HAL
open science

Access to Official Data and Researcher Accreditation in Europe: Existing Barriers and a Way Forward

Paola Tubaro, Marie Cros, Roxane Silberman

To cite this version:

Paola Tubaro, Marie Cros, Roxane Silberman. Access to Official Data and Researcher Accreditation in Europe: Existing Barriers and a Way Forward. IASSIST Quarterly, 2012, Spring 2012. hal-01371296

HAL Id: hal-01371296

https://hal.science/hal-01371296v1

Submitted on 25 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Access to Official Data and Researcher Accreditation in Europe:

Existing barriers and a way forward

by Paola Tubaro, Marie Cros and Roxane Silberman [1]

CNRS

Abstract

In Europe, national legal frameworks frequently enable research access to official statistical data, also including detailed microdata, but cross-country research remains difficult. Accreditation is a central element of the framework for access to data that currently is understood to be a barrier especially for trans-national 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 major gaps and inconsistencies across countries, and we single out best practices and new, example-setting solutions.

Overall, our key results are encouraging: almost all European countries do provide research access to their microdata, and most of them allow non-national European researchers to access their data, though under varying conditions. Some of the gaps that we have identified are relatively easy to fill, notably a widespread lack of online information, and unsystematic translation into other languages.

A small set of issues, however, will require negotiation and coordination at higher, policy-making levels: the controversial need for institutional accreditation, homogeneization of terminology, and the possibility to introduce special provisions to facilitate trans-national access. Some of these issues are under discussion today and some new solutions are being tested or piloted, so that substantial improvements can be expected in the future..

Keywords: Research access to data, researcher accreditation, trans-national access, highly detailed microdata, European Research Area.

Introduction

Today's national legal frameworks frequently include provisions that facilitate researchers' access to data produced by official statistical systems. However, existing solutions are mostly country-specific and data do not circulate easily across borders, even within the European Research Area (ERA) where regulations are very similar and a common framework on data protection applies. Difficulties are particularly acute for access to confidential (or highly detailed, as they are sometimes called) microdata. Despite progress in individual countries as well as at Eurostat level, trans-border access to country-level official microdata is still patchy and especially difficult for highly detailed microdata, thereby strongly penalizing comparative research and research on Europe as a whole.

Accreditation is a central element of the framework for access to data across borders that currently is understood to be a barrier to trans-national access. Indeed, National Statistical Institutes (NSIs) and other producers of official statistics maintain and recognize different procedures and practices in researcher accreditation, resulting in inequalities among researchers located in different countries in the ERA, a great deal of red tape, and a negative default position with respect to granting access across borders.

More precisely, accreditation can be defined as the process defining the conditions under which a researcher

willing to use official data can be considered a "fit and proper" person. In the eyes of NSIs, it would mean comparability to official statistics staff members and subjection to the same rules and penalties. In practice, accreditation involves three main steps:

- Defining eligibility criteria: who is a researcher, what is research, what is a research project;
- Establishing application procedures: how to request access, what documentation and evidence to provide;
- Setting up a service level, including: designing rules for decision-making (who approves applications, on what basis), managing and monitoring the process, ensuring good governance and transparency.

Answers to these questions contribute to NSIs' risk management framework, defining the scope for safe research access to official microdata. The problem for researchers and data users is that these answers may differ across countries, depending not only on national legislative frameworks, but also on the internal policies and established practices of each institution. Even within the same country or institution, there are discrepancies depending on type of data, mode of access, or status of the applicant. There is also some degree of variation due to fees that researchers sometimes have to pay for data provision, whether in the form of bespoke files or of secure access through high-tech facilities. Fees do not concern the accreditation process strictly speaking, but depend on the actual costs of services provided and are often set independently of the accreditation-granting authority. Be that as it may, ambiguities and incongruities arise especially in less clearly-defined cases, particularly with foreign researchers and joint (typically, cross-border) projects.

To better understand the nature and causes of these barriers and inconsistencies, and to devise potential solutions, we have set out to map the current arrangements in the different European countries - including eligibility criteria, application procedures, and organisation of the service. We endeavour to detect patterns of similarities across countries, and to identify existing best practices of how to enhance access under relatively simple and straightforward conditions for data users, while still protecting the confidentiality of statistical units. On this basis, we discuss possible approaches for the future, along two main lines. Firstly, we identify a set of simple and small-scale solutions, that may be easily transposed to a wide range of countries, and whose widespread adoption may lead to small, but tangible improvements that may make a difference for users even in the short run. Secondly, in a long-run perspective, we discuss the extent to which a future common standard for accreditation may be considered as a realistic possibility, though perhaps a distant one, and we outline open questions and issues that require negotiation at policy-making level.

Our work on accreditation is part of the Data without Boundaries (DwB) project, funded by the European Commission under its 7th Framework Programme for 2011-15, and aiming to support equal and easy access to official microdata for the ERA. It aims to map the current situation, identify and promote best practices, devise and pilot new solutions for remaining problems. Focus is on trans-national access and on highly detailed microdata.

The remainder of this paper outlines how we have undertaken this study (section 2), describes the major results we have obtained so far (section 3), and indicates directions for future development (section 4)

Methods

To answer these questions, we have started with a discovery phase aiming to collect information on current researcher accreditation arrangements in the ERA (including both the EU and the EEA countries). We have retrieved most of the information from public domain and secondary sources (particularly NSIs' websites) and existing literature (particularly Tubaro et al. 2009; UNECE 2007). We have also obtained primary data directly from representatives of Eastern European NSIs, at a dedicated workshop we organized in Bucharest, Romania, on 23rd January 2012. We have organised the information into a working document (spreadsheet) for internal use to analyze results and to identify common approaches, existing workable solutions, and areas for improvement. In a subsequent consolidation and analysis phase, we have cross-checked this information for completeness, and have identified and revealed the key messages, particularly a list of best practices.

The results presented below are the outcome of both our discovery and consolidation phases. To ensure comparability across countries, we focus on NSIs only, leaving aside other public-sector data producers (such as IAB in Germany, or the Bank of Italy), and we consider NSI data at all levels of anonymisation, not limiting our analysis to confidential data. We emphasize national rather than European data, whose access is managed by Eurostat and follows a specific set of rules and procedures. We focus on practices and procedures rather than legal principles strictly speaking, which are being investigated by another team in DwB. Practices are often part of the "tacit" knowledge of NSIs and are seldom shared or openly discussed, but have potentially strong and concrete effects, that a comparative study of legislative frameworks alone would be unable to bring to the surface. In general, laws are very similar across European countries and are often rather vague: sometimes entirely silent on matters of access to data for scientific purposes, sometimes authorizing access explicitly but without going into the details of who is a researcher, what is research, and how to establish that eligibility conditions are met. Regarding highly detailed data, laws often simply state that security must be ensured. Therefore, interpretation and practical implementation are even more crucial elements than legal rules themselves, in affecting actual conditions of access.

To reveal similarities beyond apparent discrepancies, we strive to use a common terminology here, despite national-level variations in word choice, and nuances in meaning across countries. For example, we broadly distinguish between highly detailed (or confidential) and less detailed data based on disclosure risk only, disregarding the fact that levels of anonymisation and disclosure control may differ across countries, and that modes of access to data with similar disclosure risk may also diverge widely.

Our third and final phase involves dissemination, and is still in the making. We aim to develop the working document prepared so far into a searchable tool to be made available online, to facilitate data users' search for information on accreditation in a comparable manner across countries. A repository of web pages will be built, each describing one NSI, and all linking to a data base of official statistical surveys available to researchers, that is being compiled in another part of DwB. Although this tool is still in preparation, and its technical characteristics have yet to be finalized, we outline in the conclusions how it can contribute to making the results of our study more actionable, and to improving access by making accreditation conditions more easily intelligible across borders.

Results

On this basis, we have obtained a global picture of accreditation procedures and practices across Europe, and we have identified similarities and differences across countries. We present them in the order outlined above - eligibility, application procedures, and service - and we conclude by examining in greater detail the specificity and additional problems that arise for trans-national access to data.

Eligibility

A first key question is how to define a researcher - who are the persons who, by law, can be entitled to have access to datafiles that may not be released to the general public. Interestingly, European countries' answers to this question reveal a great deal of commonality. The legal framework in most countries disallows release of data for commercial use and therefore, requires that all requests to access data are for research/study purposes (Figure 1). Affiliation to a research or higher education institution is sometimes considered as evidence of non-profit research purposes and, especially for highly detailed data, it often also acts as an additional safeguard for the data provider. Indeed some NSIs require data users to be employees of a research institution, so as to involve the responsibility of the institution (and to be able to sue it in case of breach of their terms of use, particularly confidentiality rules). In some countries (Germany) employment at a public research institution ensures subjection to the same codes of conduct as official statistics staff, and is therefore considered as a stronger guarantee against possible misconduct. The track record of the researcher (in terms of previous experience with microdata, publications etc.) is only required in some countries for access to highly detailed data (the UK's "Approved researcher" scheme for example), often allowing alternatives: for example students, who by definition have no track record, need instead some formal backing by their supervisors.

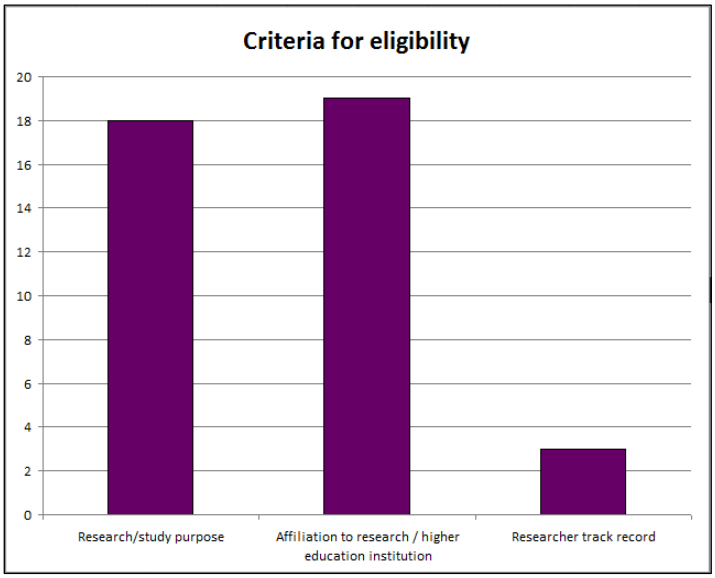

Figure 1: Eligibility criteria for research access in European countries (N = 28), absolute frequency. A country may have more than one (depending on type of data and status of users).

While requirement of a research or study purpose is widely shared, the need for institutional support is much more controversial. The reason is that it is difficult to establish which institutions are eligible, all the more so as there are a growing number of ambiguous cases: public-private partnerships, analyses undertaken by research departments of non-research bodies (OECD for example), multi-institutional research consortia of limited duration, think tanks. Another difficulty concerns the relationship between the researcher and the institution, often

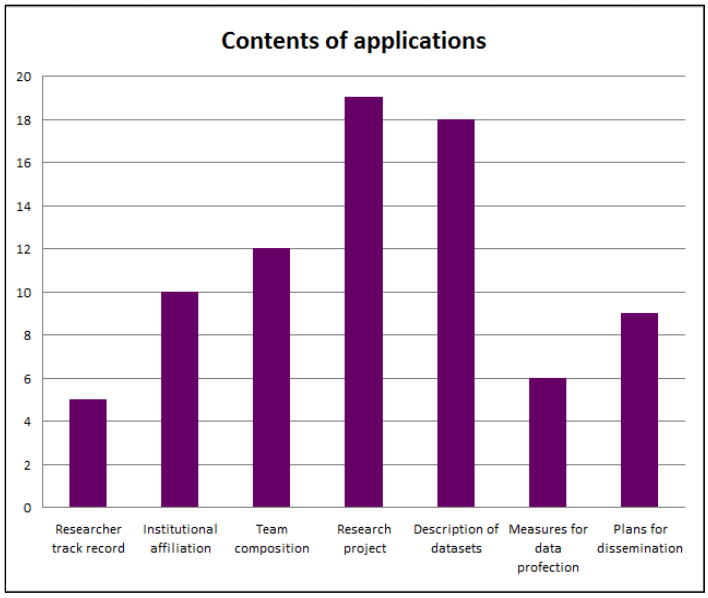

Figure 2: contents of application forms (main items), absolute frequency (N = 21). A country may have different versions of application forms depending on data types.

short-lived owing to people's career moves as well as increasingly frequent fixed-term employment contracts (of post-docs for example). Thus, a future shared system will have to carefully consider these issues and design flexible ways of ensuring institutional support, so as to accommodate for these cases, while not resulting in excessive bureaucracy and burden for data users, research institutions and NSIs alike.

Applications

The other major question is how to submit an application. An overwhelming number of NSIs require a written application, even for less detailed data; online rather than paper submissions are more and more widely accepted. Beyond this basic commonality, only about half of our sample has standard application forms: primarily large European countries (for example France, Germany, Italy, UK) that often make their forms available online. Many smaller countries instead, require

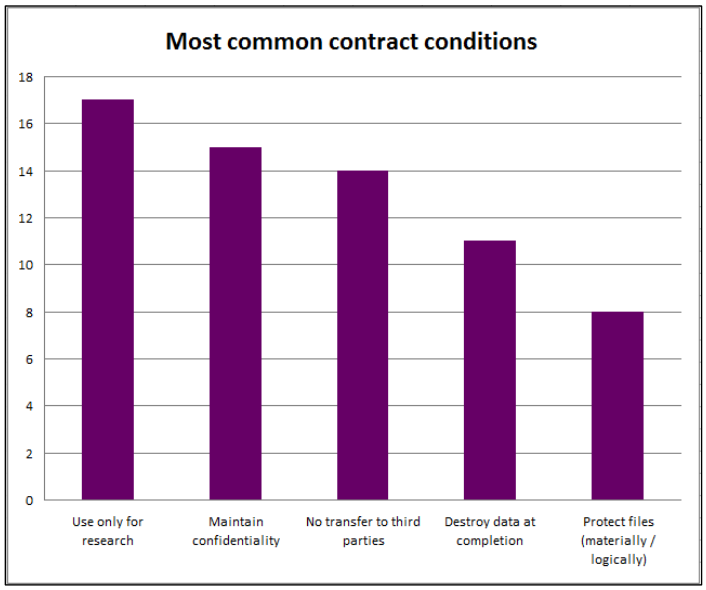

Figure 3: most common conditions in contracts between NSI and researcher / research institution, absolute frequency (N = 20). A country may have more than one depending on types of datafiles and modes of access.

a written letter or email but do not have a standardised form (for example Lithuania, Poland, Slovakia).

Regarding the contents of applications (Figure 2), most countries require a research project, though the expected level of detail may vary. It is usually necessary to include title, composition of the research team, abstract, and a comprehensive list of the data and variables needed. For access to more detailed data, a more complete description of the project is typically required (France for example); in particular to better assess disclosure risk, NSIs may ask applicants to indicate what analyses they plan to undertake or what statistical tools they intend to use. Other elements vary more widely; for example if the applicant is a team, some countries are content with just one application by the team leader on behalf of the group, while others require each team member to apply separately. Another variation concerns signatures: sometimes only the researcher or team sign an application form, sometimes an institutional representative is also required to sign (Lithuania). Finally, a condition often found in cases in which researchers receive data on their own computers (for example on CD-rom or through a ftp server) is to indicate how they intend to physically protect the data: for example using computers with passwords, keeping them in locked rooms, or avoiding copying data on laptops or portable devices (Germany).

In most cases (and almost always when data are highly detailed), if an application is approved, researchers are expected to sign a written agreement before actually starting using the data (Figure 3). This may take different legal forms, from a end user licence to a contract, which we treat as equivalent for the purposes of this study. The most interesting aspect here is the commonality of key conditions, particularly use for research only, no transfer of data to any third party (or no access outside the approved research team), and confidentiality pledges, whereby the researcher undertakes not to attempt to identify statistical units and not to publish results in forms that may enable re-identification by others. Notice that confidentiality pledges are common even when data do not present a very high disclosure risk. Other conditions are specific to the type of access requested: for example when researchers receive the data for use on their own computers, they are usually asked to destroy the files at completion of the project; but this condition obviously does not apply when, instead, they use the data on the premises of NSIs, or through a remote-access secure server where download is disallowed. In such cases, there may be other specific conditions, for example researchers much have their outputs checked for disclosure by NSI staff before being authorised to retrieve them from the system.

Service

Decisions on accreditation applications are overwhelmingly made by NSIs themselves, or dedicated internal units within NSIs, which also manage applications and monitor the whole process. In some countries, however, data archives managed by the research community take responsibility for applications concerning versions of datafiles that are not highly detailed or confidential: for example, ISSDA does so in Ireland. For highly detailed data, some countries have a dedicated authority or commission, such as Comité du Secret in France. Representatives of researchers are occasionally involved in the decision-making process, possibly as members of a Scientific Council in charge of advising the decision-maker. Though these solutions introduce further diversity in the European landscape, they empower researchers by involving them directly in the process, allow sharing costs between NSIs and the research community, and usually result in greater openness, improved efficiency, transparency, and cost-effectiveness (see for example Beagrie and Houghton 2012, for the case of the UK).

A major, widespread difficulty that hinders a smoother application process is lack of adequate communication to users and the general public. Although all European NSIs have a website, and all have an English version of (at least part of) it, many of them provide only limited or no information about existing data, and about criteria and conditions for accreditation and access. Even when this information is available, it is often difficult to locate through standard web search engines; what's more, navigation within a single NSI's website is frequently clumsy. These gaps, observed in many national-language websites, are typically exacerbated in their English translations.

Trans-national access

Can foreign researchers be eligible to access data too? We restrict our analysis to European researchers, who (regardless of their nationality or country of origin) live and work in one of the EU and EEA countries, so that they are subject to very similar legal frameworks on personal data protection. Figure 4 shows that in most cases (UK's "Approved Researcher" scheme for example), they face the same conditions as national researchers and have to go through the same application

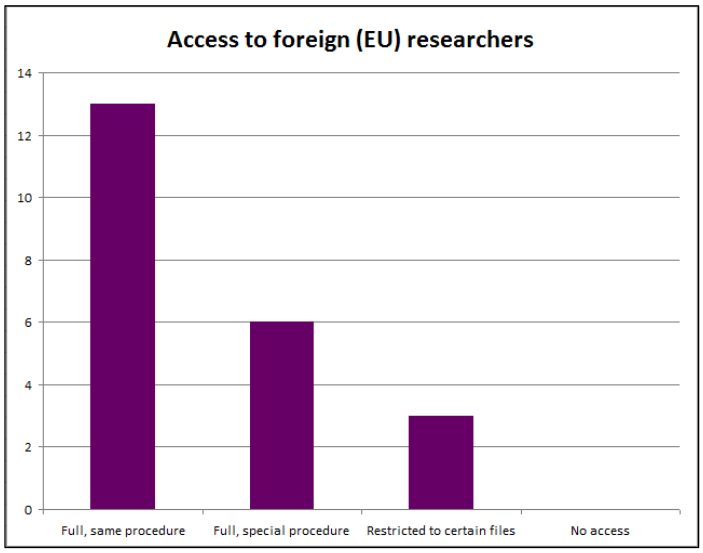

Figure 4: Conditions of access for foreign researchers (from EU and EEA countries), absolute frequency (N = 22). One option for each country.

procedure. In other cases (applications for highly detailed data in France, for example), they are entitled to the same data as national researchers, but may have to undergo some additional procedure or to produce additional evidence, for instance to prove the trustworthiness of their institution. Other countries are stricter and allow foreign researchers to access some types of data files only: in particular Germany cannot distribute Scientific Use Files (that is, data at intermediate level of anonymisation, where disclosure risk is rather low, but not inexistent) outside its borders and requires foreign users to come to its research centres to use its data. Interestingly, however, these limitations apply only to a small number of countries and most encouragingly, no country completely disallows research access to foreigners.

Nonetheless, a major bottleneck revolves around the extent to which institutions, not just individual researchers, need official accreditation. This is an obstacle especially for transnational access in that it is more difficult to ascertain the suitability of a foreign than of a national institution; more importantly, NSIs fear the technical and legal difficulty,

as well as the higher costs, of suing a foreign institution in case of breach. They thus often tend to be particularly cautious.

Other impediments to trans-national research access are subtle, and concern practicalities and procedures rather than general principles. As mentioned above, the English versions of NSIs' websites are less complete than national-language ones, further exacerbating the problem of insufficient information mentioned above. If it is difficult for a national researcher to find out what data are available and how to request them, it is even more challenging for a foreigner. More to the point, conditions for trans-national access are rarely spelled out explicitly, and applications forms, when they exist, do not always have an English version. Further, there are nuances and differences in terminology that may make it difficult for a user to understand to what extent two apparently similar national datasets are comparable (in terms of available variables, degree and methods of anonymisation, etc.). Finally, the architecture of NSIs' websites differs widely across countries, so that data access and accreditation issues are not always classified under the same headings, in a way that makes it difficult for external users to navigate through the European network of official statistics websites.

Discussion and Conclusions

Overall, our key results are encouraging: almost all European countries do provide research access to their microdata, and most of them also allow researchers from other countries within the ERA to access their data, though under varying conditions. Indeed the "open data movement", progress in IT, and pressure on NSIs to extract maximum value from their data collections, have enabled major steps forward in the last few years. In particular, availability of secure IT solutions for access to highly detailed data on the one hand, and increased production of highly anonymised Public Use Files and tabulations for distribution through the Internet, have allowed NSIs to significantly increase their offer of data of value for research. The approach to dissemination of European NSIs becomes more and more researcher-friendly, a significant evolution relative to the past.

Cross-country differences remain though interestingly, they mostly concern actual practices and concrete procedural aspects rather than general guiding principles. It is primarily because of apparently inconspicuous issues that trans-national access remains difficult: in particular differences in terminology and definitions across countries, and unclear or non-explicit rules for trans-national accreditation.

Some of the gaps that we have identified are relatively easy to fill: particularly the observed widespread lack (or incompleteness) of online information about accessible data and conditions for access, and uneven availability of English-language translations. These problems constitute a major practical obstacle for users, but can be cheaply and rapidly solved. To achieve this, we have identified a set of best practices in accreditation, that may provide guidance to all NSIs on how to make progress even in the very short run:

- Availability of complete English translations of NSI websites, particularly the pages dedicated to data access and accreditation;
- Adoption of a more common terminology, use of more similar website structures or of indicators and visual clues that help users to more easily locate information on data access and accreditation;
- Clarity and completeness of information on both general criteria and any special conditions (in particular for trans-national access, but also for a number of other less clear-cut cases such as students' access);
- Clarity and completeness of information on how to apply (including prices, if any, and expected timing);
- Standard application forms (rather than a more generic request of a written letter) with English translations, ideally downloadable from the web and allowing both online and email submission.

The database that we are building as part of DwB, and its future release through the Web, are meant to contribute to this process by further improving the capacity of the system to communicate and maintain openness, transparency, and readability.
Other issues, however, cannot be solved in the short run and require negotiation and discussion at policy-making levels. A first major question, extensively discussed above, is whether institutions need an accreditation procedure together with individual researchers' accreditation, how to recognise foreign institutions, how to assess unconventional or short-lived institutional partnerships, how to account for short-term employment contracts and researchers' career moves across institutions (and sometimes across national borders too).

A second major question concerns terminology, as much of the observed lack of clarity is due to heterogeneity of definitions and denominations. While adoption of a common terminology across Europe may be too ambitious (not least because it would involve appropriate translations into multiple national languages), perhaps a more easily applicable solution would be to create some code to "interpret" the notions defined at national level, and guide users to identifying the closest notions in different countries: for example, to understand the extent to which the "Public Use Files" of a country are really similar (for degree of anonymisation and conditions or modes of access) to those that another country may call "Campus files" or "Anonymised Microdata Files". Such a solution would also involve a great deal of preparatory work, but may be easier to implement for individual NSIs, without requiring them to radically change their operating modes.

A third major question is trans-national access and how to improve it at European level, so as to facilitate comparative research. Several options are possible. The least demanding one would require some form of mutual recognition of accreditation decisions, or perhaps some simplification of the application process in country B, if a researcher has already received accreditation for an equivalent dataset in country A. Even in this case, though, careful attention will be needed to design procedures that do not discriminate among researchers, keep track of data usage in different countries, and allow smooth and continuous communication between the NSIs involved.

A more ambitious option would be to devise a common standard for all member states - with the same criteria for eligibility, shared application procedures and forms, and possibly even similar organisation of the service level. DwB is actively engaging in discussions about a future possible standard, and an emerging concept from these discussions so far is the ambition of a "Schengen area" for researchers. As a long-run goal, this would be achieved first by harmonisation of criteria, conditions, and procedures for granting researchers access to confidential data. Once a high degree of harmonisation is achieved, the potential for integration of researcher accreditation opens up. Why maintain more than one procedure in the ERA if all procedures are the same and deliver the same results? One manifestation of a "Schengen" for researchers is the idea of a "researcher passport" whose definition, applicability and usefulness, however, are still to be assessed. Its creation might be integrated within a "European Service Centre for Official Statistics" (ESC-OS), also proposed within DwB, which would centralise information and handling of procedures

and applications on behalf of participating countries (Mack, Wolf, Esteve and Silberman 2012; Tubaro, Cros, Kleiner and Silberman 2013).

It is yet unclear which of these solutions will be preferred by NSIs and other stakeholders (particularly data archives and of course, the research community), all the more so as different countries may have different preferences. Another issue to be considered very seriously is the funding of these activities, all the more so as the budgets of most European countries are currently under strong pressure. While some of the changes we have identified can be made at relatively low cost (improvements in information in particular), other changes are more demanding. In truth, accreditation procedures are not in themselves very expensive, and any rationalization is likely to further lower down costs; but an improved and smoother accreditation process may globally increase researchers' demand for access, generating additional costs related to service provision – that is, preparation, documentation and secure delivery of data. One solution would be to involve the social science research community and share the burden, as mentioned above, particularly through data archives such as those that are part of the Council of European Social Science Data Archives (CESSDA). The experience of some countries demonstrates that this solution is not only cost-effective, but may even generate high returns to public investment, as recently shown in the case of the UK (Beagrie and Houghton 2012). In addition, enhanced coordination within the European Statistical System, with the leading role of Eurostat, may also contribute to reducing some costs, particularly those related to provision and dissemination of information in a comparable and consistent way across countries. In the years to come, DwB will continue its involvement in discussions in the hope to facilitate an improvement and possibly, a standard European model and a collaborative approach between different stakeholders to facilitate accreditation and access to official statistics.

References

Beagrie C. and Houghton J. (2012).Economic Impact Evaluation of the Economic and Social Data Service, Report for the Economic and Social Research Council. Available at: http://www.esrc.ac.uk/_images/ESDS_Economic_Impact_Evaluation_tcm8-22229.pdf

Mack A., Wolf C., Esteve A., and 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., Cros M., Kleiner B., Silberman R. (2013). Accreditation for transnational research access to official micro-data in Europe. Proceedings of the New Techniques and Technologies for Statistics (NTTS) Conference 2013. Available at: http://www.cros-portal.eu/sites/default/files/NTTS2013fullPaper_145.pdf

Tubaro P., Silberman R., Cros M., Cornilleau A., Kvalheim V., Kiberg D. and 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

UNECE - Task force directed by D. Trewin (2007). Managing statistical confidentiality and microdata access. Available at: http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Acknowledgment

This research has been undertaken as part of the "Data without Boundaries" project, which has received the financial support of the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 262608. We thank members of Work Package 3 for relevant inputs and useful suggestions

Notes

1. Corresponding author: Paola Tubaro, University of Greenwich, Room QM163, Park Row - Greenwich, London SE10 9LS. Email: p.tubaro@greenwich.ac.uk.

Paola Tubaro, University of Greenwich, CNRS
Marie Cros, Université de Lille I
Roxane Silberman, CNRS - Réseau Quetelet