Project N°: 262608

Acronym: Data without Boundaries

DELIVERABLE D3.4

Convergence in accreditation, legal framework, and information security

WORK PACKAGE 3

Enhancing legal, information security and researcher accreditation frameworks for access to data

REPORTING PERIOD:	From: Month 18	TO: Month 48
PROJECT START DATE:	1st May 2011	DURATION: 48 Months
DATE OF ISSUE OF DELIVERABLE:	April 2015	
DOCUMENT PREPARED BY:	Partners 24, 6, 17, 13, 12, 1 and 23	ONS, NSD, FORS, EKKE, SORS, CNRS-RQ and INSEE

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

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 262608 (DwB - Data without Boundaries).

ACKNOWLEDGEMENTS

The report was prepared by Paola Tubaro, Brian Kleiner (FORS), Stefan Bureli (FORS), Chryssa Kappi (EKKE), Tomaz Smrekar (SORS), Michel Isnard (INSEE), Atle Alvheim (NSD), Trond Pedersen (NSD), Paul Wearn (ONS), Nicola Shearman (ONS), Richard Welpton (UKDA) and David Schiller (IAB)

TABLE OF CONTENTS

INTRODUCTION

The FP7-funded European Commission Data without Boundaries (DwB) project aims to improve transnational research in Europe with a focus on access to the so-called confidential microdata (Secure Use Files). Work package 3 of this project addresses three elements of the framework for access to national data across transnational borders which are currently perceived to be the main barriers to international access, these are; understanding the legal frameworks across the EEA, standardising researcher accreditation practices across the EEA, and guidance on what if any are the best information security architectures that should underpin any access solutions.

WP3 had three interlinked objectives that will help remove barriers for researchers to make use of official data across transnational borders:

- To gather information on, and describe the legal frameworks for, research access to official data collected in the European Research Area. The results will be available for use in policy planning and legal changes and will work as guide for all interested parties on the legal frameworks and accreditation processes for transnational access to data across the ERA.
- To develop and promote a widely-recognised standard for researcher accreditation in order to reduce administration and access costs, and to improve efficiency and confidence in custodians of official data when providing transnational access.
- To identify the challenges and solutions for building and operating a Remote Access facility for confidential microdata in compliance with internationally recognised Information Security standards. This part of the WP also aimed to analyse suitable information security architectures for a co-operational model for linking data centres, and to propose credible information security architecture as a working model.

This report summarizes and articulates the main findings and recommendations from the work undertaken in this WP providing an overall and integrated roadmap for implementation in the future within the framework of a European Remote Access Network (Eu-RAN) and European Service Centre for Official Statistics (ESC-OS) envisaged by DwB on the basis of WP 4 and WP5 work as the necessary infrastructure aligned with the European Statistical System and the CESSDA for removing current barriers for access to official microdata across borders and allowing a major step forward for the international community of social scientists who conduct comparative research, or who focus their work within countries where they do not reside. WP3 findings and recommendations are an essential layer for such a future development.

To meet the ideals of evidence-based policy making there is a constant need for timely and relevant knowledge and analyse which is informed and underpinned by the richest sources

of data. Additionally the growth of the Information Society generates a constant need for knowledge production. The most relevant knowledge of such kind is generated through research. The research community plays an important role in stimulating policy analysis and debate, and it is often their outputs which are vital for assessing the effectiveness of government programs. Research needs data, and good and relevant data become the single most important component necessary to build a science-based understanding of the functioning of the modern society. Along with the development of new statistical tools, highly detailed microdata are increasingly necessary for cutting edge research at national and international level. Providing access to confidential data for researchers while protecting them has become a major scientific challenge that numerous countries have solved setting up specific environment and equipment that ensure access while protecting confidentiality. Many recent publications of foreign researchers in prestigious scientific journals rely now on such data.

In an ideal world the realities of researcher access should recognise that there is a general acceptance of the free movement of goods and services across borders. The mechanisms for researchers to gain access to confidential data should be the same in every country. Researchers should be allowed easy, but appropriately controlled transnational access to confidential data for research purposes. The current frameworks for legal, security and accreditation have national geography as a primary determinant of practices, this is not right and these frameworks should be altered so that they focus on the sensitivities of the data in question rather than quirks of geography and politics. The frameworks for microdata access for research purpose have developed within each national infrastructure separately. This development has rarely taken account of various different international recommendations about principles of official statistics and microdata access for research purposes, with little or no considerations about interoperability on an international level.

Current practices in research access are considered by most stakeholders in the social sciences to be a barrier to access to official data across international borders. The barriers can be “hard” or “explicit” (prohibition of trans-national access found in accreditation, security and/or legal rules) or “soft” or “implicit” (the criteria for access and/or the procedures may be invisible, ineffective, unfamiliar, or inefficient as for example when multiple and different accreditation processes are required for a single project). There is no harmonisation across the ERA for the researchers and practices and conditions for granting access vary widely among countries. Additionally the holders of data are unsympathetic to the problems faced by researchers; this can be caused by many factors such as ignorance of the difficulties, a risk adverse culture in that particular organisation, or a lack of appreciation of the benefits of the research. In any case the effect is the same: international research is often inhibited by national attitudes, criteria and procedures.
The legal framework across the ERA is open and has the potential whereas accreditation and security is closed and is set at a national level. Currently you have, for example, French

accreditation with French security. The next solution would be for French accreditation with a European common security standard. However, the ideal solution is through soft law and to head towards a common standard accreditation and common security standard, for example through the use of a European Remote Access Network (EU-RAN). In the ideal world for the researcher legal, accreditation and Legal frameworks would be all open.

CHAPTER 1 - FINDINGS

1.1 Legal Frameworks

Current Framework

The current legal framework has national geography as a primary determinant of practices. Discreet legal frameworks for microdata access for research purposes have developed within national infrastructures. These are often developed regardless of international recommendations about principles of official statistics and microdata access for research purposes. Alternately some of these legal frameworks are embedded into national infrastructures and predated the publication of these international recommendations and have not been updated to reflect international best practice.

The question is whether the similarities and the differences are essential or superficial. Medium term change is possible when the features considered to be essential are already common to each national system, and where the features that are unique or rare are superficial.

Thus change is possible when essential similarities in national systems become the core of a new standard, and superficial differences become peripheral matters to be addressed only where necessary due to local specificities.

Potential to share confidential data lawfully for trans-border access

We have discovered, in deliverable 3.2, that some NSIs have the potential to share confidential data lawfully for trans-border access where the law allows them to do so. Powers are usually found in the legislation which underpins the national statistical system, or for the mandate for the collection of the data themselves. In some cases the duty for a NSI to promote and assist research access to its confidential data may be made explicit in law.

However in many other cases the power is only implied, meaning that one of the widely-understood purposes of the NSI includes providing access to data, even if this is not written into the legislation available to the NSI. In other words, the absence ("silence") of a provision for providing access specifically to non resident researchers may mean that there is an implied power to do so.

In addition, a NSI may also need a source of permission to use it. Sources of permission may include a gateway with conditions, setting the parameters, which permit researcher access provided certain conditions are met. The requirement for permission may be written into

legislation, or may be implied. It is possible that "silent" permission exists for access to data even if it is not expressly provided for in legislation..

The discovery work conducted via a survey, in Deliverable 3.2, asked respondents to consider the potential and limitations of the use of their statistical legislation.

We have discovered in Deliverable 3.2 that in most countries direct trans-border access supply to researcher is possible in law. For example, France see remote access as an acceptable solution for sharing confidential data at national and transnational level where the non resident researcher can access the data within his own country from a research office in universities. .The report also demonstrated many countries are able to provide access to confidential data to the non resident if they come into the country, for example the UK.. Future developments in this area have been demonstrated with transnational access from another RDC in WP4 of the DwB project where IAB and the UKDA have moved towards transnational remote access with the installation of an IAB server and contract with the UKDA.

In most countries access to official microdata is well established and is now routinely available for residents. While in progress,, trans-border access to microdata, i.e. access to the official microdata that are held by NSOs and statistical agencies of foreign countries, is still an emerging activity. There are several reasons for this, including legal issues, appropriate technical infrastructure missing, standards and differences in the importance of confidentiality between statistical providers.

Remote Access Solutions

According to our findings there seems to be a solution that makes it possible to grant access to the data and keep the data still in a secure environment. Under safe settings, the possibility of remote access to confidential microdata can be considered as a secure and convenient way of providing access for scientific purposes. Virtual PCs and secure remote access can support the technology of remote access even when there are still details to solve, e.g. the assessment of IT securities for the different existing systems. However, the mode of remote access is not implemented and accepted by all countries, and just a very few have a Statistics Act or similar legislative vehicle that allows an NSI to provide access in this way. The discovery task in deliverable 3.2 established that many statistical producers thought that they had the potential to adopt this model as, for some, it was not a legal issues but rather technical issues that were preventing them pursuing this solution.

It is possible to give access to the microdata without delivering any data physically, whereas the copy of the microdata will remain in the secure environment of the data producing authority. There are different modes of access possible which require certain preconditions

for the accessing institutions or persons. Usually the access to official microdata is provided only for scientific purposes to independent research entities.

For confidential data this can be done via a safe centre inside the access facility or via remote access solutions. Scientific use files (anonymised files that are still confidential) are often not considered as sufficient from a researcher's point of view, because they are not detailed enough to run sophisticated analyses or good comparisons between countries. The future development of user demands tends to detailed microdata that can be accessed preferably from a researcher's own workstation inside his own institution. The remote access solutions take these user demands into account. The other modes of access can sometimes be justified, depending on the circumstances. Sometimes it is not even necessary to use highly confidential data or the service and support in a safe centre is just more convenient for the researcher than using remote access on his own.

For security or organisational reasons it is also possible to combine the different ways of access, such as a safe centre and remote access. Usually the microdata are physically in a safe centre, but in combination with remote access it is also possible that the data are accessed from a safe centre whereas the secure server where the data are stored is located elsewhere. This might be a good approach if the data producing authority does not send out any data and the access cannot be granted at the researchers' institution. In this case, a researcher can visit a certified safe centre that is connected remotely to the data at another countries server. Based on the different modes of access combined with a different level of data confidentiality, several zones of trust can be implemented; this concept has been demonstrated in WP4 of the DwB project.

Remote access solutions can only work when countries recognize and trust one another's accreditation criteria, so that they can be more confident when assessing how "safe" a foreign researcher applicant is. Further improvements from IT developments and the recent progress of secure remote access facilities demonstrate the need for a common adoption of a security standard, discussed further within the report, for operating research data centres to serve transnational access. The following discovery chapters look at standardizing accreditation through harmonisation of the accreditation process and the application form as well as a shared information security standard for research access data centres.

1.2 Accreditation

"Accreditation" is the term used to describe the criteria to be met and the administrative procedure to be followed before researchers can access and use data produced by others, including data from official statistics. The current framework for accreditation has national geography as a primary determinant of practices: Research accreditation has developed within each national infrastructure and separately from others. This framework strikes us today as inadequate, all the more so as research tends to be more transnational in scope and to support the policy-making needs of an increasingly integrated world. The demand for cross-national and comparative research to inform public policies is particularly strong within Europe. The core questions, then, are not only whether official statistical agencies grant access to microdata for research use, but also whether they do so in a harmonised manner throughout Europe.

To map the current situation and to propose solutions, the DwB project gathered accreditation information from different European countries to identify the essential core common content, and consulted stakeholders on how to use this material as potential foundation of a new European Accreditation Standard. It also offered active support to improve access conditions, both within each country and in view of enhancing transnational access. What follows is a summary of results obtained.

State of the Art / Common Practices

The discovery work undertaken by DwB included in its scope the criteria for eligibility, the evidence base for decision-making (i.e., the content of applications, forms and contracts completed by the applicants) and the rules and procedures for granting approval for access to data. We obtained this information partly through desk-based research (using publicly available information, primarily from the internet) and partly through questionnaires and meetings with relevant stakeholders, mostly National Statistical Institutes (NSIs) of European countries.

Reassuringly, we found that almost all countries have legal provisions to enable researchers' access to national official statistical data. But despite these globally favourable conditions, practical modalities and conditions vary widely and as a result, researchers located in different European countries have unequal opportunities to exploit the potential of national data. Access to social survey data (rather than, say, business surveys or administrative data) is most common, though with differences across countries that largely depend on their historical experience with surveys vis-à-vis registers. Scientific Use Files (sometimes called de facto anonymised data) are offered in most countries, while much greater diversity is observed concerning Public Use Files (highly anonymised, available to a large public) and

Secure Use Files (confidential versions accessible only through secure IT systems, whether on site on NSIs' premises or remotely).

Essential eligibility criteria are similar across countries, for example the "researcher" status of the applicant and the non-commercial nature of the research project to be undertaken. But these statements of principle (for example, what does "non-commercial" really mean?) are difficult to interpret. Many countries also require affiliation to a recognised research institution (but how does an institution obtain recognition? And, is affiliation sufficient to qualify an applicant as a "researcher"?). Because laws rarely offer guidance on these delicate matters, countries develop their own solutions based on their experience and culture, not always following recognised international best practices. Similarly, application forms and user licences (or contracts) share important commonalities, especially in terms of pledges and safeguards to ensure personal data protection, but again, the interpretation and practical implementation of these criteria are highly heterogeneous.

While few legal systems explicitly recognise the needs of cross-country and comparative research, they rarely ban transnational access to national data, at least within Europe. Most NSIs use this opportunity to enlarge their users' base to non-resident European researchers, but often with limitations or with more restrictive conditions, relative to national researchers. For example, non-resident researchers sometimes have access only to more heavily anonymised versions of the data, or need to travel to the country of interest to use its data. The main reason for caution is the difficulty of applying penalties in case of breach, as provisions included in statistical laws have only national validity.

Visibility of access and accreditation conditions is a practical, but endemic problem that hinders access to data especially across borders. When we first mapped access conditions across Europe in 2011, we noticed that information on data access for research was not systematically available from NSIs' websites; only a handful of them published information, and even fewer made available downloadable files such as application forms, pro-forma contracts, and guidance documents. English-language translations were rare. Conditions and procedures for non-resident researchers were even more rarely publicised on the web. Encouragingly, we notice that these problems are less severe today: a large number of European NSIs have appreciably improved their communication on the web, and essential information on access as well as data documentation, also in English language, is much more often publicly available, and relatively easy to locate online.

Another problem that we pointed to in the early stages of the DwB discovery study was terminology, highly heterogeneous across countries, so that researchers could be confused. For example, there was ambiguity as to what the term "confidential data" referred to, and some countries preferred to denote them "sensitive" or "highly detailed" data. Today, this problem is being mitigated as most countries are converging towards (or at least

understand) a classification of data files based on their degree of disclosure risk, in three categories labelled as "Public Use Files", "Scientific Use Files" and "Secure Use Files", as defined in the Access to confidential data for scientific purposes Regulation.[1]

Thus, despite favourable legal frameworks and global openness of NSIs to researchers' access and in particular, to transnational access, obstacles remain that affect negatively cross-national and comparative research. Heterogeneity of practices, procedures and forms adds burden and delays work, in that a single research project may require multiple applications and contracts, each with its own timeline; there are higher costs if, for example, travel to a data centre abroad is needed; and lack of information and terminological discrepancies discourage users.

Greater coordination across countries is necessary to remove these barriers. A full-blown transnational accreditation model would require all European countries to recognize and trust one another's accreditation criteria, and to rely on one another's decisions, so that they can be confident when assessing how "safe" a non-resident researcher applicant is. While such a scenario, involving a high degree of multilateral coordination and trust, is not yet in sight, it is interesting to notice today that some first experiences in cooperation are seeing the light, at least bilaterally (for example, an agreement for data sharing among the Nordic countries and the UKDA-IAB connection WP4), thereby setting the example and possibly encouraging further attempts in this direction.

Findings from the consultations with NSI representatives

Consultations were conducted with representatives from 22 different NSIs in Europe, following a survey developed within WP3 of DwB (see Deliverable 3.1 for details). The goal of the consultations was to have a better sense of what a European standard accreditation model for transnational access to official statistical data could look like. The model should find the balance between the legal and procedural requirements of NSIs and the needs of researchers.

Most representatives currently hold high-level positions within their institutions and so could legitimately represent the perspectives of their institutions. However, it was assumed that the opinions provided were their own and that others within their institutions might have given different responses, especially for more subjective questions, thus giving more freedom to explore possibilities of progress.

The survey was divided into three main sections. The first part concerned possibilities within NSIs for speeding up access procedures under different circumstances, as the idea was to

[1] http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013R0557&from=EN

explore how far this flexibility could be used for transnational access in case a research project requires multiple accreditations in different countries. The second part involved possibilities for developing an international standard application form for trans-national access to official statistical microdata. The last section was more forward looking and addressed the feasibility of future options at the European level to rely on European infrastructures to facilitate trans-national access to NSI microdata.

With respect to current flexibility in accreditation procedures for access to confidential official microdata, most NSIs had some room for speeding things up, especially adding researchers or variables to the same project. This often involves signing a supplementary agreement or making an amendment to the existing contract. On the other hand, the majority of NSIs declared difficulty in implementing additional flexibility in decisional procedures for already accredited researchers. This is because the focus of many NSIs is on the research project, and less on researchers.

There was however a general consensus on the five key components that are considered to be at the core of a standardized application form (information on the researcher, place where the data will be accessed/processed, requested datasets, research project, and justification). Some NSIs request additional information, especially on data security safeguards, dissemination systems, funding, etc. that could be addressed in an addendum to the standardized application form. Moreover, the assessment of the individual key components is not immediately likely to be standardized since the validation procedures and institutional restrictions are too diverse from one NSI to another.

Finally, NSIs are generally open and interested in various international initiatives for facilitating access to official microdata, while at the same time being aware of institutional and legal constraints. A European metadata database and an accreditation and access procedures database had the most support overall.

There was a certain amount of scepticism among NSI respondents concerning the idea of an ESC-OS, especially with respect to “possible delegation” in the future regarding accreditation. There were also some misconceptions among some of the respondents, regarding metadata issues (some respondents seem to think that Eurostat currently has a database of all NSI metadata) and regarding access (with the idea that an ESC-OS would control data access on behalf of NSIs). Moving toward an ESC-OS will require a step-by-step process demonstrating the utility of the services provided. It should in any case be built in cooperation with Eurostat that many NSIs consider as the natural central point also for national microdata.

1.3 Information Security

The discovery phases in chapter 1.1 (Legal frameworks) proposed potential solutions through alternative modes of dissemination like remote access or remote execution solutions. This requires a broader focus, both production and dissemination. To make data collected by the official statistical system available for analytic use, certain conditions have to be met. Safeguarding data confidentiality[2] is essential from a legal point of view and is important for building trust between institutions and between data access facilities and the population. The national laws and regulations set the premises for confidentiality. If data confidentiality is convincingly secured, trust will be built both nationally and internationally. To serve research with the required flexibility will put focus on data maintenance and data dissemination. Such services could be more decentralized we witness today, but among other things requires an adequate Information Security Management System (ISMS) as an internal bridge and facilitator.

The ISO27001 Information Security Standard is a comprehensive accreditation for information security, and the scope of this DwB WP3 task has been to present a framework for the design and management of an ISO27001 compliant ISMS that can be used across a wide variety of data providers, to achieve a recognizably high standard of information security to support circles of trust, and easing and facilitating more efficient access to data across borders.

In discussions of information security the security model developed by Office of National Statistics in the UK could be an illustrative starting point: The accepted aim is not to prohibit, but to develop safe use of data. It is recommended to move from risk avoidance strategies to risk management strategies. There are four components we have to investigate, the data, the user, technologies and analytical outcomes.

The ISO27001 Information Security Standard is a comprehensive accreditation for information security, and the scope of this DwB WP3 task has been to present a framework for the design and management of an ISO27001 compliant ISMS that can be used across a wide variety of data providers, to achieve a recognizably high standard of information security to support circles of trust, and easing and facilitating more efficient access to data across borders.

In discussions of information security the security model developed by Office of National Statistics in the UK could be an illustrative starting point: The accepted aim is not to prohibit, but to develop safe use of data. It is recommended to move from risk avoidance

[2] The problem of serving research with data collected for statistical purposes are well described here: http://www1.unece.org/stat/platform/display/confid/Managing+Statistical+Confidentiality+and+Microdata+Access

strategies to risk management strategies. There are four components we have to investigate, the data, the user, technologies and analytical outcomes.

Degree of information security is usually defined as a product of these four (or maybe seven) factors:

Input: Safe data, which usually means anonymized data, a technology commonly used for sample-based data. Anonymization solves the problem by removing information, so scientifically it becomes a problematic solution since it reduces the value.

Output: Safe results, so-called disclosure control, to prohibit those analytical outcomes are used to identify respondents. Analytical techniques differ, regression is less dangerous than tables, since there is more aggregation and less details.

Users: This is a hierarchical concept. Focus on institutions means to put focus on legal responsibility, focus on projects are more concerned with scientific justification and quality aspects, while individual users represent individual qualifications.

Technology: Technical architectures and safe communication are also important factors, included in this are all management procedures.

This model may be implemented in many varieties. The problem is to find the version that generates acceptable security at the same time as it give users flexibility and access to rich research data. Across Europe we basically find 5 varieties of data dissemination:

1. Public use files / Scientific use files, anonymized files for public or scientific use.
2. Offline Research Data Centres admitting access for authorized users.
3. Dissemination of non-anonymised data files to authorize users, users with required permissions.
4. Remote Execution input of analytical jobs and return of controlled output.
5. Remote Access, direct access over the net, with outputs returned through a controlled procedure.

Changes

The most important components of an ISMS for handling and deposit of data so that confidentiality is protected are:

- Physical barriers
- Technological barriers
- Good routines/management - creating a culture of information security'

The known desirable features related to European research data access centres need to be seen from two perspectives:

- Management: Secure storage and appropriate handling of research data, to build confidence and maintain quality of resources.
- Dissemination: Efficient law-abiding data sharing and dissemination, through appropriate information systems and transparent and clear access policies.

High level of security on one hand and easy access to research data on the other may not be compatible. The solution is to seek a balance. Such a balance must, in any case, be consistent with current rules and regulations.

CHAPTER 2 - RECOMMENDATIONS

2.1 Legal Frameworks

In an ideal world, the legal framework concerning the access of national confidential data should be the same in every country and should allow a large and easy transnational access to confidential data for research purposes.

However, we are not in an ideal world and the present legal frameworks are in the best cases different and in the worst case directly incompatible (see the result of the discovery phase in Deliverable 3.2) and, although we could recommend to NSIs what to do to achieve compatibility, waiting for this to happen will take time and would require transnational legislation changes to occur, which is not a short or medium term solution for researchers.

The recommendation is therefore not to harmonise the legal frameworks themselves, but to harmonise their use though the European Research Area (ERA). However, it should be noted, in certain cases some changes in the law can be necessary and many countries have made some changes allowing more access for the researchers.

The results of the legal framework survey in Deliverable 3.2 support this view as a possibility. The majority of countries which answered the survey thought that their legal framework allowed transnational access, even if their legal framework doesn't explicitly allow this type of access and was silent on the law. Every ERA national institute of statistics should, according to this recommendation, try to implement in the most favourable way for researchers their legal framework. For example, they should allow a request for access to be submitted in a foreign language if no national law is expressly prohibiting it.

Some prerequisites

Before, or in parallel with, improving transnational access, some steps would have to be taken by the national statistical institutes. Namely, it could be:

- Enhanced metadata should be easily available at least in English (discussed within the report through a suggested European Centre for official statistics).
- Data should be made available easily for the researcher (discussed within the report).

Some consequences

In this chapter, we demonstrate the impacts of the recommendation suggested within this report and provide a roadmap for NSIs in instances where national legislation doesn't prohibit them.

The first step is to allow the non resident researchers in national research institutes to access national confidential within the Country where the data is held. For researchers belonging to a member state of the EU, that step should not be too complicated to take, because of rules implemented in the EU. The same steps should be taken for other ERA countries.

The second step should consist of accepting access from another ERA country. To help this type of access, the use of a secure centre in which the data is being kept and remotely accessed (such as the EU- RAN concept). Note that, in this second step, accreditation is still given by the NSI.

DwB WP 4 "Improving Access to OS microdata" worked on a concept for a European Remote Access network to confidential microdata. A EuRAN could work as backbone for European research by connecting researchers with data sources all over Europe. The EuRAN supports different access points (safe rooms, institutions, and unrestricted access) adjusted to the security demands of different data sources (secure use files, scientific use files, and public use files and data documentation). The most important aspect of the network is data protection and security.

Our discovery work in Deliverable 3.2[3] revealed this kind of data access is allowed in majority of observed countries, sometimes with restrictions on administrative data. Many NSI's point of view is that secure remote execution is the safest possible solution. This solution can be seen to address the legal issues and constraints over confidentiality.

A European research infrastructure can only work, if data providing institutions (NSIs, data archives, etc.) are working together. DwB has started to create an organisational network for Europe based on understanding, trust, harmonised workflows, agreements, and contracts. Consequently, the concept of trust around institutional agreements (discussed further in the report) should be encouraged to minimise the risks around disclosures and encourage open data for researchers.

Currently highly detailed and confidential microdata can be reached through three different access ways: (1) via guest stay at the facilities of the data provider, (2) via remote execution / job submission and (3) via remote access. Remote execution or job submission does not allow seeing the confidential microdata the users want to work with, but they can work from their own workspace and receive calculations computed on the basis of the original microdata. Remote access offers access from a distant location to confidential microdata. The users can see the research data and work as if it were located on their own workspace, though not able to download the data.

[3] For further details, see D3.2 - "A report on the legal frameworks for data access to Official data and an interactive guide" (available here)

Therefore, remote access gathers two important advantages for the users: first, no travelling is needed; second, the researcher can see both data and data structure just as interim results. The suggestion for improving access to OS microdata will therefore be based on remote access solutions as suggested from WP4.1.

Findings in WP4.1 suggested access to OS microdata is improved based on remote access and further describes eight remote access solutions in Europe implemented by NSIs and CESSDA archives. The findings of this report showed that the solutions vary according to different legal, organisational and technical environments. The report proposed that the existing landscape of separated data access solutions could be improved by connecting them into a European Remote Access Network (Eu-RAN), with European Service Centre acting as an information platform. The DwB WP5 reports develops a concept for a European Service Centre for Official Statistics which could function as a research infrastructure institution for European Official Statistics microdata. The Eu-RAN would allow facilitate transnational access in cases NSIs require that access should only be given from another NSI (some other as in France may allow access from the researcher office in universities. It should also facilitate the transfer of data from one RDC to another one for research projects requiring merging datasets from different countries. The development of such a Eu-RAN and ESC-OS require a third step.

The third step could be to make specific agreements between NSIs/RDCs so that a foreign NSI/RDC could give access to data from another NSI/RDC or accept to transfer data in another NSI for combining datasets.

This step looks at the 'concept' of trust and institutional agreements. The 'concept of trust' for international microdata access relates to the different possible levels of data confidentiality and security requirements necessary to access those data. According to the level of data confidentiality there are different grades of risk assessments: while original data (non-anonymised data without direct identifiers) have a very high risk of disclosure, public use files have almost no risk at all.

The producers of official statistics in most countries would only provide data to a third party or another country if they are obliged to by law, or if it is explicitly allowed by law. In this context, there can be a big difference between delivering microdata (physically transferring microdata to the custody of another organisation based in a different country) and giving access to microdata. Even when data producers have to deliver data physically to a super-ordinate institution, they would like to keep control of who is accessing the microdata. At the same time, the data producers are willing to give access to their data to international researchers on the basis of their Statistics Act.

Depending on the national law, it is already possible for some countries to do so, according to our discovery survey. In France the CASD allows remote access from a research office in universities. This concept has been demonstrated for countries that would only allow

transnational access from another RDC in WP4 of the DwB project where IAB and the UKDA have moved towards transnational remote access with the installation of an IAB server and contract with the UKDA. For other countries it is only allowed under special conditions and for some it is completely prohibited.

The concept of "circle of trust" is based on the agreement that each member is accepted according to the same rules and conditions that are approved by all members. In the context of statistics, this mainly refers to confidentiality rules and security requirements but also to competence and legal aspects. This makes it possible to create a group/membership of trust, according to the level of confidentiality and risk of disclosure.

Why is a concept of trust needed?

If international microdata access is prohibited by a national Statistical Act or a similar instrument then a concept of trust could be seen as irrelevant. In some instances, where the law allows access and the conditions of this access are explicitly enshrined in legislation, and then there might also be no need for a concept of trust. However, there are situations where access for another country or a third party is not mentioned in the national law, implying that it is not forbidden from a legal point of view. There are two interpretations of the law in this case. Some countries treat everything that is not explicitly forbidden as allowed and some countries treat everything that is not explicitly allowed as forbidden. Therefore, even when giving access to third parties is not forbidden, some of the National Statistical Institutes still do not want to give access as long if they are not obliged to by law. Therefore, the "circle of trust concept" is needed for three reasons.

1. A "circle of trust" can bridge the gap for those countries where the law does not mention anything at all. For these countries it would be helpful to have a basic concept of requirements so that they are reassured that their data are accessed in an organised and secure manner, comparable with the security requirements in their own country.
2. A circle of trust can reassure those countries that basically allow access to foreign third parties. A law can be broken, but it is possible to prevent a breach of confidentiality with an organisational or technical solution and a reasonable amount of effort. Their data will be accessed at least under the same security standards as in their own country.
3. Countries that do not allow microdata access to foreign third parties should be encouraged to reconsider their position; if the data are accessed in a very secure way at least comparable with their own standards, there is actually no practical reason to refuse microdata access in the long run. There is always the opportunity to deliver anonymized datasets at the beginning.

Preconditions can be set up with regard to the institution requesting access and to the specificities of the technology for providing access. The minimum requirement for most data producers is that data that are taken from another body are treated at least under the same

or higher level of confidentiality as in their own premises, based on the degree of data confidentiality with their corresponding risk of disclosure.

An agreement on the rules and standards for joining the ‘circle of trust’ and a common understanding on security and accreditation standards is needed. It would also be advantageous; if a research institution that is accredited by one country for microdata would be automatically accredited in another country.

2.2 Accreditation

The findings from work within WP3 on accreditation indicate that various options are feasible, and some evening promising, for facilitating transnational access to official microdata. NSIs are generally very willing to consider these options and to weigh the benefits in relation to the costs, all within existing systems of rules, practices, and constraints. Below we describe these options regarding accreditation in the form of recommendations for the future, from the easiest to implement to the more remote, complicated, and difficult solutions.

Short-term advances

1. Provide more information for researchers about how to find and access data

An encouraging finding from the discovery work, and from follow-up meetings with DwB members, European NSIs, and other stakeholders, is that some of the obstacles to transnational access can be relatively easily removed: Indeed much progress has been made since our first information-gathering effort at the beginning of the DwB project, in 2011. We have mentioned above the lack of visibility of conditions for access and accreditation, and the improvements observed in the last four years. Partly, these developments are due to information flows and peer pressure, from Eurostat and from fellow European NSIs; seeing the example of others is likely to have facilitated transmission of knowledge, awareness of best practices, and solutions to problems, throughout Europe.

The DwB team itself has engaged in a significant effort to contribute to improvements in this sense, raising awareness of the widespread visibility problem among NSIs and other data actors in numerous meetings and conferences between 2012 and 2013. Further action will have to be taken by NSIs and other relevant data actors in the near future to pursue this process further and gradually increase the availability of information, especially in English. The work done over the four years of the DwB project suggests that it is vital to maintain regular opportunities to share information and expertise at European level, in order for each country to be able to sample experience and learn from others, as a basis to upgrade and improve their own system.

Another significant recent progress, mentioned above, is growing consensus around a more homogeneous terminology, which has mitigated, though not completely removed, the discrepancies observed even just a few years ago. In part, this is the result of convergence of debates within the European Statistical System, culminating in the approval of the European regulation of access to confidential data for scientific purposes (EC 557/2013 of 17 June 2013). The OECD has also contributed to this process with the 2014 report of its Expert Group for International Collaboration on Microdata Access.

DwB contributed too, by proposing a Glossary to classify countries' data provision, now available online. The terminology used is now familiar to all member states and it is easier now to classify national files in light of the criteria adopted at European level. In the near future, further work needs to be undertaken to remove remaining ambiguities (for example, the degree of anonymisation of Scientific Use Files sometimes varies widely even within the same country), to trigger a process of greater harmonisation and streamlining of practices across countries, and ultimately benefit transnational researchers. Again, experience over the last 4 years points to the importance of opportunities to share information and engage in multilateral discussions, among NSIs and other relevant data actors, at European level.

2. Maintain a database of accreditation procedures

To contribute to alleviating the problems arising from lack of online information, in 2014 the DwB team has publicly released the database of accreditation rules and procedures that it compiled as part of its discovery work, in order to provide data users with an extra informational tool.

This database, hosted by the DwB website[4], contains country factsheets that will hopefully support researchers in their data discovery efforts, providing essential information in a standard format and similar terminology (explained in detail in an annexed Glossary[5]). There are 32 such factsheets, one for each country covered (Austria, Belgium, Bulgaria, Cyprus, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Ireland, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovenia, Slovakia, Spain, Sweden, Switzerland, United Kingdom). Focus is on the data produced by the NSI as the leading statistical agency within a country (except in Germany where 2 data-disseminating institutions are covered: the NSI, Destatis, and IAB, a relevant international data actor and DwB member). Together with basic information (conditions, modes of access, timing, costs etc.), there are links to all the relevant websites, application forms, pro-forma contracts and other online documentation provided by national institutions.

While potentially useful, this tool needs further development and future maintenance to really serve the needs of researchers in search of information about data. It now consists of a set of static web pages with limited cross-referencing, and almost no interactive features, a set of characteristics that necessarily make it less attractive. A more user-friendly interface should be developed, possibly with the help of graphical tools and more opportunities for interaction. Further language improvements are also in order, notably to ensure any technical terms and jargon are replaced by more common words, so that researchers who are versed in the use of data, but not in the intricacies of the data access system, may still fully understand the contents.

[4] http://www.dwbproject.org/access/accreditation_db.html
[5] http://www.dwbproject.org/export/sites/default/service/accreditation_db_pdf/dwb_accreditation-factsheets_glossary_july2014.pdf

More importantly, the database will require maintenance and constant updates over time. URL addresses are replaced constantly, accreditation systems change, and even laws are subject to modification at regular intervals. For the database to be useful over time resources must be devoted to this work and a consultation mechanism must be created to allow NSIs to communicate any change that should be reported.

Mid-term options

Promote adoption of a standardised application form for transnational access

Following the findings of our consultations of the NSIs on accreditation practice, and discussions at the DwB workshop in Lausanne, we recommend a standardized application form for transnational access to official statistical data as a realistic mid-term step towards a convergence of accreditation practice.

There is indeed a general consensus that the accreditation must be based on the project for which the data is requested, rather than solely on the requesting institution or researcher. The NSIs also agree on the five key components that are considered to be at the core of a standardized application form: information on the researcher; place/institution where the data will be accessed / processed; requested datasets; research project; and justification. These key components must be addressed in the standardized application form, specifically:

- The information on the requestor(s) should include details on all researchers who will access the data;
- Extensive information on the institution(s) that the researchers are affiliated with is absolutely crucial, since a data requestor/researcher must be affiliated with a "safe" institution, whether he/she is an academic researcher or a statistician affiliated to a public institution;
- The application form must mention all requested survey datasets/years. One application form can generally include several datasets. If this is not the case, one application form must be submitted;
- The description of the research project will be an open-ended text field where the data requestor(s) should explain the purpose of his/her project, the runtime of the project, as well as the expected scientific output. It should not exceed one page A4, but should contain all relevant data;
- The justification why the requested data are needed (and in case Scientific Use Files exist why they are not sufficient) should be clear and convincing. An account of the analytic methods that will be employed in using the data should be included.

However, some NSIs request additional information, especially on data security safeguards, dissemination systems, funding, etc. that could be addressed in an addendum to the standardized application form. We recommend that the addendum be conceived in a modular approach. For example, a data protection module should explain how the data will be managed and kept in a secure environment. It must provide a detailed description of the

computer system where the data will be stored and analyzed, the locations where copies of the data will be kept, how the data will be stored (when used and when not used), a description of how the data will be transmitted between the researchers, the methods of storage of computer output and when and how the data will be deleted, for example where confidential microdata are transmitted to the researchers via CD Rom. In the case of remote access, remote execution and on site access, the access system itself provides this secure environment. With this modular approach, the NSIs can choose from a number of specific modules to request the additional information relevant to their accreditation process.

This standardized application form, as well as the additional modules, could be easily translated and made available for download in all relevant languages. The standardized application form and the different modules requested by the NSIs could then be integrated into the accreditation database, described earlier.[6]

The assessment of the individual key components of the standardized application form and the addendum will remain with the NSIs, since the validation procedures and institutional restrictions are at this point too diverse from one NSI to another. The fact that the assessment remains in the hands of the NSIs will facilitate the acceptance of a standardized application form.

The researchers would benefit greatly from such an approach when requesting data from different NSIs, since they can basically complete and submit the same application form. The additional workload to respond to more specific requirements in an addendum will be minor.

Since the validation procedures and the assessment of the individual key components will continue to be in the hands of the NSIs, at least in the short term and due to the modular nature of the addendum, we hardly see any barriers to a standardized application form and at the Lausanne workshop, many NSIs saw that as possible.

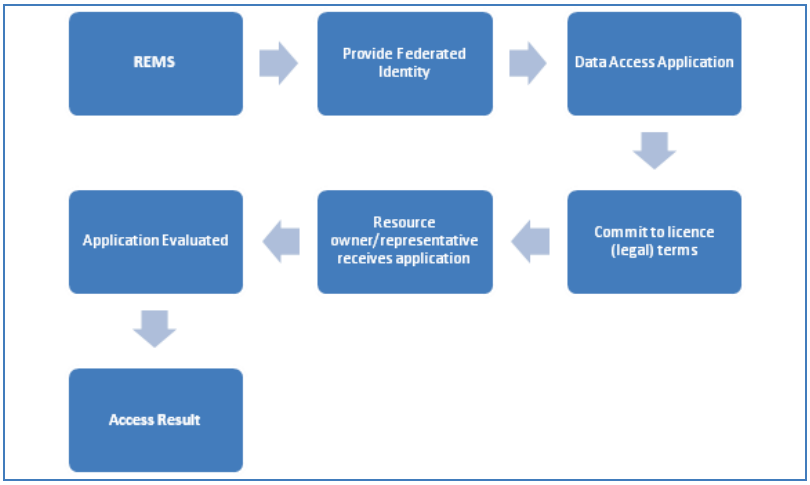
Long-term possibilities

1. *Resource Entitlement Management System [REMS]*

The Resource Entitlement Management System [REMS] is an electronic application tool for accessing sensitive data-sets for research purposes; it's design and specifications are in accordance with the requirements of biomedical data access control; nevertheless, its presenters declare that *"the tool can be used to control access to versatile resources",* such as environmental sciences, economic data-sets, humanities and social sciences (Linden, Nyrönen and Lappalainen, 2013)[7]

[6] http://www.dwbproject.org/access/accreditation_db.html

[7] Linden Mikael, Tommi Nyrönen and Ilkka Lappalainen. "Resource Entitlement Management System." Trans-European Research and Education Networking Association. 24 April 2013. pp.12.



The path to data access through REMS

REMS prerequisite is that applicants need a *Federated Identity* as authentication means. The question to be asked is whether the current *Federated Identity* concept covers the accreditation needs discussed.

Based on the research and consultations findings, although NSIs evaluate affiliation to an approved home institution for granting access, the prevailing trend is to base their decision on evaluating the project for which data is requested. A major shift to the concept of federated identity is therefore needed, for the adoption of REMS or a similar system.

REMS features which are functional to the microdata access case

Feature 1	Cross-organisational identity management
Feature 2	Infrastructure for e-signature for commitment to the licence terms (acc. to EU Directive 1999)
Feature 3	Supports research group applications, as well as, individual members application
Feature 4	Application is "open" in the system for the life-span of the (approved) project, thus allowing addition of new research members and/or more datasets.
Feature 5	Reporting on data requests and data use statistics is provided by the system (audit trail); the system requests the data user(s) to provide information on publications arising from the data.
Feature 6	Data-set download AND remote access to data are supported, depending on the restrictions set by the data owner
Feature 7	Authentication of all parties involved (applicants / data owners / data providers / approvers) includes: name; title; e-mail address; home organisation; type of affiliation.
Feature 8	REMS is NOT a resource discovery tool. It mediates the application process once resource is discovered through other infrastructure(s).
Feature 9	The tool developers plan to implement it on an open source licence.

In chapter 3.4 of D3.1 a Statement of Requirements and Design Principles is presented as the minimum for the ERA research accreditation standard. This 'ideal' standard programmatically "ensures the acquisition, compilation, maintenance, presentation, and preservation of authenticated baseline evidence relevant to access decisions that uphold the UNECE Core Principles for access to confidential microdata. A REMS catered to these particular accreditation needs must provide these requirements. In the table the existing REMS provisions are tested against the baseline evidence required for an ERA research accreditation standard.

Authenticated Baseline Evidence Required for an ERA Research Accreditation Standard	**Does REMS provide?**
Identity of the researcher(s)	YES
Identity of relevant employer (legal personality)	YES
Identity of the owner of the data (legal personality)	YES
Identity of data access service provider (legal personality)	YES
Record of previous uses of ERA official microdata	YES
Record of breaches of terms and conditions of use of ERA official microdata	YES
Indicators of qualifications and career status	Enhanced Federated Identity Data ?
History of training in confidentiality risk management	Enhanced Federated Identity Data ?
Descriptive indicators of statistical purposes, and a generic statement	Possible Addition
National territory of the law applying to the access	Possible Addition
Reference to other relevant binding agreements	Possible Addition
Identification of the relevant licence applicable to the data	YES
The sanctions and penalties applying in the event of a breach of rules	Possible Addition
Applicable information security standards	Possible Addition
Applicable rules of conduct for the research facility	Possible Addition
The purpose of the research	YES
The data sources used	YES
The research facility used	YES
The location and availability of the research results	YES
The location and availability of derived data generated by the research	YES

Evaluation: would REMS do the job?

It seems that in some cases the Federated Identity Concept, as it is currently adopted, needs further specialisation, if possible. Further, provision of additional documentation is needed to cover for the enhanced needs of the type of accreditation under discussion.

Does the current REMS cover for the proposed DwB-WP3 design principles?

Design Principles	**Description**	**Provision status by existing REMS**
Design Principle 1	Data access decisions are made by the data owner or a specific committee, being the owner of the risks in the data. Decisions are to be made on a project by project basis for confidential data. In all instances the Data Owners should seek any relevant views / advice before reaching a final decision	Provided
Design Principle 2	The evidence for a data access decision should, wherever possible, be obtained from independent and authenticated sources. Self-reporting by the researcher, in particular free-text entry should be kept to a minimum.	Provided
Design Principle 3	Relevant evidence, and no irrelevant evidence, should be used to make a data access decision. The relevance or otherwise of evidence is a matter for the owner of the risks in the data, provided the Core Principles are respected.	Cannot be controlled by REMS
Design Principle 4	The evidence used to make a data access decision should be preserved until the liability for the decision elapses.	Provided
Design Principle 5	The evidence should be precompiled into the relevant categories determined by the data owner and available at the moment the access proposal is submitted.	Provided
Design Principle 6	Researchers should be able to see and seek correction of their own information.	Provided
Design Principle 7	The public should be able to see and use freely some of the evidence, in particular that relating to the purpose of the research and the location and availability of results.	Provided (in the form of audit statistics)
Design Principle 8	For efficiency, reliability and authenticity purposes, all evidence should be compiled in a single pan-ERA European Research Accreditation Service, maintained by a trusted third party.	Provided
Design Principle 9	The European Research Accreditation Service should support a process flow for accreditation decisions that enables all parties to use some, or all, of the administrative procedure in place of existing local administration.	Provided

Design Principles	Description	Provision status by existing REMS
Design Principle 10	The European Research Accreditation Service should be designed in such a way as to allow for its integration into pan-European Resource Discovery, Pan-European Web Single Sign On, Metadata, and Remote Access Network architecture.	REMS is not part of resource discovery tools

2. *Integrate a transnational accreditation system within a European Service Centre for Official Statistics (ESC-OS)*

While implementation of the short- and mid-term accreditation recommendations would certainly improve the situation considerably for researchers with respect to transnational access to official microdata, they would leave still other difficulties concerning access. Most importantly, researchers would still have to pass by each relevant NSI from which they are requesting data, even if a common application form was used. Then their applications would have to be reviewed separately by each NSI. In the end, without a more centralised system, researchers could still have to wait long periods to finally have the data they need for their comparative analyses.

In our view, what is needed in Europe is a centralised system where researchers can discover and then apply for data from different NSIs *all in one place*, following the same procedures and using the same forms, through the use of a European Service Centre. Such a central transnational accreditation data access system for official statistics would save time and effort for researchers, would reduce the burden on NSIs, and would facilitate and strengthen comparative analysis internationally.

The European Service Centre for Official Statistics (ESC-OS), conceived within WP5 of DwB, is envisioned as a way to facilitate access to NSI data across national borders. The ESC-OS would be a research infrastructure institution aligned with the European Statistical System and the CESSDA network of European Data Archives, and would promote and "broker" the scientific use of European official statistical microdata by providing enhanced metadata (in English) and a central data discovery and access facility. Through the ESC-OS, researchers would be able to locate relevant data for their projects, and then order the data from the respective NSIs.

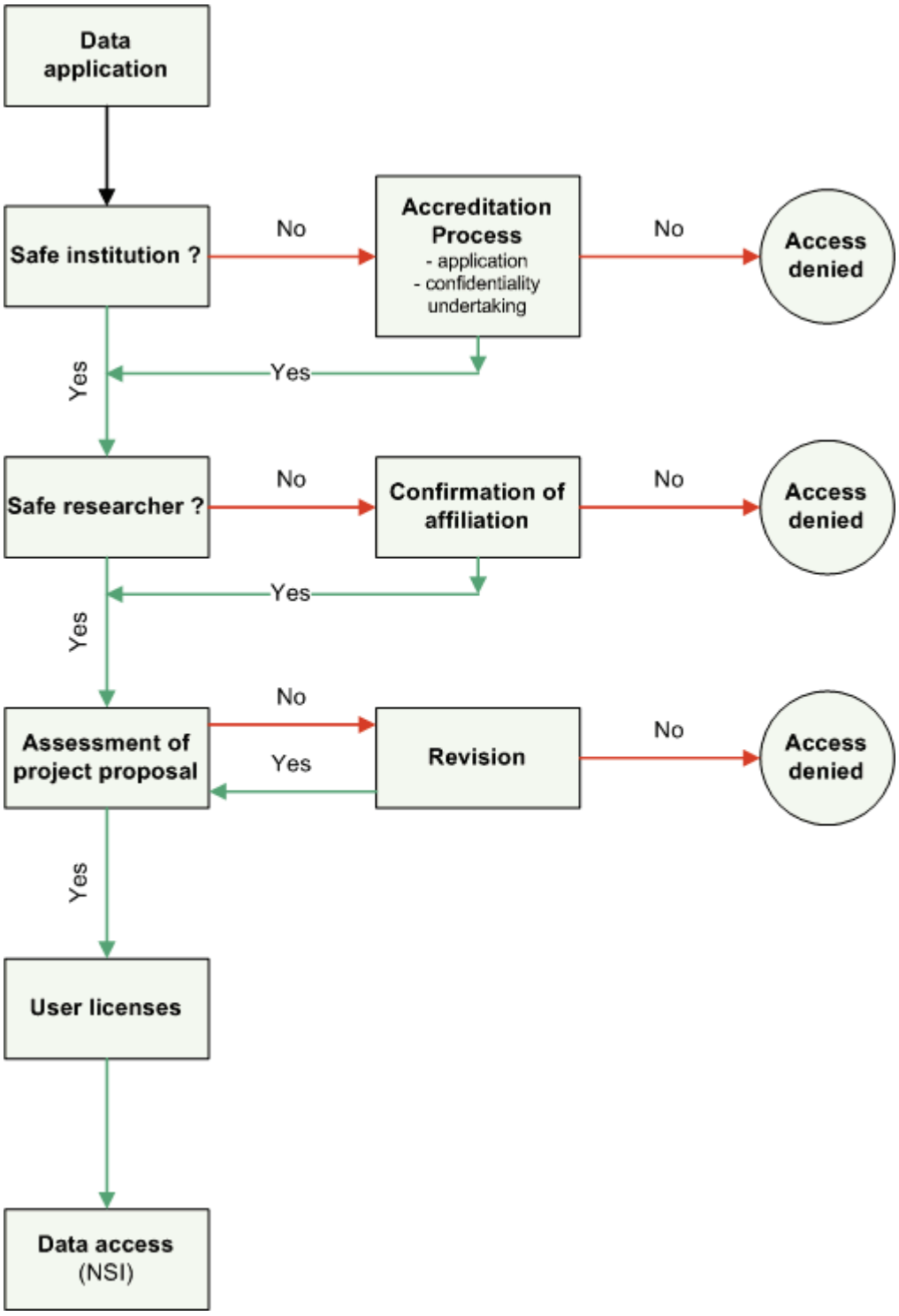
A transnational accreditation and access system built into a future ESC-OS would be an important step forward for the international community of social scientists who conduct comparative research, or who focus their work within countries where they do not reside. Such a system, represented in Exhibit 1, would have the following general features:

1) A researcher interested in data from a particular survey would complete a standard application form online.

2) Upon application, the ESC-OS would assess the eligibility of the researcher based on his/her affiliated institution and the features of his/her project proposal.

3) If all eligibility criteria are satisfied (as determined by the ESC-OS and/or the NSI), then a standard user license is prepared and sent to the researcher and institutional representative for signing.

4) Upon signing of the contract, the NSI is notified that the data may be released.



There are different possible configurations with respect to such a transnational accreditation and access system, depending on the extent of involvement of individual NSIs in the ESC-OS. Assuming that an NSI participates in the ESC-OS and is open to making certain data available to researchers outside of the country, there are three options:

- First, an NSI can "opt-in" fully, meaning that the ESC-OS would process and evaluate all incoming requests for data, and would manage the application and contractual requirements. If a decision has been made to provide the requested data, the data requestors would sign a user license agreement, and then the NSI would be notified and would make the data available to the researcher or team of researchers.
- Second, an NSI can opt-in partially, meaning that the ESC-OS would still process and evaluate the request, as well as manage the application and contractual steps, but that the NSI would have the final word on whether access should be granted.
- Third, it is possible that the NSI would prefer to evaluate all data requests using its own terms, forms, and decisional procedures. The ESC-OS would simply transfer researchers requesting data to the appropriate NSI contact person.

As part of the collaboration between the NSIs and the ESC-OS, there should be a negotiated agreement on the rights and obligations of each partner. Within this agreement, the extent of involvement of the NSI should be stipulated (e.g., full, partial, or no opt-in). In addition, the agreement should specify any special conditions insisted upon by the NSI regarding accreditation and access. For example, an NSI may insist on an application form that includes certain features that are not contained in the standard ESC-OS form. The agreement thus would cover all aspects of the accreditation process to the satisfaction of both the NSI and the ESC-OS.

Such a system for transnational accreditation and access would benefit NSIs by providing an external front office for handling international data requests, where applications and contractual matters would be carried out by a third party (the ESC-OS) following agreed-upon steps and terms. This would free up resources for NSIs that would otherwise be devoted to addressing the data requests of international researchers. Another benefit to NSIs of a collaboration with the ESC-OS would be increased visibility and discoverability of their data, and increased institutional visibility at the international level (e.g., in scientific journals that cite data sources, or by European-level policy-makers who orient toward scientific findings). However, such a recommendation would require an extensive knowledge. The ESC-OS is envisaged as a distributed system where national data archives are engaged and provide the necessary competences and knowledge required at national level.

Further, researchers would benefit greatly from such a system. First, after identifying the relevant data, they could apply all at once at the same place for data from multiple NSIs using the same forms, rather than applying separately at multiple NSIs using diverse forms. The data request process would thus be simplified and faster. Also, within a centralised

system, researchers would not need to be accredited separately by different NSIs according to different rules, but once only according to an accepted international standard. Stored information about researchers could be used by the ESC-OS to expedite even more the accreditation and access process for future data requests.

Of course, such a centralised solution for speeding up transnational access to official microdata would require a level of harmonisation and standardisation that currently does not exist. There would be many barriers, including the often cited fear of “delegation” to a third party, especially to a third party in a foreign country. It should be reiterated however that all data would remain in the possession of the NSIs, and that only the NSIs would ultimately provide access to the data. In this sense, the centralised management of data requests would stay within the bounds of current national data protection laws and security requirements. Most likely, such a centralised approach would be implemented gradually, beginning as an initiative among several NSIs within a circle of trust and linked to a EU-RAN where NSIs involved in such a secure system will concretely face the needs for an harmonisation and simplification of the accreditation process and could come up to multilateral agreements.

2.3 Information Security

Ideal practice

As trans-border supply of confidential micro level data is not possible by law in some countries, Remote Access solutions (RA) or Research Data Centres (SDC) comes up, in DwB discovery work, as a good tool for improving the situation.

From an information security perspective, such solutions require the following:

a) Safe environment
b) Safe technical facilities
c) Safe users
d) Safe data

In order to control the 'human component', a satisfactory accreditation system should be in place. Even though all thinkable precautions are taken, one cannot have a fully robust system. By controlling the users you can minimize the risk of events to an acceptable level.

Safe data essentially means to ensure that the confidentiality is not being compromised, given the measures a) through to c). This is done by controlling the analytical outputs or the actual data if direct insight to data matrixes is permitted, i.e. disclosure control.

Safe data may not be that relevant as the physical and user components, as RA and SDC typically give researchers direct access to microdata through legal requirements that are to be fulfilled. However, access may in some cases also be given to restricted or anonymized data only. This may apply to data regarded as sensitive by the data owner. Examples may be state secrets, health related information, criminal records etc.

Both RA and SDC give researchers the opportunity to analyse, or even see actual microdata (if permitted) in another country, which opens for increased researcher collaboration across borders and overall increased use of micro level data. A combination of RA/SDC (or EU-RAN) may be the optimal solution, as this ensures that the highest level of information security. In practice, a researcher will be located a safe data location in his own country while accessing data located in a safe location in another country. This gives optimal user control as well as ensuring information security through safe environment/facilities. Not only will the researcher be screened and accredited, they may also be monitored. Benefits for the researcher is that he doesn't have to travel abroad in order to work with his data, as would be the case with a SDC solution. RA and an extended solution as an EU-RAN based on RA would also open possibilities, depending on national constraints, for transferring data for research projects requiring combined data. Consequently, researchers will then be monitored by the safe location in their own country instead.

An Information security standard is a prerequisite for this solution. The adoption of such ISO norm would facilitate inclusion in a EU-RAN of RDCs held not only in the premises of NSIs but also in Data Archives, thus contributing to the construction of a single secure infrastructure in the ERA available for all sorts of confidential microdata and national delegation systems. Currently in 2 countries, NSIs have delegated to another institution, for example UKDA and CASD/GENES for secure access for official microdata.

The internationally recognised information security standard ISO27001 presents guidelines for implementing a satisfactory level of information security for RA and SDC solutions, and it is recommendable for microdata access entities in the ERA, mainly NSIs and DAs, to implement it. This will lead to harmonization and transparency with respect to information security, which will lead to a "Circle of Trust" across borders. Without trust between institutions, bilateral agreements for remote access to data across borders will be difficult to achieve.

The main problem today, as discussed in chapter 2.1, is that microdata access is regulated by National Statistical Acts which differ from one another. Some countries allow microdata access to foreign based researchers, while in others foreign access is not mentioned in the law. In the latter case, some interpret this by giving access, while others close their doors for foreigners. If information security and accreditation procedures are regulated / handled on a national level as well, then further insecurity is inflicted when faced with data requests from researchers abroad. The data owner may ask: How can we be certain that our data will be properly secured? Through researcher affiliation with an ISO27001 certified institution (NSI or DA) such uncertainty will be minimized, contributing to a higher level of trust.

Introducing the ISO27001 information security standard

The ISO27001 standard is a generic process- and lifecycle-oriented information security management system (ISMS), which can be used across a variety of businesses. It is considered as a specification of best practice for an ISMS, and contains the following 3 elements:

- Introduction and scope description
- How to implement the ISMS
- In-depth description on management roles and responsibilities

Adapting an ISMS such as the ISO27001 should be a strategic decision depending on the needs, objectives, security requirements and the size and structure of an organisation. Small problems need simple solutions, while the complex ones require a more widespread solution. The supplementary ISO27002 is a code of practice for information security and is meant to be used as an implementation support, presenting an extensive list of specific information security measures to implement.

The measures presented by the ISO27000 family of security standards cover most of the needs for information security related to the business of archiving, handling, processing, creation and dissemination of data.

The ISO27001 information security standard is introduced and explained in the DwB Deliverable D3.3: 'Research data centres and ISO27001 – a guide'[8]. The standard in itself is a generic information security standard. The deliverable focus is on what is relevant for handling and dissemination of micro level research data. It also introduces a specific risk management model called OCTAVE Allegro.

Deliverable D3.3 singles out three components that need to be controlled through appropriate measures in order to secure research data and confidentiality in practical use situations:

- The data (the actual data and derived analytic output)
- The user (employees/data providers or data users/researchers)
- The practical/technical environment (the surroundings, equipment, systems etc)

Current ISO27001 providers

The UK Data Archive (UKDA) is the first academic department in the UK to have achieved ISO 27001 certification. In doing so, the Archive, via the UK Data Service, has been able to provide secure access to data deemed too confidential/sensitive to allow researchers to download. Certification to ISO27001 has been instrumental in giving prospective data depositors of confidential/sensitive data additional confidence that the Archive can provide access to such data in a safe and secure manner.

Ensuring adequate information security measures are in place, reviewed and developed requires much time and devotion among Archive staff at all levels of seniority. ISO27001 certification is considered a major asset to the Archive, particularly as data security continues to increase in importance in the UK data landscape. For example, the ESRC's Big Data Network, including the Administrative Data Research Network, focuses heavily on the implementation of data security measures.

While the process of developing and implementing information security measures would anyway be incorporated into our processes, ISO 27001 provides a formal framework by which these processes can be examined against, the results of which can easily be judged by data depositors.

[8] D3.3 is available here

According to UKDA, achieving certification of ISO 27001 is not a trivial task. Archive staff spent a significant amount of time ensuring operations comply with the elements of ISO 27001, and consider methods by which procedures can be improved. However, the relatively recent organisation of the GOC and ISMG structures, and the streamlining of management for Controlled Documents, has reduced the cost/burden of ensuring compliance and continuous improvement.

It must be emphasised that without the certification, the Archive would not contractually be permitted to hold and provide secure access to, the confidential sources of data that it can. In this sense, the benefits to the Archive accruing from ISO27001 certification far outweigh the costs of implementation. However, there is no real reason why an organisation without receiving certification should not be trusted to do the same. However, certification inspires confidence, increases trust, and promulgates awareness across the organisation. It is the Director's opinion that the cost of implementation and certification is little more in real terms than the cost of implementation alone, and the benefits accrued by additional certification outweighs those additional costs.

More generally, the UK data landscape is increasingly changing and becoming shaped by organisations that maintain 'safe settings' for data, and those that do not. As more sources of confidential data are considered for access, and with an increase in public consultation about the use of personal data, so organisations that can demonstrate competence with handling confidential data find themselves in positions of advantage. Indeed, one of the 'visions' of the UK Data Service is to maintain a status as a 'trusted source of data', and ensuring that we operate 'safe settings' is an important avenue for achieving this status.

However, one issue about ISO 27001 auditing that the Archive has recently addressed is the 'relevance' to providing access to confidential/sensitive data, through the Secure Lab. Previously, audits did not examine closely the procedures in place for managing access to these data, or for managing data, user accounts etc. within the Secure Lab environment. As a result, data depositors who supply the Archive with confidential/sensitive data have visited the Archive to undertake their own audits'. This should not be encouraged due to the duplication of resources. The external ISO 27001 audit should prevent any duplication of effort in governance.

The reason for accrediting to ISO 27001 originally was to ensure the Archive could store and disseminate secure access to confidential/sensitive data. That the original audits did not scrutinise these operations too closely, resulting in further audits from data depositors, implied further resources were used for auditing than were necessary. The Archive should bear the costs of certification to ensure it can hold and disseminate access to these data, but in a way which is efficient given the limited resources of the Archive.

More about the UKDA experience with the ISO27001 can be found in DwB Deliverable D3.3.

Other options to ISO27001

Alternative standards do exist which should also be considered before a judgement on ISO27001 can be made.

1. ISO 16363 Trusted Digital Repositories

A relevant set of controls for the storage and provision of access to confidential data is ISO 16363 Trusted Digital Repositories standard (hereafter TDR). This standard sets out a number of criteria which accredited institutions should adhere to, and include:

- Organisational infrastructure relating to preservation of data
- Digital Object management
- Technical infrastructure and security risk management

The latter of these criteria clearly have considerable overlap with principles of ISO 27001, although perhaps not as extensively as the set of controls included in ISO 27001.

The Archive has already undertaken a test audit against ISO 16363 as part of the EU-funded APARSEN project. At the time of writing this standard has no formal auditing mechanism which means that it is impossible to be certified against this standard. However organisations wishing to apply the controls described by the standard can do so, on an informal basis.

2. Internal Government security standards

Research Data Centres (RDC) providing access to confidential sources of microdata from within UK government departments comply with security standards set by CESG (the Communications-Electronics Security Group of the National Technical Authority for Information Assurance).

All government departments must adhere to these standards and by virtue of location within a government department, an RDC located in one of these departments will necessarily comply with the standards.

Broadly, the standard covers risk management from a technical perspective. Many of the controls contained within the CESG framework are also included in ISO 27001. However, an RDC, such as the HMRC Data lab (which is the RDC based at the UK government department that collects tax revenue) is not required to meet ISO 27001 because it complies with the CESG requirements. This scenario is almost certainly true of other RDCs that are located within government departments.

3. Meeting security controls without formal certification

It is possible that an organisation can meet security requirements (controls), without receiving formal certification. For instance, an organisation could implement an ISMS,

document procedures and demonstrate continuous improvement by conducting internal audits, without receiving certification to one of the standards mentioned above (indeed the Archive originally undertook this, see Section 2).

However, the benefit of undertaking formal certification with the accompanying auditing requirements is the ability to demonstrate continuous improvement through audit compliance and the implementation of recommended audit outcomes. The resulting certification assures current and prospective data depositors who supply the Archive with confidential data for dissemination of access, enabling smoother negotiations and increasing our ability to maintain our status as a 'trusted source of data'. One could argue therefore that implementing a robust ISMS without formal certification to a recognised standard is not actually an alternative given the benefits which certification confers.

In particular, certification to a formal standard (not necessarily ISO 27001 specifically), necessarily requires the development, implementation and review of 'quality' procedures (i.e. procedures that have been formally considered, tested and reviewed as 'fit for purpose'). Indeed, compliance with a formal framework often leads to:

- improvement in work processes
- ability to meet contractual obligation (e.g. with data depositors)
- a publicly understandable statement of quality and reliability.

Shared Information Security standard

ISO27001 or not, a shared information security standard for research data centres must emerge that is fit for its purpose as a new social science infrastructure for Europe.

There are two important requirements for a common approach to Information Security in Research Data Centres:

i) The adoption and implementation of information security risk management should follow a common model, developed through exchange of expertise and a transparent architecture.

ii) The 'circle of trust' concept should be used to ensure that standards in information security in RDCs become neither a race to the bottom, nor gold-plating, but a common core infrastructure standard that is fit for its purpose.

1. *Shared information security standards and interoperability does not mean all RDCs are the same*

   - A diversity of modes of service is healthy.
   - Different users and different data sources require different modes of service.

- In a one-stop-shop of social science data access services, each different mode provides a different service for any ERA user.
- However, diversity that is within the ecosystem is most desirable. Diversity that is incompatible with the ecosystem can be a threat rather than strength.
- The aim is a network of equivalent services which can build and extend organically
- We must ensure that our recommendations will encourage an adoption of a common understanding of the need for and character of information security across organisations.

2. Establish RDC/ build a network of equivalent services

The necessary expertise to build a shared information security standard, and the mechanisms for promulgating it, already exists in the ERA.

- It is a responsibility of established RDCs with an ISO 27001 accreditation and with the confidence of one or more National Statistics Institutions, to form partnerships with other RCDs with the express purpose of assisting the partner RDC service to achieve compliance with the ISO standard.
- The established RDCs should recover the costs of that assistance. These costs represent a very substantial saving compared to external consultancy costs. Further, the investment stays within the ERA and builds the capacity and capability of the social science infrastructure of Europe.

Benchmarking and audit is essential to retain the confidence of data depositors:

- If the adoption of a common approach to information security is achieved, the costs of audit and benchmarking are reduced.
- A coordinating authority, such as CESSDA, should make the necessary arrangements for the spot auditing of the members of the emerging network of RDCs in the ERA.
- Transparency of benchmarking and audit is essential. An annual report on the changing threats, standards, and service delivery should be produced by the relevant coordinating authority.

De-mystification is essential:

- Coordination, partnership assistance, circle of trust recognition, transparency and coordinated benchmarking and audit, should enable new RDCs to be established with necessary background information about time, cost, and service constraints. The establishment of a new RDC should become routine. New RDCs should be built within the Circle of Trust from the outset. It is more efficient to build a new RDC to a common standard than to build a unique RDC and then re-engineer it to fit the standard at some later date.

CHAPTER 3 - THE FUTURE, GOING FORWARD

3.1 Removing Obstacles to Trans-National Access

Some major obstacles to trans-national access require Europe-wide agreement, but others are easier to address, such as standardized accreditation and security. However, the main obstacle to smoother trans-national access concerns penalties to be applied in case of breach, as provisions included in statistical laws have only national validity. How then to pursue a person based in another country? This is a major reason why many NSIs have so far been reluctant to authorize the physical transfer of data across borders.

To solve this problem the following is required:-

- Share a common understanding of legal obstacles and possible solutions
- Adoption of a standardised application form for transnational access.
- Adoption of a shared information security standard for research data centre's
- Propose an architecture for a European Remote Access Network (EuRAN)
- Integrate a transnational accreditation system within a European Service Centre for Official Statistics

The aim is not for a commission led pan European law as it cannot protect confidentiality, as this is a matter for national governments. The desired outcome ought to be achieved using national laws. The best way for data to cross borders is as open data. This solution is linked with supplying open data and / or providing a European Remote Access Network solution, linked to the findings of DwB work package 4. Just because remote access is used, the legal problems haven't gone away. Its low risk, but accreditation / security standards, disclosure and institutional agreements will still be required.

International agreement will have to be found – Linked to Institutional agreements and Concept of Trust

This option relies on institutional agreements between statistical organizations that include undertakings to enforce criminal and regulatory sanctions for breach of statistical confidence occurring within their own borders irrespective of the origin of the data. In effect, the parties to the agreement agree to place their data in each other's' care, and to use, protect, and sanction the use of the other party's data as if they were their own data.

The biggest gains in cross-border collaboration in microdata access are when two or more statistical offices in different countries agree, objectively, that they are partners in each other's central circle of trust. The concept of "circle of trust" is based on the agreement that

each member is accepting the same rules and conditions, confidentiality rules and security requirements.

International agreement is already happening with the Nordic cooperation in the statistical field between Denmark, Norway and Sweden.[9] This cooperation is a joint initiative with bilateral agreements. This has mainly worked due to the link in similarities in legislation and shared practices in accreditation security.

The DwB project has discovered a "circle of trust" can bridge the gap for those countries where the law does not mention anything at all. It is possible to prevent a breach of confidentiality as data will be accessed at least under the same security standards as in their own country to agreed sanctions and penalties for breaches of confidentiality. The concept of trust proposes that a statistical authority should not be inhibited to exchange data where it is assured that the agency at the destination of the data is both willing and able to sanction any breach of confidentiality to a standard that meets or exceeds its own public laws and regulations.

This concept has been demonstrated in WP4 of the DwB project where IAB and the UKDA have moved towards transnational remote access with the installation of an IAB server and contract with the UKDA.

Countries recognize and trust one another's accreditation criteria, so that they can be more confident when assessing how "safe" a foreign researcher applicant is

Proposals for further mid-term steps through harmonisation of the accreditation process and application form, bilateral, multilateral agreements, in the case of multiple accreditations, should be used for future projects for implementation of remote access networks solutions. Statistical organisations need to recognise and trust one another's accreditation criteria (standardised accreditation) so that they can be more confident when assessing how "safe" a foreign researcher applicant is.

Accreditation is fundamentally a human-driven process based on compliance, transparency in the establishment and enforcement of rules and procedures, and the creation of a "circle of trust" in which the researcher is accepted as partner into the official statistical system, albeit temporarily, and is made contractually to share its obligations and responsibilities.

[9] I. Thaulow (DST), Access to Nordic microdata for research purposes, in European Data Access Forum, 24 & 25 March 2015, Luxembourg. See http://www.dwbproject.org/events/edaf2.html

Further improvements from IT developments and the recent progress of secure remote access facilities

WP4 of the DwB Project looked at the feasibility of a European remote access network (EU-RAN). The project looked at a Proof of concept for an EU-RAN with a secured infrastructure based on 3 nodes and a central node with CASD, GESIS and DESTATIS.

The Concept for an EU-RAN would allow researchers to work on datasets located in different countries with other researchers. Remote access involves a great degree of trust-building between the data user and producer, requiring, beyond the IT infrastructure, engagement through a contract as well as methodological and ethical training of users. Remote execution virtually eliminates the reliance on legalities and puts the burden on technological means for protecting confidential data. With remote execution, a researcher never sees confidential data and therefore cannot breach statistical confidentiality through permitted use of the remote execution system. The responsibility of ensuring confidentiality falls on the National Statistical Office (NSO) allowing access to its resources, which must vet research output to ensure that it does not violate the confidentiality laws of its own country.

Work package 11 of the DwB project is developing statistical disclosure control measures that are suitable for checking the outputs of research data centres, for preparing the datasets they hold, and for the training of researchers in disclosure control measures. The obligation to complete training and to submit outputs for checking can form part of an accreditation decision and standard.

This concept works when linked to a common adoption of a security and accreditation standard as demonstrated in WP4 which examined the issue of linking research data centres and models for operating data centres to serve transnational access.

Integrate a transnational accreditation system within a European Service Centre for Official Statistics

In our view, what is needed in Europe is a centralised system where researchers can discover and then apply for data from different NSIs all in one place, following the same procedures and using the same forms. Such a central transnational accreditation data access system for official statistics would save time and effort for researchers, would reduce the burden on NSIs, and would facilitate and strengthen comparative analysis internationally.

Working with data is not the beginning of a research project. Before accessing data, researchers need a variety of preliminary information. First, they need to know which data sources are available, and from where. Then they need detailed information about the datasets and the variables (preferably standardised for all datasets, collected amongst all

countries). On the basis of these sources of information, the researchers can decide whether their project is achievable with the existing data. If so, information about the conditions of access to data are needed, including information about the accreditation process, contracts, legal restrictions, statistical disclosure control of results and other research outputs, and technical access modalities.

The discovery and exploitation of available data is very time consuming, in so far as it implies understanding different application and accreditation procedures. An integrated portal service for data discovery across a multiplicity of data providers and metadata systems is currently being developed within DwB WP5, WP8 and 12. Such a service would substantially reduce the effort related to data discovery and exploitation, by providing the researcher with consistent information and support as well as documentation and metadata in standardized formats. A group of researchers working on the same project who would live in different Member States ought to be able to consistently access high quality information.

It would be desirable for the Eu-RAN, to have a facility for managing online accreditation applications. This would provide convenience to researchers submitting applications to access the data, and for data producers who grant permission to access the data. The ESC-OS would be a research infrastructure institution aligned with the European Statistical System and the CESSDA network of European Data Archives, and would promote and "broker" the scientific use of European official statistical microdata by providing enhanced metadata (in English) and a central data discovery and access facility. Through the ESC-OS, researchers would be able to locate relevant data for their projects, and then order the data from the respective NSIs or other government agencies. A transnational accreditation and access system built into a future ESC-OS would be an important step forward for the international community of social scientists who conduct comparative research, or who focus their work within countries where they do not reside.

3.2 Summary

A European research infrastructure can only work, if data providing institutions (NSIs, data archives, etc.) are working together. DwB has started to create an organisational network for Europe based on understanding, trust, harmonised workflows, agreements, and contracts. Consequently, the concept of trust around institutional agreements and standardised accreditation should be encouraged to minimise the risks around disclosures and encourage open data for researchers. Further improvements from IT developments and the recent progress of secure remote access facilities are linked to a common adoption of a security standard and accreditation. This was established in WP4 of the DwB project which examined the issues of linking research data centres and models for operating data centres to serve transnational access, within a European Remotes Access Network.

The internationally recognized information security standard ISO27001 presents guidelines for implementing a satisfactory level of information security for RA and SDC solutions, and it is recommendable for microdata access entities in the ERA, mainly NSIs and DAs, to implement it. This will lead to harmonization and transparency with respect to information security, which will lead to a "Circle of Trust" across borders. Without trust between institutions, bilateral agreements for remote access to data across borders will be difficult to achieve. If data confidentiality is convincingly secured, trust will be built both nationally and internationally.

Over the last four years Work Package 3 has seen many changes which are encouraging around the concept of removing the barriers for researchers, to make use of data across transnational borders. From various consultations and the proof of concept in WP4 there is a continuous trend of improvement in the legal frameworks through remote access. Statistical providers understand the need to move away from multiple accreditations and the need for harmonisation through a standardised accreditation. The collaborative working with the research data centres in WP4 has highlighted the need for an international security standard, through a notion of the circle of trust concept. Consequently, many of the recommendations and tools created within WP3, once implemented, will remove and open the barriers for researchers to official data across international borders.

REFERENCES

Linden, Mikael, Tommi Nyrönen and Ilkka Lappalainen. "Resource Entitlement Management System." Trans-European Research and Education Networking Association. 24 April 2013. pp.12.

Dale A. (2003). Research access to microdata: an attempt to provide a context", IASSIST Quarterly, pp. 16-19.

OECD (2012). Microdata Exchange and the Challenges of Open Data and Transparency, Report of the Expert Group for International Collaboration on Microdata Access, chaired by P. Jackson, n. STD/CSTAT/MICRO(2012)14, 3 December.

Tubaro P., M. Cros, B. Kleiner, R. Silberman (2013). Accreditation for trans-national research access to official microdata in Europe. Proceedings of the NTTS Conference, 5-7 March, Brussels.

Tubaro P., M. Cros, R. Silberman (2012). Access to official data and researcher accreditation in Europe: Existing barriers and a way forward. IASSIST Quarterly, Spring, p. 22.

Tubaro P., R. Silberman, M. Cros, A. Cornilleau, V. Kvalheim, D. Kiberg, P. Farago (2009). Audit of access mechanisms and official statistics in the European Research Area, Technical report D10.1 for CESSDA PPP.

UNECE (2007a). Managing Statistical Confidentiality and Microdata. Principles and Guidelines of Good Practice, handbook prepared at the request of the Conference of European Statisticians (CES) – Task Force on Confidentiality and Microdata chaired by D. Trewin, New York and Geneva, United Nations Publications.

UNECE (2007b). Register-based Statistics in the Nordic Countries. Review of Best Practices with focus on population and social statistics, New York and Geneva, United Nations Publications.