Accreditation for transnational research access to official micro-data in Europe

Paola Tubaro[1], Marie Cros[2], Brian Kleiner[3], Roxane Silberman[4]

[1]University of Greenwich, London, and CNRS, Paris, e-mail: p.tubaro@gre.ac.uk
[2] Université Lille I, e-mail: marie.cros@univ-lille1.fr
[3]Brian Kleiner, FORS (Swiss Centre of Expertise in the Social Sciences), e-mail: Brian.Kleiner@fors.unil.ch
[4] CNRS - Réseau Quetelet, e-mail: roxane.silberman@ens.fr

Abstract

Accreditation is a central element of the framework for research access to micro-data that currently is understood to be a barrier for transnational access. To better understand the nature and causes of the problem, and to devise potential solutions, we have mapped current arrangements across European countries. We identify similarities and differences as well as areas for improvement.

Our key results are encouraging: almost all European countries do provide research access to their micro-data, and most of them allow non-national European researchers to access their data, though under varying conditions. However, some obstacles remain, and some of them require negotiation and coordination at policy-making levels.

To overcome existing barriers, we propose some potential options for the future and concrete steps towards improvement, which if explored from now on, could substantially improve access while still ensuring safe and lawful conditions, and reducing the administrative financial burden of data providers.

Keywords: cross-border access, data archives, National Statistical Institutes.

1. Introduction

Accreditation is a central element of the framework for access to official statistical micro-data for scientific purposes, as authorised by European legislation and the statistical laws of a growing number of countries. It contributes to enabling safe research access to official data, and it is part of the risk management of National Statistical Institutes (NSIs). It includes defining the conditions under which a researcher can be considered a "fit and

proper" person, and thus be allowed to access data; sometimes NSIs regard it as a way to ensure that a researcher is comparable to official statistics staff members and subject to the same rules, codes of conduct, and penalties in case of breach. Accreditation is a complex procedure with three major components: the definition of *eligibility criteria* (who is a researcher, what is research, what is a research project); the set-up of *application procedures* (how to request access, what forms to fill out, whose signatures to include); and the design of *rules for decision-making* (who decides, and on what basis). The relative weight that is placed on each of these components often depends on the degree of disclosure risk, typically distinguishing between data with high anonymisation allowing wide dissemination, data with an intermediate degree of anonymisation (with removal of direct identifiers and moderate use of disclosure control tools, or "de facto" anonymised), and highly detailed (confidential) micro-data.

Accreditation is currently understood to be a barrier to access across borders, even within the European Union and neighbouring countries, which share fundamental data protection and statistical legal principles. While there is growing interest in the scientific community for comparative or Europe-wide studies that require access to datasets from different countries, it remains very problematic to obtain them. The different procedures and practices in researcher accreditation that NSIs maintain and recognize produce inconsistencies, often resulting in increased burden for researchers who, for example, often have to submit several applications for data to be used in one single project, each with differing forms to complete, evidence to produce, and timing. Nor is the task of NSIs easy, owing to the greater difficulty of assessing whether researchers from a foreign country (and their affiliated institution) can be considered safe, compared to nation-based applicants. Of course, the higher the disclosure risk of the datasets for which access is requested, the more crucial it is to ensure that all conditions for safe access are met. Thus, trans-border applications often entail either an increased administrative burden for NSIs, if they undertake a more thorough examination of eligibility criteria, or a negative default position with respect to granting access to researchers across borders.

The Data without Boundaries (DwB) project, funded by the European Commission under its 7th Framework Programme for 2011-15, aims to promote equal and easy access to official micro-data in European countries. One part of it addresses accreditation, with particular emphasis on transnational accreditation and highly detailed micro-data. The objective of our work is to better understand the nature and causes of existing barriers, mapping the current arrangements – including eligibility criteria, application procedures and rules for decision-making – in the different European countries. We aimed to detect similarities and differences, as well as to identify any best practices – enabling enhanced access under relatively simple and straightforward conditions for data users, while still protecting the confidentiality of statistical units. On this basis, we are now proposing potential solutions, to be discussed with all stakeholders (particularly NSIs and representatives of the research community), in an effort to devise solutions that if adopted might improve aspects of the accreditation process.

In the remainder of this paper we present the key results of our overview study of accreditation conditions and procedures in the different European countries, outline ongoing tendencies and efforts to make improvements, and propose some potential directions for the future, as a basis for discussion with NSIs and other stakeholders.

2. Accreditation to official microdata in Europe: an overview

As a first step in our study of cross-border researcher accreditation, we collected information on current arrangements in Europe. To ensure comparability across countries, we focused on NSIs only, leaving aside other public-sector data producers (statistical administrations such as IAB in Germany, or central banks such as the Bank of Italy), and we considered NSI data at all levels of anonymisation, not limiting our search to highly detailed micro-data. We retrieved most of the information from NSIs' websites but also from some existing sources (Tubaro et al. 2009), and obtained some information directly from representatives of Eastern European NSIs at a dedicated workshop organized in Bucharest, Romania, in January 2012. Our key results are on *national-level* accreditation procedures, for access to *national-level* datasets by researchers based in the same European country or in another one (what we call "transnational access", strictly speaking). To have a broader picture, we compare below these results to current conditions for accreditation and access to *European-level* micro-datasets at Eurostat.

2.1 National-level accreditation: policies and practices, progress and remaining barriers

Our key results can be summarised as follows:
- o *All countries have provisions to enable researcher access, but practical modalities and conditions vary widely.*
  Reassuringly, we have found that all European countries allow some form of researcher access to their microdata. The open data movement, pressure on NSIs to maximise the value of their data collections, and progress in IT have enabled major steps forward in this respect in a large number of countries. However, differences remain. Available data are mostly social surveys, while business data are less widely accessible, and register data are most commonly found in Northern Europe, where they traditionally form the basis of official statistics. Most countries release de facto anonymised data, usually in the form of Scientific Use Files (SUFs); comparatively, both highly anonymised and highly detailed datasets are more rarely found, though their availability is growing, driven by both demand from data users and IT developments. In some cases, research institutions and data archives (such as NSD in Norway, Réseau Quetelet in France, and the UK Data Archive) participate in the dissemination process and share part of the administrative and financial burden of assessing and managing applications, and subsequently providing data services.
- o *Basic conditions (in particular, eligibility criteria, contents of applications and user contracts) are common, but obstacles remain to a more homogeneous process.*
  Interestingly, key basic principles are held in common: for example, almost all NSIs, for almost all types of data, will check the non-profit research purpose of any applicant; and a large number require affiliation with a higher education or research institution. The strictness of the latter requirement varies, though: some

countries assess institutional suitability on a case-by-case basis, while others such as Norway and the Netherlands have an approved list of institutions, with some differences in procedure for applicants from non-listed (for example, newly-created) institutions.

In practice, most NSIs require a written application: a dedicated form in some countries (such as France, Germany, Italy, UK) and a simple letter in others (e.g., Estonia, Poland, Slovakia, Spain), often with a description of a research project, datasets and variables to be used, where more details are usually demanded when access to more detailed datasets is requested. Almost all countries also require signing of a contract (or undertaking or user licence) once access has been granted, the only few exceptions concerning highly anonymised data files that are freely downloadable from the web, such as Public Use Files in France and "Campus Files" for teaching in Germany. All in all, it is encouraging to notice that cross-country differences usually concern actual practices and details of procedures rather than general guiding principles.

o *There is widespread openness to granting transnational access, but with limitations or with more restrictive conditions, relative to national researchers.*

We were pleased to discover that most NSIs allow researchers from other European countries to access their data. However, there are differences in the way this principle is put into practice. While access is often granted under the same conditions as national researchers (UK, for all types of data), sometimes foreigners have to undergo some additional scrutiny to prove their eligibility (in France, for example, for SUFs). Heavier restrictions are in place in some countries, for instance when foreign data users are prevented from receiving the data at home, having to visit onsite data centres instead (Germany, for all types of data except Campus files), or have to register with a host national research institution (Denmark, for detailed micro-data).

o *Some major obstacles to transnational access require Europe-wide agreement, but others are easier to address.*

The main obstacle to smoother transnational access concerns penalties to be applied in case of breach, as provisions included in statistical laws have only national validity: how, then to sue 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 (especially detailed micro-data) across borders.

To entirely solve this problem, some form of international agreement will have to be found, where 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. In addition, further improvements are coming from IT developments and the recent progress of secure remote access facilities, involving no actual transfer of data. Norway, for example, used to provide foreigners with fully and de facto anonymised data, and is now piloting a new remote access system (RAIRD, Remote Access Infrastructure for Register Data) that aims to enhance access to more detailed micro-data for the whole research community, both national and international.

There are also some subtle issues that hinder transnational access, in particular differences in terminology and definitions across countries: for example, the term

"Public Use Files" does not always refer to datasets that are available to the general public through the web, and may be subject to some restrictions; conversely, secure "remote access" computing facilities are sometimes used to offer access to data that are *not* highly detailed, and would elsewhere be distributed as SUFs. It is clear that misunderstandings make transnational access practically more difficult, though legally authorised. Again, however, clarification requires some degree of agreement and at the Europe-wide level.

Our study also shows a widespread lack (or incompleteness) of online information about accessible data, conditions for access, criteria and proceedings for accreditation. Transnational access is rarely explicitly mentioned, even when the countries do allow it, and availability of English-language translations is uneven. These problems constitute a major practical obstacle for users, but can be easily and cheaply solved – while the issues mentioned above require some degree of negotiation at higher policy levels.

Overall, these results suggest overall openness of NSIs to the needs of researchers, a high degree of commonality in guiding principles and general conditions, and significant improvements in recent times. However, differences and obstacles remain; though mostly at the practical level, they currently constitute major barriers to transnational access and require concerted solutions.

2.2 European-level accreditation at Eurostat

The existence of barriers and the need for concertation is flagrant when it comes to the European-level process for granting research access to the European datasets held by Eurostat. The current system shares its main guiding principles and implementation details, with the national-level access and accreditation procedures outlined above. This is a natural result of the design of European institutions, as common regulations are the result of discussions and agreements among Member States (here, represented by their NSIs). In particular the distribution of European SUFs requires institutional affiliation backing based on a list of approved higher education and research institutions provided by Member States, and held and managed at Eurostat. Similarly, negotiations among countries will drive the future regulation and guidelines that are currently under discussion, with the aim of facilitating research access to highly detailed European-level micro-data. At the moment, discussions revolve around the possible requirement of an accreditation procedure for each research institution prior to individual researcher accreditation. Such a procedure may result in increased burden for all parties, but may ultimately be imposed for access to all types of data, including highly detailed microdata. Another controversial issue is the possibility to delegate access to de facto anonymised data (SUFs and their equivalent) to research institutions and data archives. While some countries have experience in delegation at national level, others are new to such a process and more reluctant to extend it. These are very sensitive questions that, as mentioned above, affect the risk management of NSIs and have to comply with the strict requirements of the statistical and data protection laws of all countries. Because European regulations require consent of all Members States to be approved, a researcher-friendly regulatory change at this level may be a lengthy process.

3.   Ultimate goals and current restraints

Looking towards the future and possible solutions to improve this situation, what would be the optimal state from the perspectives of researchers and of NSIs? Where do interests converge and where are they at odds? In the end, what would be the ideal solution, given these various perspectives?

From the researcher point of view, official micro-data first must be discoverable. This means that there should be sufficient information online to direct users towards relevant datasets, in a comprehensible language, and (most importantly for transnational access) with English translations. Once relevant data are discovered, researchers would like to be able to place a request for access: hence, application forms should be easy to download, complete, and submit, again in a familiar language. Importantly, researchers do not want to have to complete multiple application forms for different datasets or for datasets in different countries. Finally, once having submitted a request for data, researchers expect a prompt response and quick and easy access to data, after having signed a user license agreement with the data provider. Ideally, the researcher would like to have one single form for any request, possibly usable for multiple requests, and one single interlocutor.

From the point of view of the providers of official micro-data, access for research purposes is a secondary goal, after production of information for governments and protection of the confidentiality of the individuals who provide them with information. It is essential for NSIs to fully meet the legal requirements and to maintain the trust of policymakers and the public. Release of data requires strong legal expertise and experience as well as substantial material resources: even just for accreditation, NSIs have to implement multiple checks and controls, all the more so when their policy requires thorough investigation of researcher or institution eligibility. They have to adequately evaluate requests for data, process licenses, prepare datasets, address questions from users, and make metadata available in English following international documentation standards. Lack of resources may prompt some of them to make some of their data unavailable to researchers, even when access would be lawful.

To satisfy all these needs simultaneously, one would need countries to 1) agree on shared eligibility criteria, 2) agree to adopt a standard application form, 3) delegate decision-making power to a third party, possibly a central body, and 4) delegate management of the procedure too. In principle, a unified system built along these lines may both meet researchers' demand for better information, rationalization of procedures and simplification, and NSIs' need to stick to high standards of data protection while reducing some of the administrative and financial burden associated with data release.

Such a solution, however, faces a number of obstacles. They can be divided into two main categories: first, overcoming existing differences in eligibility criteria and requirements to establish to obtain accreditation, and second, acceptance of delegation to a third party for decision-making and management of applications.

The results of our study, described above, suggest that differences in criteria and requirements are in fact easier to surmount than it might seem at first sight, in that they mostly concern practical matters and specific implementation of regulations, but not the general principles and guiding lines. These similarities may be taken as a basis for an

agreement among countries, at least on some key points. However, more extensive negotiation is likely to be necessary on the issue of whether research institutions need some form of accreditation together with individual researchers, and even the very definition of a research institution – whether it must be based on some form of national recognition, or whether it must meet some given requirements (such as non-profit purpose). One solution for transnational access to national data would be to start from less ambitious agreements between clusters of countries that share a similar approach to this issue, and can therefore more easily come to an agreement. This could be a first step towards implementing larger-scale agreements, and may improve access at least in some countries. A common, standard application form may easily be derived once these key principles are agreed upon, and translations in different languages could be provided. This would be a significant improvement for researchers, who would have better access to data held nationally, which often include administrative data and other datasets that are not available at European level at Eurostat. Thus, such clusters would contribute to creating a more researcher-friendly environment parallel to developments at Eurostat.

Delegation to a third party for decision-making and application handling is a different and more sensitive choice, particularly for highly detailed micro-data. As mentioned above, encouraging signs come from the experience of countries where NSIs have delegated some of their functions to data archives. Although most of them have done so only for highly anonymised and sometimes de facto anonymised data (for example Ireland, the Netherlands), there are experiences at least of cooperation, if not full delegation, for highly detailed data (for example France, Norway). However, greater efforts towards trust-building will be necessary before this model can be generalized to the whole of Europe – or at least to smaller clusters of volunteer countries.

Overall, principled support for supporting access to official micro-data for scientific purposes quickly runs up against real-world constraints and considerations. Given the challenges of these constraints, it is most likely that changes towards easier transnational access to official micro-data will be incremental. NSIs and other providers of official micro-data will need time to adjust their policies, procedures, and systems, and will need assistance in overcoming resource limitations for a service that is not their primary concern. So what can be done in the short- and long-term to ease access to official micro-data for research? The following sections describe and propose a set of potential workable solutions focused on accreditation, some currently under discussion within the DwB project.

4.   Moving forward: potential solutions and work in progress

4.1   Short and medium-term improvements

A first, and relatively low-cost, improvement could be obtained if NSIs would ensure availability of complete English translations of their websites, particularly the pages dedicated to data access; use visual clues and an appropriate structure of hyperlinks and sidebars to help users locate access-related information; verify clarity and completeness of information on both general criteria, any special conditions (particularly for transnational access), and how to apply; and systematically provide application forms,

ideally with English translations. These are simple steps that could, however, remove many practical obstacles and greatly facilitate the discovery task of researchers interested in using data from different countries.

Within DwB, we will contribute to this process by making available the information on accreditation criteria and procedures that we have collected (and that we are using for this study). We aim to develop a searchable tool to be made available online, providing data users with comparable cross-country information on accreditation. It will be a repository of web pages, each describing accreditation in one NSI and linking to relevant information and forms on the website of this NSI, and all also linking to a data base of official statistical surveys available to researchers, which is being compiled in another part of DwB.However, maintenance of the repository after DwB comes to an end will require some degree of cooperation and perhaps of delegation – as the task could be co-handled by NSIs and data archives, or even taken over by data archives. A potential benefit of a common repository is that, though only providing basic information and links to NSIs' own websites (or those of archives), it will contribute to making countries more aware of similarities and differences among them. On this basis, they may be more willing to start talks on how to homogenize and rationalize the system, for instance through common forms, or some procedural simplification for researchers wishing to access data of the same type from different countries. Again, this could occur within smaller clusters of countries first, piloting the model for others that may want to follow suit at a later stage. Acceptance in one or few countries may indeed accelerate acceptance in other countries too. Further extensions and development of this tool, however, will require higher-level policy negotiation, presumably in the long run.

4.2 Long-term possibilities regarding European accreditation processes

While a stepwise approach to improving accreditation to official micro-data is most realistic, a long-term sustainable solution will eventually be required to ensure sufficient stability and coverage across Europe. The current situation, as described and discussed above, suggests that a long-term solution for transnational access may require a higher level of international standardization and centralization of accreditation issues than has until now been the case. As we will argue below, moves towards greater standardization and centralization on the part of data providers will likely lead to a significant reduction of burden in responding to the needs of researchers.

- o *One option would involve establishing a centralized database of accredited users and institutions within Europe.*
  The payoff of such a solution would be considerable: rather than having to evaluate the eligibility of individual researchers and their institutions (foreign or otherwise) for each data request, NSIs could simply check whether these are authorized in the database, and then concentrate solely on the safety of the specific proposed project. However, such a database may not significantly shorten the accreditation process if specific elements (like a research project) cannot be transferred from one access request to the other; the database will require require continuous updates and maintenance to be really useful, taking into account

researchers' mobility; and its setup is likely to require a significant up-front investment.

o *There may be a "researcher passport", a recognized international document carried by vetted researchers to prove eligibility for access to micro-data.*
Once established by a central commission according to international standards (approved by NSIs), the passport would allow researchers to apply for official micro-data, by-passing identify checks and controls. It might be that the passport would rely, at least partly, on this central database and would be filled in with information about accessed data, their provenance, and so on, much in the way that a passport is stamped by different countries that have been visited. Again, updates will be needed regularly and frequently, and the extent to which such a passport really simplifies the process will need to be carefully evaluated.

o *Another idea is to build a central database of information about research institutions, by housing all of the institutions that have been evaluated and accredited across countries, according to some approved international standard.*
The advantage for data providers evaluating requests for transnational access is that they could simply rely on this database for institutional information, and focus their assessment on the specificity of each application (the research project for example). Such a solution, however, requires an agreement about the types of credentials institutions must provide, and about the very need for an accreditation of institutions – which may in itself be burdensome. In all cases, the database should not duplicate the work already done at Eurostat, and might in fact rely on it as a starting point.

If such a central service is established, it will be paramount to properly design its governance system. It will be necessary to ensure that the database is up-to-date and implemented lawfully and strictly in accordance with the accepted criteria. For this reason, it should be run by an internationally recognized and trusted organization or infrastructure that is responsible for accreditation of researchers and institutions. Many decisions will have to be made in this respect: in particular, this responsibility could be delegated to a supra-national institution such as Eurostat, shared by NSIs, or co-handled by NSIs with data archives (at least those that have experience with official statistical data). The latter solution would have the advantage of reducing the administrative and financial burden.

The DwB project is considering the creation of a *European Service Centre for Official Statistics (ESC-OS)* (see Mack et al. 2012). While an ESC-OS primarily aims to provide an online portal of metadata for discovering official micro-data, it might also serve as a central database for accreditation, possibly linked to a European remote access network. The Centre could conceivably assume responsibility for processing requests for data (using a standard application form), and then dispatch completed forms for review by data providers. Its governance will be shared between NSIs (within the European Statistical System, ESS) and data archives (CESSDA, *Council of European Social Science Data Archives*). In this way, much of the burden on data providers would be reduced, and researchers would have the benefit of discovering and requesting data all in one place, using a single application form. Over time, once trust has been built, NSIs might delegate more decisional responsibility to the ESC-OS, leading to greater homogenization and rationalization, though this may only be decided at a later stage.

To discuss these options and concrete steps to take, DwB is now in the process of preparing consultations with individual NSIs to gain insight into their openness to any such solutions. The ultimate accreditation model that we would like to put forward must be consistent with national legal frameworks, more researcher-friendly, and without being overly burdensome for data providers. NSIs' views are important to help us strike a balance between these potentially conflicting needs.

Conclusions

Overall, our study suggests that a standard European model for accreditation and access to official statistics is a realistic possibility, though adoption requires time. The upcoming consultations will discuss potential solutions, some of which have been outlined above. We aim to refine them in light of feedback from NSIs, in order to propose a set of recommendations that may really make a difference in the years to come. It will be necessary, in particular, to establish the extent to which a "passport" would actually facilitate transnational access to official micro-data, and how an ESC-OS could be helpful for facilitating researcher accreditation across borders. Many of the answers to these questions will depend on the degree to which NSIs would be open to delegating part of the accreditation process to such a body, and on how its governance structure is designed. These are some of the key issues we aim to clarify with stakeholders in the round of consultations to come.

References

Mack A., Wolf C., Esteve A., Silberman R. (2012) Report on concept for and components of European Service Centre for official statistics. Available at: http://www.dwbproject.org/export/sites/default/about/public_deliveraples/d5_1_european_service_centre_report.pdf

Tubaro P., Silberman R., Cros M., Cornilleau A., Kvalheim V., Kiberg D., Farago P. (2009). Audit of access mechanisms and official statistics in the European Research Area, Technical report D10.1 for CESSDA PPP. Available at: http://www.cessda.org/project/doc/D10.1_Audit_of_access_mechanisms_and_official_statistics.pdf