handled by CESSDA partners the maintenance, updating and administration of the online platform should be handled by central staff. This staff could also coordinate training courses and organize an annual conference and potentially also oversee researcher requests for cross border access. However this would require those partners responsible for preparing and updating metadata and routines to commit considerable resources to these tasks.
- Establish and expand a European Remote Access network: Within the DwB project currently institutions from three countries are pioneering a secure remote access network which connects them between each other and will allow researcher to access sensitive microdata from abroad in safe centres. The utility of such a service for researchers however would increase manifold if the network could be expanded to include a greater number of data providers from throughout Europe.
- Full integration of online platform services: In its most basic form an ESCOS online platform could consist of a mere collection of links to different databases and other services and resources. However in order to increase the utility for users the full integration of these services would be desirable. From a technical perspective standardization of services is a prerequisite for establishing a search and browsing function. Furthermore an integrated service would provide stability of services over time and after the initial implementation would require fewer personnel for technical maintenance.
- Establish regular dialogue and cooperation with ESS partners: One important step towards achieving a tighter cooperation with ESS partners and Eurostat in particular will be establishing regular user conferences and training courses. A further activity which would prove useful in this regard would be the continuation of European Data Access Forum, where representatives of NSIs, data archives and the research community come together to discuss policies on data access.
- Finally, the role ESCOS can play within the realm of official statistics in Europe will heavily depend on how NSIs and Eurostat are involved its establishment and operations. One option would be to include representatives of NSIs and Eurostat in an advisory board.

5 Conclusion

In this report we have developed a concept for a service centre for European official statistics. This service centre should be implemented as a subunit 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.
- 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 anonymization throughout Europe and work towards standardization of accreditation procedures.
- Assist researchers in cross border data access and lobby for harmonization of access conditions and procedures 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 and a secure remote access network which connects these safe centres.
- Possibly assist the ESS in dissemination of microdata either by providing expertise on data harmonization and documentation or even by handling user accreditation.

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 instrumental that:

- A constant dialogue between data producers, service providers and data users is implemented and maintained and that this dialogue incorporates all parts of the (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. The next step to establishing ESCOS would be to extend the sketch presented here and describe in more detail the implementation options and their consequences. This should be followed by the development of a business plan for the preferred option(s).