Distributing Access to Data, not Data

Providing Remote Access to European Microdata by David Schiller and Richard Welpton[1]

Abstract

Data about individuals and organizations are routinely collected across the Member States of Europe, through surveys, administrative and business transactions. Yet access to these microdata for research purposes, particularly across national borders, is often restricted for confidentiality or legal reasons. Despite the benefits that accrue to society from allowing comparative research to be undertaken using cross-national data sources, such as policy evaluation, researchers face significant barriers in making comparative analyses of data collected in more than one Member State. Legal restrictions on the dissemination and transfer of data, and the consequential cost of visiting the data within its country of origin, prohibit access. We present a new initiative to build a European Remote Access Network, as part of the Data without Boundaries programme, which arguably represents one of the greatest efforts of recent years to allow researchers from across the European Union to access data collected in more than one country.

Keywords

European Research Infrastructure, Comparative research, transnational research, data access, data security, remote desktop

Introduction

Data about individuals and organizations are collected throughout the Member States of Europe, such as National Statistics Institutes and Research Institutes, for a variety of purposes. These include the production of aggregate population and economic statistics, such as unemployment measures, but also for a number of administrative purposes too. As an example, to calculate state-provided benefits such as retirement or unemployment payments. Yet considerable demand also exists from the research community, to access such data for research purposes, for which microdata are required.

Some data sources are readily available to researchers, often via internet download. These data are heavily anonymised, by means of perturbation, e.g. removing variables, top-coding and aggregation, to protect the confidentiality of subjects within the data, for example, individuals and organizations. Access to detailed, confidential non-perturbed data have previously been restricted, due to the potential risk of identifying the subjects. Yet at the same time, accessing detailed data is becoming more and more important for modern research methods.

In recent years, not only has access to confidential data improved within countries, but exciting developments are afoot in terms of international data access. The Data without Boundaries (DwB) project (dwbproject.org), involving Data Archives, National Statistics Institutes and Universities from around Europe, has published recommendations for cross-border data access (Schiller 2013). The Research Data Centre[2] (FDZ) of the German Federal Employment Agency (BA) at the Institute of Employment Research in Germany (IAB) has established access points in the USA, e.g. at the University of Michigan, enabling researchers in the US to access detailed confidential German microdata (Bender and Heining 2011).

> An RDC, often referred to as a 'secure enclave', is a centre where researchers can access detailed confidential microdata.

Without such initiatives, access to cross-country detailed and confidential data will remain constrained, forcing researchers to visit a Research Data Centre (RDC) in person within each country, pertaining to the

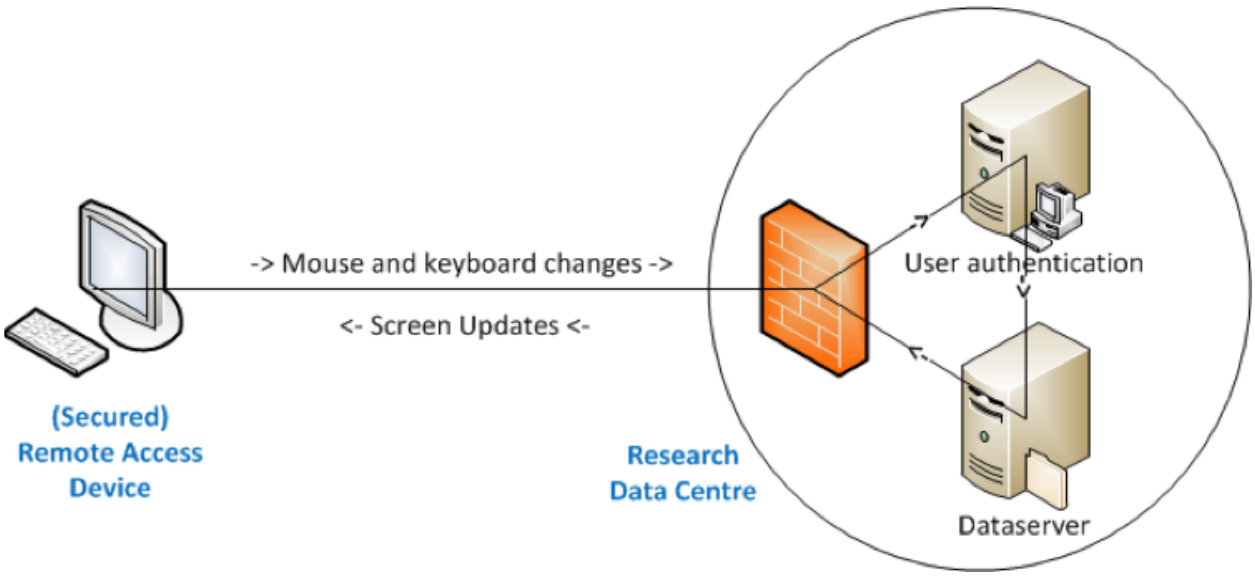

Figure 1 - secure remote desktop

data they require. An RDC, often referred to as a 'secure enclave', is a centre where researchers can access detailed confidential microdata, which are never removed from the RDC, but where researchers can undertake statistical analyses of the data. For example, a researcher wishing to compare the effects of trade unions on productivity in UK and Germany would need to visit the UK to access the UK data, and then visit Germany to access the German data.

This paper explains how the proposed European Remote Access Network (EuRAN) will enable researchers to access data from different countries throughout Europe, from a single location convenient to them. We begin this paper by explaining the concepts of remote desktop access and its alternatives. We then evaluate the demand by researchers for access similar sources collected in different countries, proceeding to explain the concept of the EuRAN and how it might work.

What is remote access?

Generally, remote access refers to controlling an application package remotely. It neglects aspects of security issues or what can be done from the remote locations. Therefore a more detailed description is needed. In the context of this article we talk about the concept of a secure remote desktop to access confidential microdata, and we distinguish between remote desktop and job submission (or remote execution). However, the terms job submission and remote execution are often used synonymously. The table below provides a closer examination of the terms that illustrates their distinctions.

Remote execution would therefore describe the more automated solution, with reference to the table above, we take remote execution to include both job submission and remote execution. For the purpose of this text, the distinction between remote desktop and job submission (remote execution), both as sub-units of remote access, shall be enough.

Remote desktop:

Figure 1 presents a remote desktop solution. To enable work with a remote desktop solution the user needs an access device(this could simply be a plug-in software for their internet browser on their own computer), or a 'thin-client', which is a small computing device configured to access a particular network, to send enquiries to a distant server via keyboard, mouse, or a touchscreen. A secure encrypted tunnel or virtual private network, VPN is established through the internet: enquiries (e.g. commands to analyse the data) travel in one direction; screen updates or results travel in the other. Behind a firewall, secure servers for authentication and working with data are accessed. Since data remain in a secure environment and data modification can only happen under controlled circumstances a secured and controlled environment to undertake research with confidential data is in place.

Only screen updates, such as pictures from a graphical user interface that is used to work with the data on the secured servers, e.g. the statistical package "R" are routed through the already mentioned VPN tunnel to the screen of the access device. This means that encrypted pictures are routed and scrambled across the internet. If a hacker were to crack the encrypted connection he or she will only see views of scrambled pictures of the graphical user interface that wouldn't reveal anything intelligible, thus ensuring security.

Job Submission	Remote Execution	Remote Desktop
A user submits a program syntax (e.g. via email attachement) to an RDC and a staff member has to execute the program on the data server. In the US and Australia, this may also be known as 'remote analysis server'.	A user can start an application package (or at least an process) at the data owners' facilities by themselves, without seeing data.	A user can log into a regular network computer account (e.g. Windows), access familiar statistical software, and access data on-screen.

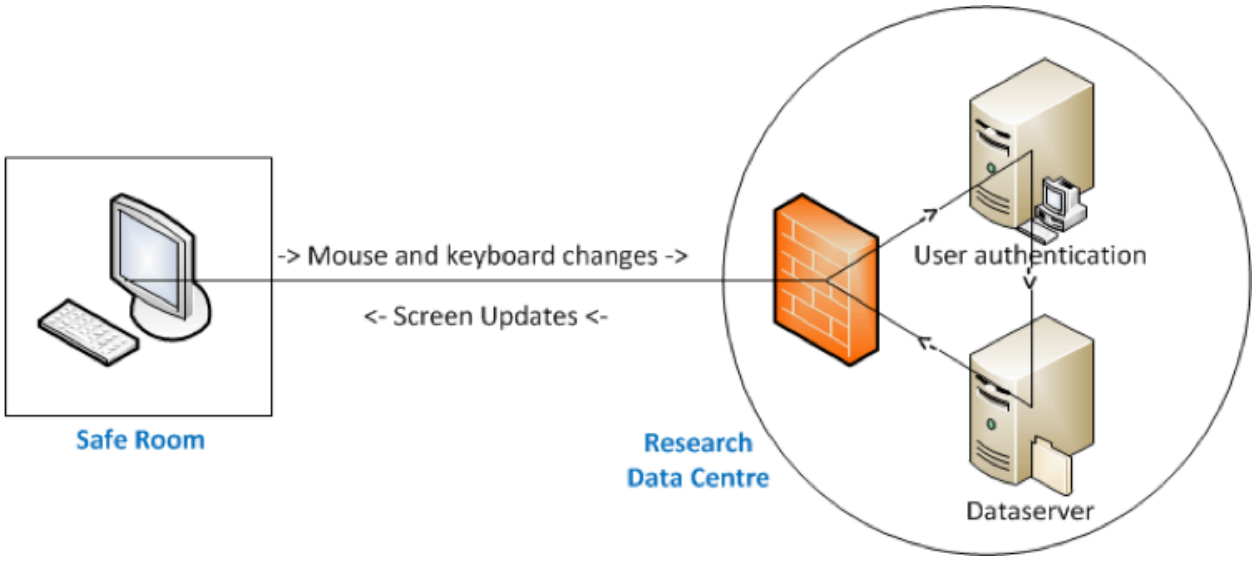

Figure 2 - - secure remote desktop using safe room

The only potential security gap remaining is that the screen of the user can be 'overlooked'. In general everybody within close proximity to the computer can look at the screen, no matter who has entered the accreditation credentials when opening the connection. It has to be kept in mind that, due to technical restrictions, no part of the screen can be extracted from the accessing device. The work undertaken on the server via the computer in use is completely sealed off. For example, it is not possible to copy and paste text or files from the server window on to the local computer. The ability to connect local computer devices such as hard drives to the server should be disabled (likewise, printers, USB sticks etc). Intruders must take pictures of the screen or scribble down what is shown on the screen, a significant effort for little reward for researchers. This already implies an active action by the user that is, under normal circumstances, not to be expected, since proper training and management of researchers (see Desai and Ritchie 2009) will foster correct forms of behaviour by the user.

Additional security measures to control the remote desktop device can also be implemented. These may include technical controls such as incoming IP-address restrictions, validation of hardware certificates, GPS information, as well as biometric authentication (e.g. finger-print recognition). Safe Rooms can also be set up in order to have a controlled environment around the screen (see Brandt and Schiller 2013).

Remote access using Safe Room:

Figure 2 presents a remote desktop solution using a Safe Room. If the data are deemed too confidential to allow remote desktop access from a researcher's institution, a Safe Room is required to allow access to the data. A Safe Room is located in the facilities of a data owner, such as the ONS Virtual Microdata Laboratory or a trusted partner, who are enabled to run the service on behalf of the data owner, such as the UK Data Service Secure Lab. It is controlled and only devices to access the confidential data are located in the room. Adding the physical secure room to the technical security measures creates a complete secured environment for the confidential data, since researchers are identified by staff before access to the data is permitted. In practice from a technical perspective, Safe Room access works identically to secure remote desktop (a computer securely connects to a server via encrypted VPN). What is different is that a physical wall is built around the equipment, and procedures for signing in researchers exist, among with other similar access protocols. In the diagram above, the only difference is highlighted by the square box around the terminal, which didn't exist previously in our illustration of secure remote desktop access.

Job submission:

Job submission (or remote execution, see Figure 3) differs from remote desktop access. The researcher cannot see the data. He or she can only submit program code (syntax) to a data owner, to be run on the data (see Figure 3). The data remains in a secured environment, where data owner staff runs the syntax on the data. The user does not immediately see the results; they are sent to the user after confidentiality checks are undertaken by the data owner. Depending on the specific realization of the job submission solution, the single steps can be automated and subsequently quickened. Due to the setting of job submission solutions neither secured network connection, nor a secured access point is needed.

Both remote desktop and remote execution solutions have their assets and drawbacks. Remote desktop access allows researchers to run calculations in "real time" and to browse the microdata; however, data owners often perceive risks about researchers being 'overlooked' by unauthorised individuals. Job submission provides the highest level of data security because the microdata will never be seen by the user; at the same time analysing data is very inconvenient for the users, due to the fact that they can only "throw" their syntax into the black box and "guess" the next steps by looking into the provided output files. To make the job submission process efficient, investment in 'fake' or synthetic data must be made by the data owners, to allow researchers to write and practice their syntax. Nevertheless, a large staff must be employed to receive and run the syntax, and process the requests for statistic outputs. According to the mentioned assets and drawbacks the best option to support the complete lifecycles of a research projects is a combination of both data access solutions; resulting in enabling researchers to choose their preferred access way depending on the current phase of their research project.

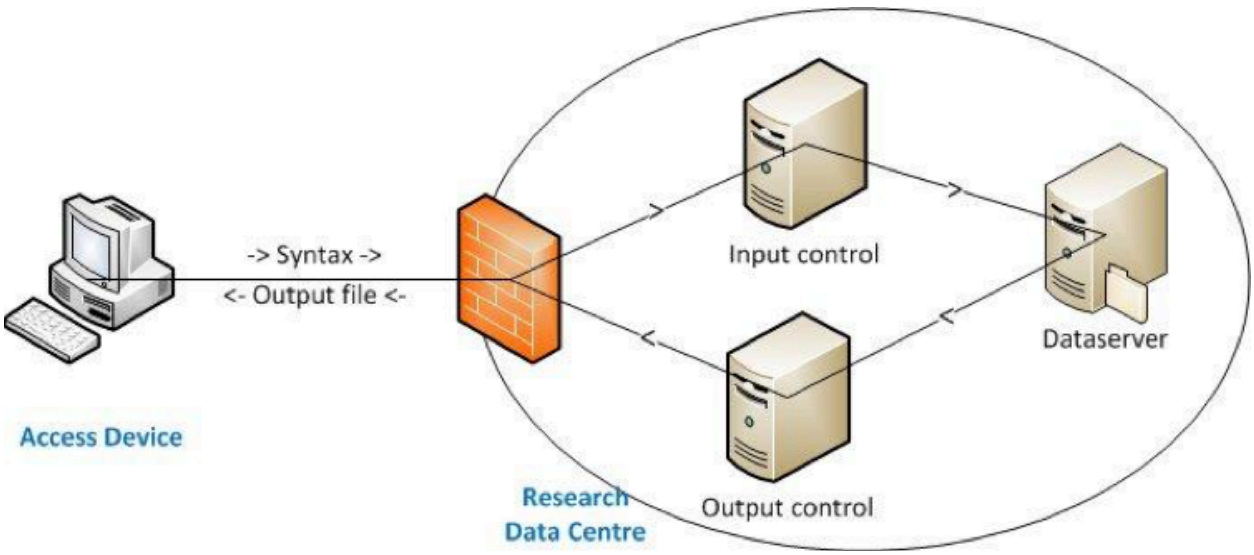

Figure 3 - - job submission solution

In addition, data protection measures described within this article reflect only a part of the possible ways to protect confidential microdata. In order to find the best mixture between protecting microdata and supporting research a portfolio approach consisting of technical, organisational, legal measures and effective researcher management solutions are required.

Assessing the demand for data collected throughout Europe

In this section, we briefly summarise some of the reasons why access to data from more than one country is desirable, and consequently demanded from researchers. Where data are collected by official agencies (e.g. Government, National Statistics Institutes) the quality-assurance processes embedded in data collection methods satisfy the demand from the research community for robust data. In addition, data collected by these means usually provide large sample sizes, a particular feature of data collected through transaction and administrative processes such as taxation.

But regardless of how the data are collected, whether by official agencies or Research Institutes/Universities, we cannot understate the desire of researchers to compare the different institutions, markets and policy implementations across the different Member States of the European Union, and how the outcomes of individuals and organisations differ with respect to these different environments. We believe that more comparative research would be undertaken more frequently if detailed microdata from the different Member States were more easily available. In the current economic climate for example, comparing the effect of unemployment insurance on the employment outcomes of the UK and German workforces requires comparative labour market data from both countries.

In addition, where small-scale survey data are collected by each Member State, researchers could capitalise on the larger sample sizes that may be obtained by pooling observations from different countries. Data on business organisations that operate across the European Union, currently fragmented only by access but not necessarily by sampling frame (Eurostat regulations ensure that data about business organizations are consistently collected throughout the Member States), could even be combined: a researcher could, for example, potentially compare the productivity of UK and German manufacturing plants that belong to the same company.

Finally, data collectors/producers can take advantage of a potentially free source of statistical validation and quality assurance: when researchers analyse data collected by the European Member States, they will easily spot sources of varying quality (particularly in terms of documentation, metadata, and often the robustness of the data themselves). Data collectors/producers can exploit this knowledge to improve collection methods, and enhance the quality of aggregate statistics that they are duty-bound to produce.

For these reasons, access to sources of detailed confidential microdata collected throughout the European Union is a prize highly sought after. The potential for pan-European research, and the implications to test an array of public policies, is enormous.

Distributing Access to Data, not Data

Before the advent of secure network technology (providing secure access as we describe below) data could only be transferred to the researcher. Yet one of the major barriers to accessing detailed microdata throughout the EU is the legal imposition that data cannot be transferred among Member States. Researchers are therefore forced to travel to the destination country where the data reside. However, researchers have many demands on their time. Even without departmental and teaching duties, family commitments coupled with the expense of travelling, dissuade all but the most determined and well-funded researcher from accessing data from more than one Member State.

A similar situation confronts researchers even within their own country – access to the most detailed data may only be accessible by travelling to an onsite RDC; equipped with a lockable secure room where researchers can undertake their research. Even within small countries such as the UK, travelling places a burden of constraint on the researcher.

However, the above mentioned recent advances in secure network technology have the potential to ease this problem considerably. The solution is to distribute access to data, instead of data. The main advantage can be seen in the fact that data owners responsible for confidential data are always in control of their data. The data remain physically in the place they want it to be and they can immediately cut the connection to the researcher, if deemed necessary. While this reassures data data owners that access is secure, researchers realise huge benefits from using remote desktop solutions. There is no need to have confidential data stored on private computers of researchers. They can do the same analysis remotely. Sophisticated working environments enable user friendly solutions. In addition, data sources needed for modern research become available, where they were previously much harder to reach without remote desktop access and the approach of distributing access to data, not data.

Other examples of distributing access to data instead of data exist: for example in the Big Data world. The solutions are similar, even if the objectives are different. Data remain in the same place during Big Data analysis because of network problems that would accrue when moving data from one place to another, primarily due to their size; confidential microdata for social science research are not transferred because of legal issues. This fact opens up the possibilities to exchange new developed solutions for data access, storage and analysis between the two worlds; and bring them closer together by doing so. For the time being developments from single remote desktop solutions to networks of remote desktops are needed.

In the UK, a remote desktop access solution is provided to researchers by the UK Data Archive[3] , and this is known as the Secure Lab. It currently provides access to detailed and confidential economic and social microdata to some six hundred researchers across the UK. And as previously mentioned, the IAB in Germany provides access to German microdata to American and German researchers via Safe Rooms located throughout the USA. Remote access solutions such as these prove that it is possible, with collaborative efforts of the countries involved, to overcome the legal barriers of moving data, by distributing access to data, not data.

Principles for Access to European Research Data

Since we cannot anticipate future changes to the technological and legal environments, we advocate that European data access is founded upon a set of principles (following Ritchie 2005), which are designed to satisfy data collectors/producers and the researchers who wish to access the micro-data. Providing these principles are met, future technological solutions can be implemented, in whatever form they take, and these may differ to the solutions we propose below. We set out the following principles:

1 Access must be distributed. It may not be possible or desirable to physically move data between Member States. But providing access is still possible, this should not matter.
2 Access should come from a single point. The researchers should be able to access all available data from a single point, rather than accessing multiple sources of data from multiple points which is time consuming and costly.

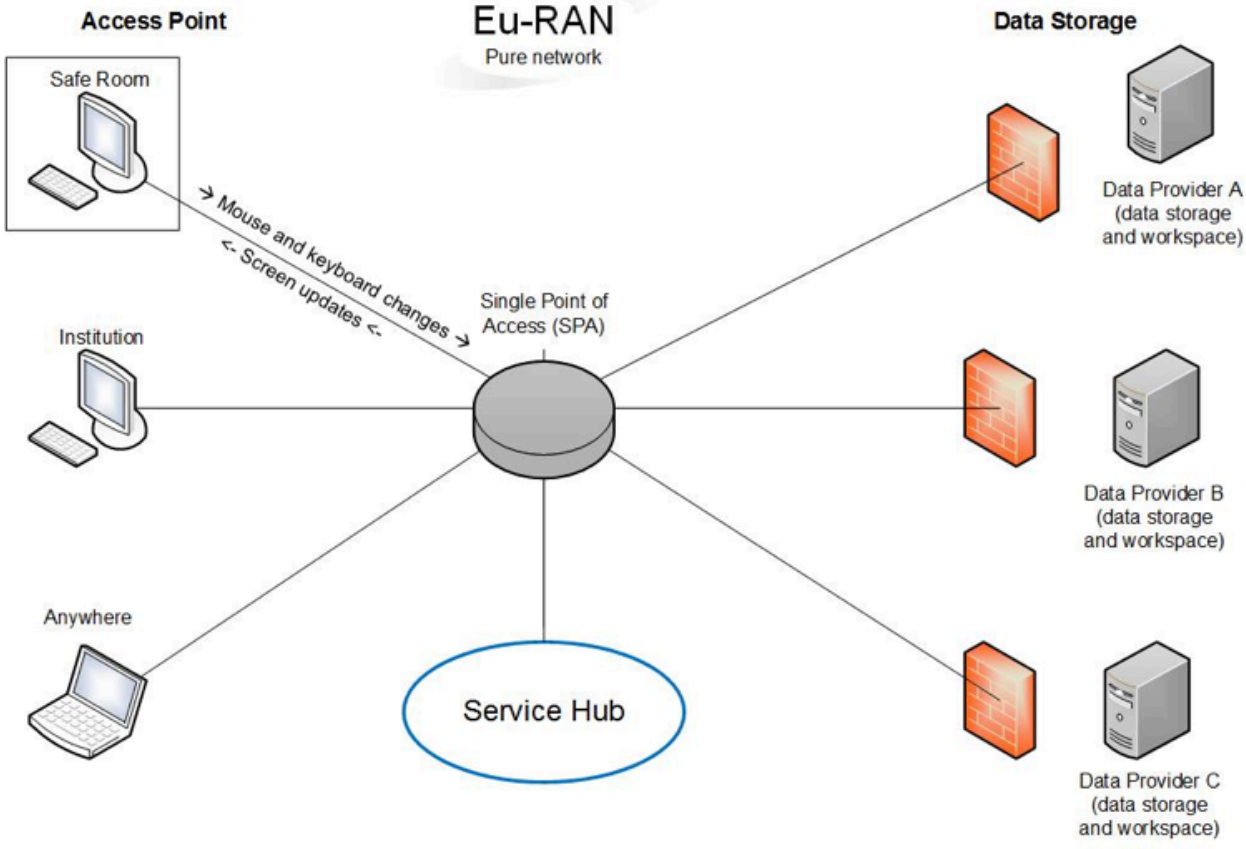

Figure 4 EuRAN

3 Access must be secure. The connection between the researchers accessing the data, and the location of the data, must be secure.
4 Access must be compatible. The access infrastructure that is provided to researchers should be compatible with technological systems used by researchers and data providers alike.
5 Researchers must be able to work collaboratively

We feel strongly that the last principle should be met. As we envision researchers from different Member States will work together on the same research project.

European Remote Access Network (EuRAN)

In addition to the UK Data Service and IAB in Germany, there now exist RDCs that provide access to detailed confidential microdata in the Netherlands, France, Sweden and Denmark, to name but a few Member States. See also 'Data without Boundaries deliverable 4.1: Report on the state of the art of current SC in Europe'[4].

The European Data without Boundaries (DwB) project announced work to create a European Remote Access Network (EuRAN). The aim of EuRAN is to allow researchers based in one Member State to access detailed confidential data from other Member States, without the need to travel to those Member States and to use services that support research on a European level. This section provides more information about how EuRAN will work, and broadly describes its key features.

Many of the Member States already have at least one RDC providing access to detailed confidential data. The EuRAN can build on this existing infrastructures and experiences. The architecture of EuRAN is illustrated in Figure 4.

Suppose we have three Member States (data providers A, B and C). The data would always remain within the Member States, to comply with legal requirements. A researcher, in principle, located anywhere within the European Union, would be able to access the data by a number of methods. In the figure above, these are specifically: from a Safe Room, from their Institution, or even from anywhere (for example to look at data documentation). While the connection between the access points and the Single Point of Access (SPA) always uses a remote desktop approach, the connection between the SPA and the data providers can consist of a remote desktop, remote execution or a job submission solution.

How will the 'user experience' appear? Imagine a researcher based at a UK institution (e.g. university). Depending on the extent of confidentiality of the dataset, the researcher would log into their account either from a Safe Room, their institution computer, or from a laptop at any convenient location. By logging into their account, they will be provided with a simple account, for example a Windows server account, and access data from one or more of the three data providers (which represent a different country). The data are stored in each of the three countries, but the account is set up such that the researcher, from the single account, can click onto one or more data storage devices that takes them to the server which resides in the respective country. For example, if the researcher is granted access to UK data (Data Provider A) and German data (Data Provide B), then they will be able to click and enter those respective storage devices from within their single account.

We now describe the key components of EuRAN, which are illustrated in the Figure 4.

Access Points for Distribution

From a technical perspective, working with data from many sources across the EU remotely is only limited by the possibility of using a device that provides access to a network, usually the internet. However, access nodes, the physical location where a researcher may access data, are often more narrowly defined by legal restrictions and the enforcement of data protection principles according to one or more Member State. For example, in the UK, statistical legislation prohibits access to data only for government staff and 'Approved Researchers', a legal entity created which describes some trustworthy person who has been approved to access data collected under the legislation. Therefore, one cannot access such detailed data using a laptop in a café where many non-approved individuals are located close by. The access nodes for the EuRAN must be agreed by participating Member States.

As mentioned earlier using a remote desktop solution with access device located in a safe room offers a completely secured environment to access confidential microdata. EuRAN supports this access solution. While this approach still forces the user to travel to a location where a safe room is available, a Safe Room Network (Brandt and Schiller 2013) would reduce the need to travel.
A more convenient way to work with data is the possibility to access from the home institution of the researcher or even from within their office. This type of access is now provided by a number of operational remote desktop services in Europe, e.g. the Secure Lab at the UK Data Archive or the CASD in Paris[5] . Researchers can access the secure environment after they have gone through a two-factor authentication, e.g. password and finger print recognition. Finally the EuRAN also supports access from "anywhere". This is possible and needed, when accessing non-confidential services, like data documentation or a wiki shared with other project members.

The principle here is that access is now distributed, the data only remain in the country from where they are collected, but the researcher can access data from all Member States from their home country, rather than travelling to each country. This is essentially an exercise in minimising the movement of the data, while maximising access to the data, subject to legal, organisational and technical constraints. The EuRAN solves this problem for the given set of constraints until further agreement among Member States is such that access can be distributed further. We hope that the future access landscape will be more flexible, 'disaggregating' access to the point where researchers can access the data from a location convenient to them.

Single Point of Access (SPA) and Service Hub

A major design aspect of EuRAN is that researchers should be able to access the detailed European data through a single point. This central access point enables the use of a number of functionalities, e.g. a centralized user authentication system, centralized information platform, workspace for cooperative works, such as projects or the scientific community, storage platform for multinational datasets, and a secure trusted third party environment.

Whether the principle of single access is achieved via a Safe Room or a more distributed access method (e.g. from the researcher's own institution), is not important. However, the single point of access must first authenticate the researcher, and therefore needs a sophisticated rights management system in order to provide the

researcher with all functionalities needed and ensure security and compliance with the restrictions of the data owners.

The user should experience a familiar and customizable working environment. 'Back office operations' via a service 'hub' manages databases and applications that offer additional services demanded by researchers, such as documentation, metadata production etc. (see Burghardt and Schiller forthcoming for more details of such an operation).

Secure Connection

A third principle of establishing a EuRAN is that the connection to the data must be secure. As noted above, it would be wise to take advantage of evolving technologies.

Implementing this in practice can be achieved by using encryption techniques and secure Virtual Private Network (VPN) technology, which is currently used by e.g. the UK Data Archive Secure Lab and the IAB. This provides a secure encrypted connection between the user at the access node and the computer server, which stores the data. Such technology is widely used by the banking and military sectors which rely on confidential up-to-the-second data.
In addition to meeting the secure connection principle, a compatibility principle must ensue, whereby the connection interface, the point at which a researcher logs into an account is compatible with all suppliers of the data and users of the data: otherwise it will be impossible to join the network with the various data suppliers, and researchers themselves will not be able to use the interface.

The single point of access and its connection to the servers will always occur via a secure remote desktop solution. The EuRAN will support a combination of remote desktop access, remote execution and job submission solutions as may be necessary depending on the researcher needs and data owner requirements. These solutions can work together via a single point of access.

Data Storage

The extent to which data can be stored outside of the Member State in which they are collected depends upon national legislation and the disclosure risk of the data. Typically, interpretations of national legislation prevent the distribution of confidential data outside of national boundaries. Our final principle therefore is that data do not travel. But this is not necessary. Modern technology, as shown above, allows access to data without the need for data to be physically moved. Data can be accessed using existing storage facilities that are currently provided by the RDCs of the Member States. Using the model of a single point of access, the researcher simply authenticates themselves when logging into through their access point, and will securely access storage facilities at the various RDCs of the Member States.

If security restrictions ask for it, parallel data storage infrastructures can be established within the existing RDCs. This would result in two data storages: one isolated within the RDCs and the other one as part of the secured EuRAN infrastructure.

While data must remain in the Member States where they were collected, it is up for discussion and agreement as to the circumstances that output files, containing statistical results from analyses are allowed to be moved to and stored in the single point of access.

Microdata Computation Centre (MiCoCe)

The EuRAN can provide access to data storage systems of different RDCs, where researchers can work with confidential microdata and save their work. One of the services provided by the service hub in the single point of access could be a Microdata Computation Centre (MiCoCe). It would provide storage space, statistical software and above all computational power required to bring data together from different countries for comparative analyses in a secure IT environment.

As mentioned previously, confidential data cannot be transferred from one Member State to another. When trying to promote European research instead of national research, a solution to send enquiries from one single point and run calculations on multiple data sources stored in different physical locations is needed. A workshop[6] held by DwB (in late April 2014) demonstrated some of the potential approaches that could act as a solution. However, harmonization of data sources across Europe is needed to push transnational research in Europe, at least the storage of interim outputs from calculations with multiple data sources should be possible within the single point of access of EuRAN.
Furthermore, the MiCoCe could also function as a storage and computation centre for new RDCs that do not want to invest or do not have the resources to build the whole infrastructure by themselves. If legally possible these RDCs can use the capabilities of MiCoCe.

Virtual Research Environment (VRE)

We understand that access to data by itself is not a means to an end, and that in order for researchers to undertake scientific enquiries of the data, they require tools in which to do so. These are encompassed within a 'Virtual Research Environment', a workspace provided to each researcher who accesses the EuRAN. This workspace can be protected by different security levels, depending on the disclosure risk of the data source involved.

The basic requirement is a workspace which includes analytical software and applications to generate, prepare and present results. However, we believe that a EuRAN must be built with collaboration in mind. We anticipate that researchers from different Member States will work together on projects. A 'collaboratory' must be available, similar to that which will be available in the UK Data Archive Secure Lab, which allows researchers working on the same project to securely share and discuss results. IAB is also involved in a project developing such an environment. This again is subject to evolving technology. At a basic level, a shared project area, accessible only by a group of researchers working on the same project, should be available. But more advanced solutions, including instant messaging to allow real-time communication between researchers, would aid research productivity and are therefore highly desirable.

The availability of tools such as MethodBox[7] can provide a one-stop data support solution for researchers. This tool can enable access to documentation and metadata relating to the data the researchers are using, can allow communication between researchers and data owners on queries, and provide support for one another. Such tools foster understanding of the data, and the data producers can view community discussions about their data with a view to improving data collection and preparation methods for the future. In addition, the burden of research support for the data producers and RDCs will surely be minimised if researchers can support each other through online forums, as an example.

This provides but a flavour of the Virtual Research Environment. As technology develops, future devices for enhancing the working and support environment ought to be provided, indeed it is likely that researchers will drive the demand for collaborative tools..

Information Platform

Building a data infrastructure network such as EuRAN must be complemented with dedicated support. By this, we mean user support functions that researchers can avail themselves of. These support functions provide help to users for accreditation, applying to access the data, including support for finding and selecting appropriate data sources and completing relevant application documentation; and support while analysing the data, which may include the production and promotion of documentation and metadata. These support functions constitute the 'Information Platform', and is likely to be the first point-of-contact a researcher will have with the network. This is also an opportunity for the Member States to harmonize these support activities, which are currently provided by the individual Member States. Researchers should receive access to the same information and support, regardless of the Member State where they are based. A platform such as this is described in 'Data without Boundaries deliverable 5.1: Report on the concept for and components of European Service Centre'[8].

Developing EuRAN

To demonstrate how the EuRAN could be established, three RDCs delivered by the CASD (France), IAB (Germany) and UK Data Service, have begun a project to provide access to each others' detailed, confidential data. This will begin with the installation of thin-client terminals in the safe rooms of each RDC. Each thin-client will provide direct remote and secure access, using the VPN technology described previously, to each RDCs' collection of detailed data. For example, a thin-client installed in the UK Data Archive Safe Room, can be used by researchers based in the UK, to access data available from the IAB. This will be a pilot project, which can be developed into the full EuRAN that we have described in this paper. The pilot project will provide pragmatic solutions to many of the technical, legal and organisational issues, which will surely need to be addressed as the EuRAN begins to emerge.

Although the pilot will be established by these three particular Member States, feedback on development and progress of this pilot will be shared and discussed with other Data without Boundaries participants and external experts. The network will therefore evolve with successive iterations, to help improve the concept.

Thus far, our overview of the development of the EuRAN has focused on information technology. At this point, some digression is required because we recognise that an access network which relies on security, dependence on technology alone is not sufficient. As Desai and Ritchie (2009) point out, if researchers are treated as a risk by data producers and/or data providers, researchers have little incentive to consider themselves as responsible for the security of the data for which they are accessing. Achieving 'buy-in' from the researchers accessing the network will therefore be crucial, not just in terms of establishing an effective easy-to-use network, but also for achieving data security. Part of the role of the EuRAN will be to foster such engagement by actively managing researchers to achieve the involvement of the community of researchers in protecting the confidentiality of the data.

Conclusion and Outlook

This paper has summarised the advantages of a more integrated network of access to detailed and confidential data can bring about. We have presented a technological solution, supported by principles of European microdata access to ensure that collectors of data and researchers who analyse the data are equally satisfied. Technological solutions will evolve in the future: but the underlying principles required for secure and collaborative access can be met by an array of solutions.

The Data without Boundaries project consists of many projects, of which the EuRAN development is only one. Other hard work, including an examination of legal issues, statistical disclosure control of results, and training of researchers, to name but a few, has also been undertaken by National Institutes and Data Archives throughout Europe, and will directly contribute to the future development of the EuRAN.

As a result of this project, we anticipate that the landscape for accessing detailed European data will soon be very different, to the advantage of the research community, and to society which benefits from the comparative research that can be undertaken using data collected throughout Europe.

References

Bender, S. and Heining, J. (2011), „The Research-Data-Centre in Research-Data-Centre Approach: A First Step Towards Decentralised International Data Sharing", IASSIST Quarterly, Fall 2011, p. 10. http://iassistdata.org/iq/research-data-centre-research-data-centre-approach-first-step-towards-decentralised-international

Brandt, M. and Schiller, D. (2013), "Safe centre network – need for a safe centre to enrich European research", presented to the Joint UNECE/Eurostat work session on statistical data confidentiality, 2013, available at http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_3_Brandt_Schiller.pdf

Burghardt, A. and Schiller, D. (forthcoming),"Introducing the Service Hub for Remote Data Access".

Ritchie, F. (2005), "Access to Business Microdata in the UK: Dealing with the Irreducible Risks", presented to the Joint UNECE/Eurostat work session on statistical data confidentiality, 2005, available at http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2005/wp.29.e.pdf

Desai, T., and Ritchie, F. (2009), "Effective Researcher Management", presented to the Joint UNECE/Eurostat work session on statistical data confidentiality, 2009, available at http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2009/wp.15.e.pdf

Schiller, D. (2013), "Proposal for a European Remote Access Network (EuRAN) – main components", presented to the Joint UNECE/Eurostat work session on statistical data confidentiality, 2013, available at http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_3_Schiller.pdf

Acknowledgement

We acknowledge the work and contributions of the Data without Boundaries Work Package 4 participants: Atle Alvheim (NSD), Steve Bond (ONS), Anja Burghardt, Iris Dieterich (IAB), Leo Engberts (CBS), Kamel Gadouche (CASD), Maurice Brandt, Christopher Gürke (destatis), and Roxane Silberman (CNRS). 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). We also acknowledge previous unpublished work by Felix Ritchie and Richard Welpton presented at the 2011 IASSIST conference entitled "Access without Borders". Some of the ideas of this unpublished work are presented in this paper.

Notes

1. David Schiller, Institute for Employment Research (IAB), Nuremberg (Germany); email: david.schiller@iab.de
Richard Welpton, UK Data Archive, University of Essex, Colchester (UK); email: rwelpton@essex.ac.uk
2. http://fdz.iab.de/
3. Information about the UK Data Service Secure Lab is available at http://ukdataservice.ac.uk/use-data/secure-lab.aspx
4. Visit http://www.dwbproject.org/about/public_deliverables/d4_1_current_sc_in_europe_report_full.pdf
5. http://casd.eu/
6. http://www.dwbproject.org/events/workshop-micoce.html/
7. https://www.methodbox.org
8. Visit http://www.dwbproject.org/export/sites/default/about/public_deliveraples/d5_1_european_service_centre_report.pdf.