Project N°: 262608

Acronym: Data without Boundaries

DELIVERABLE D4.4

System for Accessing Transnational Microdata and Description of Implementation Procedures

WORK PACKAGE 4

Improving Access to OS Microdata

REPORTING PERIOD:	From: Month 37	TO: Month 48
PROJECT START DATE:	1st May 2011	DURATION: 48 Months
DATE OF ISSUE OF DELIVERABLE:	September 2015	
DOCUMENT PREPARED BY:	Partners 9, 18 and 28	IAB, UKDA and GENES

Combination of CP & CSA project funded by the European Community
Under the programme "FP7 - SP4 Capacities"
Priority 1.1.3: European Social Science Data Archives and remote access to Official Statistics

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

TABLE OF CONTENTS

INTRODUCTION

Easy transnational access to national official microdata, particularly for the highly-detailed ones (Secure Use Files) that are under specific legal frameworks, is one of the most important challenges when it comes to improving European research. The Data without Boundaries work package (WP) 4 (improving access to OS microdata) is responsible for the technical improvement of European data access. In deliverable 4.1 a state-of-play of remote desktop solutions provided by European NSIs and CESSDA data archives was evaluated and presented. Building on those findings, deliverable 4.2 proposed a European Remote Access Network (EuRAN) to scale up transnational European research. With deliverable 4.3, WP4 moved from evaluation and concept work to real implementations. While 4.3 showed an exemplary implementation of a one-way solution (a server was installed at IAB to allow access to IAB data); deliverables 4.4 to 4.6 describe first steps on the way from isolated access applications to a real remote access network (by establishing access solutions that depend on more than one partner; e.g. sharing services like dealing with the data access point and the data storage solution). Therefore deliverable 4.5 focuses on the technical implementations of a secure remote access test environment between IAB and UKDA to prepare access to real confidential data by real researchers based on a RDC-in-RDC approach. It is complemented with a Proof-of-Concept (PoC) demonstrating the feasibility of a real European remote access network between 3 RDCs (GENES-CASD, GESIS-GML and Destatis) in 2 countries (France and Germany) representing 3 national nodes federated by a central node through highly-secured tunnels. As the PoC was designed to evidence the potential of such a RA network in terms of IT security, dummy (i.e. anonymised, non-confidential) data were used; but it is fully functional when it comes to accessing data from different partners within a secured network. To that end, IAB and UKDA were focusing on the RDC-in-RDC approach and CASD on delivering the PoC for the network approach. This sharing of tasks and responsibilities did already start within this deliverable: the first part (description of partners) focuses on CASD, IAB, and UKDA to assess each potential partner's requirements and possibilities when it comes to accessing data across borders; while the second part (the Memorandum of Understanding) focuses on IAB and UKDA as only those institutions require a RDC-in-RDC approach in order to allow accessing their highly-detailed data from abroad. Deliverable 4.6 will describe how actual research projects have started working with the newly-developed access connections set up between IAB and UKDA. All developments should be put together in the end and built the basis for further implementation work.

WP4 is mainly about technical solutions that do not care about borders whereas negotiations about organisational, IT and legal aspects do. Therefore, deliverable 4.4 builds the bridge between technical solutions conducted by a single institution and technical solutions developed by a network of partners. In addition, deliverable 4.4 takes into account different national environments and different specific institutional requirements by proposing and implementing organisational measures to overcome brakes to data access. Therefore, the discussed access solutions are built on remote desktop access to confidential (secure use files) microdata, taking into consideration the constraints faced by several RDCs / countries and requiring a RDC-in-RDC approach (which is one possible definition of a network) to allow transnational access. The novelty in this approach is that two partners (namely IAB and UKDA) are not only sharing services - like data providers giving access facility responsibilities to other partners - but are also establishing a reciprocal agreement between them. Such an agreement between two partners located in different countries of the European Union will also work as a template for other partners willing to join in or start networks by

themselves. In other words, an organisational basis for a European Remote Access Network (EuRAN) is established.

First the involved partners are described in detail before conceptual approaches to support cooperation of different institutions can be evidenced. The final chapter shows, as an example, the establishment of an agreement between two partners requiring a RDC-in-RDC approach to allow transnational access to their highly-detailed data; thus resulting in a Memorandum of Understanding (MoU) that can also be used as a template for other institutions willing to cooperate in this field and for further developments towards a real network.

1. THE PARTNERS OF THE PILOT

Work package 4 “improving access to OS microdata" consists of eight partners (CASD, CBS, CNRS, DESTATIS, IAB, NSD, ONS and UKDA). The pilot implementation carried out in task 4.3 and resulting in deliverables 4.3 to 4.6 are done by three partners – CASD at GENES (France), IAB at the Federal Employment Agency (Germany) and UKDA at the University of Essex (UK). In order to understand the different institutional settings, the three partners must be described in a structured way. First, a general description of the partner is given. Then the following questions are answered: What is it (kind of institution)? Where and under which law (location and responsible national law)? Which data (types of data, especially level of disclosure risk)? Which access ways (are offered by the institution)? Requirements to allow remote access to data stored in the home institution (from same country, Europe, elsewhere)? Requirements to allow data to be stored somewhere else (same country, Europe, elsewhere)?

1.1 CASD at GENES (France)

The “Groupe des Écoles Nationales d'Économie et Statistique” (GENES) is an independent research & training institution, composed of 5 entities:

- a research centre: the “*Centre de Recherche en Économie et Statistique*” (CREST) with 9 laboratories;
- a school of continuing education: the "*Centre d'Études des Programmes Économiques*” (CEPE) ;
- a research data centre: the “*Centre d'accès sécurisé aux données*” (CASD), i.e. Secure Remote Access Centre for confidential microdata for social sciences and economics;
- 2 “Grandes Écoles” (higher-education) :
  - the “*École Nationale de la Statistique et de l'Administration Économique*” (ENSAE, Malakoff), that trains statisticians, economists and INSEE’s administrators,
  - the “*École Nationale de la Statistique et de l'Analyse de l'Information*” (ENSAI, Rennes), that trains statisticians and INSEE’s associates.

GENES trains a total of about 600 students per year, who will join the civil service (employees) or large private companies, banking and insurance, international institutions, etc. GENES also includes hundreds of permanent researchers and associates.

What is it?

Since 2007, GENES - in close collaboration with INSEE (the French NSI) - has been in charge of developing the first remote secure access centre for confidential data to be introduced in France.

Following the classifications of Eurostat and the European Commission, GENES is a National Statistical Authority (NSA).

Where and under which law?

GENES and INSEE are located in the same building in Paris.

For a long time, access to highly-detailed microdata, especially the official microdata produced by INSEE and the SSM (Ministerial Statistical Services) was very restrictive due to both the Privacy

Protection Law of 1978 ("Loi Informatique et Libertés") and to the strict Statistics Law of June the 7th, 1951 when it comes to personal data.

In 2004, changes occurred in the 1978 law; and the recent 2008 amendment to the 1951 statistical law (law no. 51-711 dated June 7th, 1951) provided conditions to set up a highly secure equipment allowing wider access for researchers to highly-detailed and sensitive data. The 2008 amendment provides an exemption for academic research purposes and offers researchers an access to sensitive data - such as households data or business data coming from statistical surveys or administrative documents, for research purposes exclusively. Researchers' access authorisations are issued by law after advice by the Committee for Statistical Confidentiality, and agreement of the competent administrative authority.

CASD has been built upon the long cooperation history between research and INSEE, which has led to the creation of Réseau Quetelet, the French Data Archives that provides access for researchers to an important collection of anonymised official microdata (SUF). This cooperation is organised under the aegis of the CCDSHS ("*Comité de Concertation pour les Données en Sciences Humaines et Sociales*"), the French national Data committee for social sciences and humanities that was created in 2001 by the Ministries in charge of Research, Education, Employment and Economy. Under the presidency of the MESR ("Ministry of Higher Education and Research"), the Committee gathers representatives of research institutions and universities with INSEE and governmental statistical services. The Committee that is in charge of a data policy for research in social sciences pays specific attention to access to official microdata in respect of good practices and protection of confidentiality.

Which data?

Researchers can use detailed statistical data through the following channels:

- by using aggregated data and some Public Use Files (for INSEE only) that are widely available on official websites (INSEE, etc.). These data are not detailed enough and are therefore not so useful for researchers;
- by using specific files for researchers, SUF ("Scientific Use files" or "Fichiers de Production et de Recherche", FPR in French) and bespoke tabulations distributed only for researchers through the Réseau Quetelet in cooperation with INSEE, several statistical services and other data producers.
- by using CASD to get access to confidential microdata (highly-detailed and sensitive data) concerning numerous domains:
    - INSEE Household Data: Demographics, Salaries, Employment, Transportation, Housing
    - INSEE Business Data: Annual Corporate Surveys, Financial Links
    - Agriculture Ministry: Agricultural Census, Crop-growing Practices Surveys
    - Forthcoming: Individual and Corporate Tax Data (law currently being changed), Labor Ministry Data, Ecology Ministry Data

These data are often administrative data that offer wide possibilities for cutting-edge analysis but are difficult to anonymise. They are produced in France by a large number of organisations or institutions (INSEE, SSM, academic institutions, etc.).

Which access ways?

CASD provides only access to confidential microdata but in two ways: remote access from the researcher's institution (that entails contracts with both the institution and the researcher) and transfer to their own on-site access for specific OS data producer (the statistical departments in the ministries). It is currently being envisaged that Eurostat data could be accessed remotely from accredited RDCs in the countries where the researchers would have to sit, as advocated by the ESSnet DARA project. However, at the moment, a solution for accessing Eurostat data is not implemented yet and it is not clarified whether CASD will be allowed in the first implementation phase by Eurostat to provide this access (at the moment only RDCs from NSIs should be accredited) or whether INSEE will have to implement a specific solution.

Access way	**Provided by**	**Anonymization level of Data product**
Free download	Insee website (no license)	PUF (Public Use Files also usable as Campus File)
Restricted download and use	Réseau Quetelet	SUF and bespoke tabulations
On-site stay at GENES - Paris (safe room)	CASD	Highly detailed data from specific data producer (stays in Paris)
On-site stay in researcher's institution via secure remote access	CASD and researcher's institution (via contract)	Highly detailed data from Insee, SSM

Requirements to allow remote access to data stored in the home institution (from same country, Europe, elsewhere)?

Same country: To get access to the microdata, the researcher has to use a specific access unit, SD-Box, guarantying imperviousness as the sole means for accessing the central infrastructure where the data are stored. A contract is signed between GENES and the institution hosting the SD-Box:

- Unit placement constraints (i.e. not in a high-traffic area, screen not readable by other people, room locked with a key)
- Establishment identified by its postal address and network address.

In addition, the researcher must attend a half-day enrolment/training session at GENES; promoting awareness of legal and computer system aspects and regarding statistics confidentiality. During this session, researchers record their fingerprints on their access cards. GENES then supplies an SD-Box. The researcher must also sign the 'User charter' (Confidentiality Agreement, Terms of use: not transferring the access card, etc.)

Europe: Same as above.

Elsewhere: Same as above and access only from trusted countries.

Requirements to allow data to be stored somewhere else (same country, Europe, elsewhere)?

Currently, there are no possibilities to store confidential microdata somewhere else in case researchers would like to combine them with other data for running single analysis.

It could be discussed between the data producer and the partner institute.

1.2 IAB at the Federal Employment Agency (Germany)

The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) provides the documented research data, coming from administrative process of the BA and from surveys carried out by IAB, in a structured way to the scientific community. The FDZ is financed by the BA; and uses the infrastructure of the IAB.

What is it?

The FDZ is a unit of the German Federal Employment Agency. It is directly subordinated to IAB's Directorate. Within IAB, the FDZ is part of Research Area F "Methodology and Data Access". The Federal Employment Agency is a self-governing supreme federal authority under the legal supervision of the German Federal Ministry of Labour and Social Affairs, with currently about 115.000 employees. Following the classifications of Eurostat and the European Commission the Federal Employment Agency is a National Statistical Authority (NSA).

Where and under which law?

FDZ and IAB are located within the headquarters of the BA in Nuremberg (Bavaria), Germany. Data collected through the BA and the IAB are governed by the German Social Code Book (SGB). If SGB does not apply, the German Federal Data Protection Act (BDSG) is applicable.

Which data?

Research data provided by the FDZ are administrative data, deriving from the notification process of the social security system and other administrative processes of the BA, and data from surveys carried out by IAB. While the administrative data cover the complete work force that is subject to social insurance contributions (exception: Self-Employed People , civil servants and family workers) and measures by the Employment Agency (only sub-samples are provided to researchers), the survey data is collected based on a given sampling-frame. All research data are microdata; on individuals/households and/or establishments. The original administrative data have an extremely high disclosure risk. This risk is reduced by the sub-sample; while the disclosure risk of the survey data is already lower due to the sampling-frame. The resulting highly-detailed research data still contain a substantial disclosure risk. This risk is minimized by anonymization techniques like aggregation resulting in (1) Scientific Use Files (SUF), that are available via restricted download for restricted use by a named group of researchers for a given project; and in (2) Public Use Files (PUF, also Campus File), that have no disclosure risk at all and are freely available on the internet. The highly-detailed versions are only available during a guest stay in one of the locations of the Research Data Centre of the BA or via job submission. BA administrative data are also available merged with IAB survey data and/or additional information from other data sources.

Which access ways?

The FDZ offers a variety of access ways fitted to the disclosure risk of the regarding data product.

Access way	Anonymization level of Data product
Free download	Public Use Files (Campus File)
Restricted download and use	Scientific Use Files

Job submission[1]	Secure Use Files (data stays in Nuremberg)
Onsite stay in Nuremberg (safe room[2])	Secure Use Files (data stays in Nuremberg)
Onsite stay in different locations[3] (safe rooms) via secure remote access	Secure Use Files (data stays in Nuremberg)

Requirements to allow remote access to data stored in the home institution (from same country, Europe, elsewhere)?

Same country: Secured connection and access from a safe room that offers the same security standard as the on-site stay room in Nuremberg (physical access control, only devices to access the research data in Nuremberg, etc.) located within the facilities of a trusted partner (in general organisations that are experts in dealing with confidential data). In addition, contracts that fix agreed-on responsibilities between IAB and the partner institute are needed.

Europe: Same as above.

Elsewhere: Same as above and access only from trusted countries.

Requirements to allow data to be stored somewhere else (same country, Europe, elsewhere)?

This is currently only possible for SUF and done at IZA in Bonn (Germany) and CISER in Ithaca (USA). Requirements are: a trusted institution and contracts that fix agreed-on responsibilities between IAB and the partner institute.

1.3 UKDA at the University of Essex (UK)

The UK Data Service provides Secure Access dissemination methods to enable secure access to confidential data. The secure access methods provided by the UK Data Service include Secure Remote Access and Safe Centre access.

What is it?

The former mechanism began operation in August 2011, and now allows 400 researchers (as of May 2013) across the UK to access 45 sources of confidential business, economic and social data that previously could only be accessed by an onsite facility. Researchers are provided with access to a standard windows platform that is accessed remotely using Citrix encryption technology. Essentially, the researcher is allowed to log into a server based at the UK Data Archive. Inside this environment, the researcher is provided with the data for which they have been approved and statistical software and text editing software. Therefore, they never download the data, or access the data on their own computer, but instead undertake all of their research work on the UK Data Archive server. Statistical

[1] Job submission: The user sends his program code; the staffs of the FDZ runs it on the highly detailed data, checks the output for disclosure risk and sends it back to the user.

[2] Safe room: A room with physical access control and equipped only with devices to access the highly detailed research data remotely.

[3] Currently: Berlin, Bremen, Dresden, Düsseldorf, Mannheim (all Germany) and Ann Arbor, Berkeley, Ithaca (all USA). UKDA is the first European access points to secure use files outside of Germany.

outputs are released to them via email after a check for confidentiality (known as statistical disclosure control) is conducted on the results by UK Data Service staff.

Researchers make an application to access the data that are passed on to the relevant data depositor(s) for approval. Once approved, the researcher attends a mandatory half-day training course organised and run by UK Data Service staff. The course covers: how to use the Secure Remote Access system; the researchers' role in data protection; and an introductory course on applied statistical disclosure control (enabling researchers to prepare non-confidential research outputs themselves from within the Secure Access system that require the minimum of vetting by UK Data Service staff). After successfully completing the course, the researcher is provided with their credentials, which they can use to access their account, following the installation of a small Citrix software application that launches the secure connection to the UK Data Archive secure server.

The second mechanism provided by the UK Data Service for accessing confidential data is the Safe Centre located at the UK Data Archive at the University of Essex in Colchester, UK. The system works exactly as described above with Secure Remote Access. The vital difference is that researchers must access the data from within a secure room at the Archive. As well as making a successful application to access the data, and complete a training course, the researcher will visit the Safe Centre. As with Secure Remote Access, the researcher is provided with an account allowing them to remotely log into the secure server. However, monitoring by staff is much more frequent than using the Secure Remote Access. For example, the Safe Centre can only be accessed by authorised personnel using swipe cards, which logs entries and exits from the room. The Centre is monitored by CCTV. In addition, researchers must provide a form of identification when they visit the Centre, and they are prevented from taking personal items into the Centre. For these reasons, the data made available are considered far more sensitive or confidential than the data provided through the Secure Remote Access route. This justifies the extra security mechanisms that are in place.

Where and under which law?

The SDS is located at the campus of the University of Essex in Colchester. Researchers accessing data using the Secure Remote Access option can do so from their institution. Until October 2012, this option was only available to academic researchers based at Higher Education Institutes and Research Institutes in the UK. Since this time, the remit of the UK Data Service has expanded to include access to non-education and private sector establishments. At the time of writing however, the UK Data Service does not currently provide access to individuals working in these sectors.

The restriction to access data only at the researcher's institution is based on agreement made between data depositors and the UK Data Service, which at the time the service was established, represented a great leap forward in terms of access to confidential data in the UK. Another reason for providing access to researchers at their institution is the assurance that institutes will have up-to-date anti-virus and anti-malware software installed on their networks, and that in the UK, the JANET network provides some isolation from the general internet. A two-step authentication is required: researchers must enter a username and password to access their account; researchers must provide a static routable IP address for the computer at their institution that they wish to access the Secure Remote Access route on. This IP address is checked by UK Data Service staff to ensure that access is provided to a computer at a UK higher education institution, and provides the second authentication. This two-step method of authentication has recently been re-approved by the UK Government as a secure access verification method. As discussed previously, access to the Safe Centre occurs at the UK Data Archive.

Access to data provided by the UK Data Service Secure Access routes is governed by a number of statutes. The relevant legal framework depends upon the source of the data. For example, access to data for research purposes collected by the Office for National Statistics is governed by Section 39 of the Statistics and Registration Services Act 2007. Data collected by the Department of Work and Pensions are governed by the Social Security statutes. The Data Protection Act 1998 also governs access to all data. It is beyond the scope of this report to detail the legislation, but it is worth mentioning that the various legislations share the same concepts with regard to accessing confidential data:

1. Access may be provided for research purposes
2. The research purpose must in some demonstrable way, serve a 'public good'
3. The confidentiality of the data must be protected
4. Breaching the confidentiality of the data, and especially where distress is experienced by a data subject as a consequence of this breach, is a personal criminal offence
5. The 'public good' derived from accessing the data, must exceed the private cost to a data subject were information exposed.

In addition, it is worth mentioning that the individual legislative frameworks are subject to interpretation by the data depositors. The legislation themselves do not prescribe how data are accessed per se. Instead, data depositors decide how data access mechanisms ensure that the principles enshrined within legislation are met.

Which data?

The UK Data Service provides access to social and economic microdata, and a number of macro data too. These data vary in:

- Source (they are deposited by many data producers, including Government Departments, Research Institutes, and by researchers themselves)
- Access (researchers access data via a spectrum of access mechanisms, from download to restricted access only through our Safe Centre at the University of Essex, with intermediate options in between. Each access option depends upon the extent to which the data are considered confidential and or sensitive, with the least easily downloadable.)
- Type (the majority of the data are quantitative, but a significant number of qualitative datasets are also available in different formats, for example, transcripts, audio recordings etc).

A range of datasets from a number of sources are currently available through the Secure Remote Access route. These include:

- Business microdata, including firm-level data on productivity, innovation, employment, skills, industrial relations, electronic transactions, trade, investment
- Social survey data, including the Labour Force Survey with detailed geographies and other individual characteristics not available for download
- Data about children and young people, including sensitive data that would likely cause distress if exposed
- Combined administrative and survey data

Which access ways?

Access way	Anonymization level of Data product
Free download	PUF (Campus File)
Restricted download and use	SUF
Secure Remote Access (accessible to researchers from their institution)	Highly detailed confidential data – only direct identifiers are removed
On-site stay in Colchester (Safe Centre)	Highly detailed data which include direct identifiers

Requirements to allow remote access to data stored in the home institution (from same country, Europe, elsewhere)?

Secure Remote Access provides remote access to data housed in our Data Centre, at the UK Data Archive. Researchers using this service must be based in the UK. This is because data depositors have interpreted the legal framework that provides access to these data, such that access is only permissible in the UK. In addition to this interpretation, Secure Remote access requires that the researcher signs a User Agreement, which stipulates the conditions under which their access is governed. This is counter-signed by the researcher's institution, binding the institution to terms and conditions of the researcher's access. This is necessary, because in the event of a serious breach, the Economic and Social Research Council, the funding council of the UK Data Service, may withdraw funding from a researcher's institution. This penalty has been established to deter a researcher from breaking the terms and conditions of their access. This is not a penalty that could be applied to institutions outside the UK, which are not funded by the council.

Requirements to allow data to be stored somewhere else (same country, Europe, elsewhere)?

Currently, the restricted collections of the UK Data Service are stored at the UK Data Archive under contract with the data depositors. These data depositors may decide that the data could also be stored in another country. However, it would not be possible for the UK Data Archive to physically store data in another location (within or outside the UK). This is because we are required by data depositors to hold the ISO 27001 accreditation in order to store, process and disseminate restricted data. The UK Data Archive has recently received its second certification for this standard. Placing data in an environment that does not meet the ISO 27001 would not be acceptable. The costs of accrediting another location are likely to be prohibitive.

Summary

The DwB project aims to provide access to data, which are generally considered restricted, due to their confidential and/or sensitive nature (for example, it may be relatively straightforward to identify an observation, either by detailed geographic reference or by sufficient detail and number of variables to allow identification). All three partners of the pilot provide highly-detailed data under restricted conditions because of the legal frameworks and data security issues. They developed secure access routes that enable them to make this confidential data usable for researchers. At the same time the organisational structures, the respective data protection law, the data sources provided, and access point restrictions differ between the three institutions. All partners are currently not allowed to store data outside their facilities, due to the responsibility to care about data protection issues, which can be done more easily within the borders of the own institution. In the Work Package 4, and for the purposes of the pilot, IAB, UKDA and CASD have worked together to find common ground that enables them to provide secure access to confidential data from each other. While CASD does not require non-resident researchers to access data from another RDC, this is currently one solution for IAB and UKDA. Therefore, for this first implementation, while IAB and UKDA will provide a 2-way connection to each other's RDC, CASD will only function as a point of access to IAB and UKDA.

2. FROM ISOLATED APPLICATIONS TO A REMOTE ACCESS NETWORK (RAN)

All three partners have running remote desktop solutions; enabling researchers to work with the confidential data of the corresponding data provider. DwB deliverable 4.1 showed examples of other remote desktop solutions currently operating in Europe. All of them enable remote access to one data storage facility. Research projects willing to work with data from more than one data provider - may it be from one country in order to use a wider range of information on individuals or from different countries to widen the sample size of the surveyed group or to do comparative research- have to use more than one access solution. In the example of the three partners involved that would mean: using an SD-Box in your institution to access CASD data, go to a Safe Centre in Germany to access IAB data, and go to the UK to access UKDA data. To harmonize data access, conserve resources, ease work with more than one data source, and enable transnational research; the isolated remote access applications have to be advanced into a remote access network (RAN). The concept of such a European Remote Access Network (EuRAN) was explained in DwB deliverable 4.2.

Although the concepts and architecture suggested for the EuRAN are based on an actual network of RDCs. Referred to as a "pilot", the work that will be further described in this report is based on a RDC-in RDC approach, particularly useful immediately for RDCs that can only allow remote access from accredited RDCs. It is therefore a first form of a RDCs network with connections working in parallel, however different from the pan-European network envisaged for a EuRAN. That being said, it represents a first and necessary step toward the implementation of a two-way transnational secure remote solution to confidential microdata and provides useful lessons in that perspective. This is complemented by a Proof-of-Concept demonstrating the feasibility of a true network of RDCs with a central node in terms of IT security (see deliverable D4.5 for further details).

Beside the technical implementations needed to build a EuRAN, organisational and legal aspects have to be taken into account in order to build a basis for cooperation between partners. In the end, the best technical solution will not be accepted if it is not embedded into a working organisational structure and if it is not in line with legal procedures.

3. BASIC BUILDING BLOCKS FOR A RAN

A running RAN needs to be built on organisational concepts of cooperation. Three of them are described below; although none of them can be used as an isolated solution. They can be used as complementary approaches, which all build on trust and agreements.

3.1 Shared services

The concept of shared services is based on agreements between partners regarding given tasks. This concept is not limited to data access procedures but can also be used for data documentation, data archiving, or data anonymisation tasks. The idea is that a given task and the related responsibilities are clearly defined. Based on agreements containing those definitions, partners can take over the given tasks. When talking about data access, the Safe Centre Approach is one example of shared services. In this remote desktop setting, data stay in the servers of the data provider but the physical access room that also contains the access device is hosted by another party. Responsibilities of the partner hosting the access room are set by agreements with the data-providing party. This concept of shared services can easily become more sophisticated when a real EuRAN with multiple partners is implemented and/or a secure virtual research environment is built to host services and act as an access infrastructure.

3.2 Circle of Trust

In the context of the DwB project and in parallel with the OECD discussions between NSIs, the notion of a "Circle of Trust" was outlined to refer to the need to create circumstances where different parties, such as research data centres and data archives or universities, can rely on each other. Mutual trust is needed for sharing microdata services, i.e. exchanging microdata or providing access to confidential microdata. A concept for basic requirements is necessary for the data providers, and essential for transnational microdata access, but relevant also for the researchers seeking access to data.

When creating a Circle of Trust, each member joining the Circle should be accepted according to the same rules and conditions which are approved by all members. These would cover confidentiality rules and security requirements, but also competence and legal aspects. There would also be set preconditions for the institutions themselves or for technologies providing the access.

Measures needed to gain trust among the actors and to establish a Circle of Trust are:

- shared best practices
- collection and documentation of rules and protocols for transparency
- cooperation agreement
- harmonised contracts for microdata access
- guidelines for the treatment of microdata requests
- catalogue of rules to check which institution can be approved to access microdata
- Researchers' Passport
- security concept and accreditation guidelines for safe centres

- list of security and user demands for a remote access system
- anonymisation concept for scientific use files
- rules and protocols for the transmission of microdata
- guidelines for statistical disclosure methods and output checking, and
- common understanding of responsibilities and similarities.

The concept of Circle of Trust helps to define reliable rules for a mutual trustworthy cooperation between partners with different protection needs, due to different laws, data, or technical implementations.

3.3 Equivalence

The concept of equivalence directly derives from the trust-building tasks initiated between partners. The idea behind it is that organisations that are working more or less in the same business, that have the same responsibilities, and share the same tasks should trust each other because they understand what the others do, how they do it, and why they do it; simply because the organisations, their requirements, their procedures and their legal frameworks (including penalties in case of breach of confidentiality) are equivalent. Focusing on Research Data Centres (RDC) that are made to give access to confidential research data, equivalence would help to easily evidence that the main tasks of such organisations – like secure data storage, disclosure control, access rules – are done in a similar, understandable and trustable way. According to that the concept of equivalence enables cooperation between organisations by building trust upon shared understanding and a basis for establishing mutual responsibilities.

In sum, those three mentioned approaches for cooperation between partner institutions need to be evaluated, criticised and put into practice. The following chapter shows the procedure followed by IAB and UKDA in order to get an agreement about accessing each other's confidential data via a Safe Room approach. It can actually be seen as a kind of "proof of concept" for the Circle of Trust.

4. A MEMORANDUM OF UNDERSTANDING BETWEEN IAB AND UKDA

IAB offers remote desktop connections from physical safe rooms to the confidential data stored on the secured servers at IAB. Safe Rooms are located at the facilities of trusted partners in Germany (currently six locations). No Safe Rooms have been established in Europe outside of Germany yet. UKDA offers remote desktop access to confidential data from UK institutions. No remote access from outside of UK was possible.

Within the DwB project, the two partners started a pilot to enable access to each other's data. Before caring about technical solutions; agreements and an organisational framework had to be put into place. This chapter describes the procedure, the results, and the challenges faced.

4.1 Legal aspects

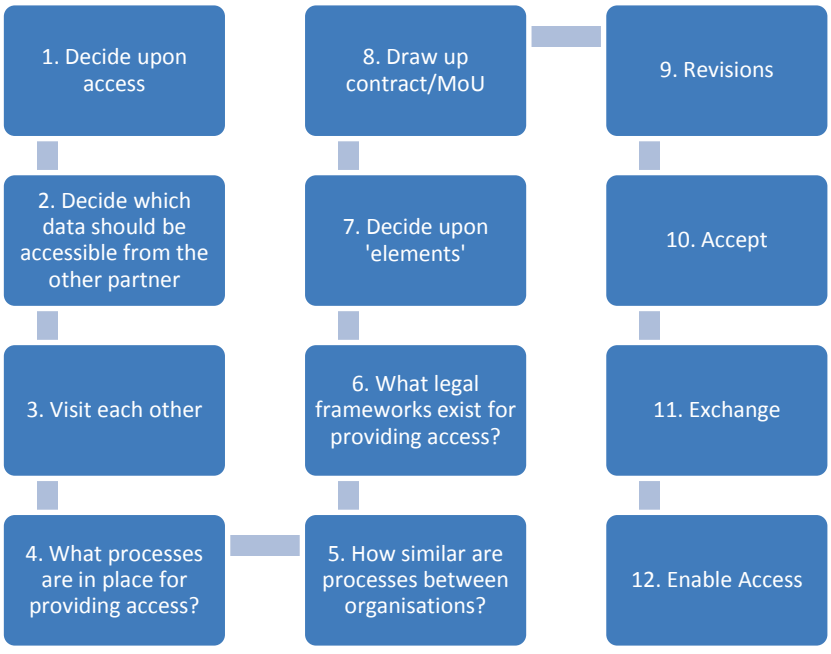
This section presents information about the legal arrangements that have been established to enable secure access between IAB and UKDA. These can be considered as 'prerequisites' for enabling access between the two organisations: given the sensitivity and confidential nature of the data involved, it was important that a number of arrangements were agreed upon by both parties.

The aims of establishing an agreement between two organisations include:

- To ensure that each party understands its responsibilities with regard to providing access
- To equip each party for carrying out its responsibilities
- To avoid misunderstandings about roles and responsibilities
- To foster an environment of understanding and collaboration between the two organisations

Summary of workflow

In order to establish an agreement between the two organisations (known hereafter as 'the parties'), a number of stages were undertaken. These are summarised in the diagram below:



These aspects are summarised in more detail as follows:

1. Decide upon access	The two parties will need to agree to provide access to each other's secure service through their onsite facility. This is potentially a significant undertaking in terms of staff resources (managing bookings, supervision etc.) so is important to assess the likely resource implications.
2. Decide which data should be accessible from the other partner	Both parties will already provide secure access to confidential sources of microdata. In the case of IAB, all data accessed through the RDC belongs to the organisation. At UKDA however, the organisation essentially acts as a 'third-party' service provider for a number of 'data owners' (e.g. NSIs). Therefore UKDA spoke to one of these data owners (the Understanding Society Data Access Committee) to obtain agreement that access to these data could be made available through the connection. Upon completion of a successful pilot, other data owners will be approached and asked to consider allowing their data to be accessed through IAB this way.
3. Visit each other	Understanding how each party works is crucial to this agreement. Therefore, it was necessary for staff from each party to visit each other, and to find out how they provide secure access to their confidential data sources. The visits help to establish a basis of trust, and allow the parties to compare their working practices. This will enable them to ascertain any differences in working practices, which may need to be harmonised in order for the effective operation of the service.
4. What processes are in place for providing access?	Understanding how each service provides access is essential. That way, when the legal agreement states how the parties will operate, there will be understanding on both sides (critical given that both parties are based in different countries where language could be misinterpreted).
5. How similar are processes between organisations?	If the parties provide access in similar ways, then providing access should not result in any additional costs in terms of resources. Any differences may entail resources, or may also affect information security standards applied by both parties. So, at this stage, differences should be identified and efforts to harmonise or at least understand the differences and take them into account can be undertaken.

6. What legal frameworks exist for providing access?	This is important to understand, as any national legislation that is applicable to access to personal data (which confidential data normally are considered to be) must be adhered to.
7. Decide upon 'elements'	A legal agreement will have a number of 'elements'. By this, we refer to the aspects that both parties are expected to undertake to meet the terms of the agreement. For example, under the element 'safe setting', one party may declare they will provide a safe room, while the other party will provide the necessary secure IT equipment to enable the connection to their server. Each element represents an area of activity, and for each element, each party states the actions they will agree to undertake.
8. Draw up contract/MoU	Using the elements, and including the legal information, this stage involves putting together the material into a formal agreement.
9. Revisions	The agreement is exchanged and the two organisations suggest revisions.
10. Accept	Once these revisions have been accepted by both organisations, then the agreement can be exchanged.
11. Exchange	Representatives of the organisations sign the agreement, copies of which are sent to the corresponding organisation.
12. Enable Access	Once the signed agreements have been exchanged, then the technical connection between the two organisations is established.

Demonstrating Equivalence

Demonstrating that the services operate in an 'equivalent fashion' was key to designing the agreement, and reassuring data owners in both countries. If a UK data owner 'trusts' the UKDA to deliver a secure service, then in theory they should also trust IAB, particularly if IAB operates along similar lines as UKDA. This was the basis for sharing services.

Two visits were arranged between UKDA and IAB staff. The purpose of these two visits was to ensure that both organisations were able to offer suitable access to each other's data. The visits included a review of arrangements, which in turn were translated into 'elements' of the agreement. These are explained in more detail in "Elements", the next section.

Where such a visit identifies equivalence or differences in operational procedures, these should be documented and assessed to identify whether solutions are required. Where possible, and with the will of both parties, convergence should be encouraged.

4.2 Elements

Section	Elements	Description
Definitions	All terms used throughout the document to be defined here.	**Objective: Avoid misunderstanding about terms.** In order to also avoid misunderstandings about operational issues arising later. This is crucial for what is a relatively fledging industry where new terms are created and used interchangeably. These terms may relate to technological or procedural aspects, but also most importantly, have legal definitions. They should be defined tightly and without ambiguity to avoid misinterpretation.
Services	This refers to all the aspects of operating an RDC.	**Objective: Ensure clarity about what services each organisation can expect the other to provide.** It is important to list the various operational aspects of operating an RDC. Firstly, it allows both parties to be aware of each others' activities. Secondly, it allows both parties to consider which activities they will be responsible for when providing access to each others' data (and again, should avoid any misunderstanding about which party is responsible for undertaking which activity).
Set-up / installation	Provision of equipment and installation	**Objective: Specify the responsibilities of each organisation for establishing the secure connection.** This section defines which party is responsible for providing equipment to allow access to the data. It should also specify that the equipment will remain the property of the party providing the equipment. In addition to simply providing the necessary equipment, this section should also define which party is responsible for installing it, and how the work should be allocated. For example, the receiving party should be aware of roughly how many resources will be required in order to install it, and how much support they can receive from the sending party.

Application (to access)	Applications relating to access to the data	**Objective: Establish which organisation should handle applications by researchers to access the data.** This section describes which party is responsible for the promotion and administration of access to the data. This refers to the process by which researchers order and apply to access the data.
Account	RDC IT accounts for researchers	**Objective: Establish which party is responsible for setting up the IT accounts that will be used by the researcher for logging into the thin clients provided.** It may be superfluous to add this since a remote access system can only be maintained at source, which includes setting up and providing IT accounts. Therefore the party based at the location of the stored data can only set up accounts. However, this technical establishment may not be widely understood and should therefore be explained.
Access control and monitoring	Providing access to researchers & Monitoring use	**Objective: Establish which organisation is responsible for critical security procedures.** For example, 'signing-in' procedures may be in place (e.g. a researcher has to provide photo-ID when visiting the RDC). Since access occurs in a different country to where the data are stored, the responsibility of applying these procedures may fall on the RDC staff hosting the thin client. In addition, the conditions under which researchers who have been admitted to the RDC are to be monitored, may be stipulated. For example, the data-sharing organisation may require that researchers access the data in a room where it is not possible for them to take photographs, use writing materials etc. These criteria should be explicitly stated.

Support	Supporting users with: • Access; • Using the data; • Documentation;	**Objective: Allocating responsibilities for supporting researchers (users).** It should be defined which party will take responsibility for supporting users, and to what extend this support covers. This is particularly important because the party providing access to the other party's data will also be providing access to its own data collection. Therefore, explicitly defining what support researchers visiting the RDC can expect from onsite support staff is important, to avoid duplication/use of resources not planned for. In fact, the party providing the thin terminal will be the 'owner' of the data, and should therefore undertake all support for users with regard to access (applications, but not entry to the RDC they are visiting), using the data, documentation of the data, and SDC.
Statistical Disclosure Control	Undertaking SDC checks Delivering outputs to researchers	**Objective: Allocating responsibility for output checking.** Similar to the IT account creation and issuing element, SDC on the statistical results can of course only be undertaken by the party for which the data are stored and which provides the thin client. As before, it may be superfluous to include this element, but for the purposes of individuals not familiar with the concept of remote access, it is important to include a statement here about SDC, particularly since this is one of the most resource-intensive parts of running a service.

Training	Training for researchers Training for RDC staff	**Objective: Allocating responsibility for training researchers.** An agreement should specify whether the receiving RDC staff should be trained. This might include: • How to sign-in researchers accessing the terminal • How to switch on the thin client • How to monitor researchers • Any other procedures for providing access It may be the policy of an RDC that researchers accessing data through remote secure access are provided with training. This normally covers: • Using the service • Legal responsibility/ethical issues • SDC This section provides an opportunity for the parties to specify whether such training is to be provided, and which party will provide the training to researchers.
Maintenance	Maintenance tasks	**Objective: Allocating responsibility for maintaining the connection.** This includes: which organisation is responsible for maintaining the thin client; what work may be expected of the host organisation for undertaking any maintenance, or help with maintenance, of the thin client, and ensuring the secure connection works.
Safe Room Security	Authentication of devices Applicable laws and procedures Internal auditing External auditing	**Objective: State and allocate responsibilities for ensuring Safe Room access is secure.** It is clearly important to set out what each organisation expects of the other in terms of achieving Safe Room security. This should be clearly stated without ambiguity. In addition, 'auditing' requirements (i.e. the 'right' of one organisation to inspect the other organisation's Safe Room to ensure security is maintained) should be specified.

Period of agreement	Duration	**Objective: Specify the time period for which the agreement is valid, and how much notice should be provided if one or both parties decide to withdraw.**
Available research data and fees	Available research data and fees	**Objective: Specify which datasets are to be made available through the secure connection, and whether any fees for access should be charged.** In the experience of this project, both organisations believed that it would not be beneficial to be specific about which datasets would be made available. This may change in the future as new datasets become available, and if specified in the agreement, then the document would need to be renewed every time a new dataset were made available.
Appendix: Description of services		**Objective: ensure the organisations understand the services.** Secure remote access to confidential data is a recent development, and unless personnel have been involved in running such a service, they may not be familiar with the concept. A description of the secure access mechanism is advisable. For example, a legal representative may not be familiar with the technical or organisational aspects, and this section should provide information that can easily be understood.

Memorandum of Understanding (MoU) vs. Contract

It was decided between the IAB and UKDA to draw up a MoU rather than a legal contract, to cover the IAB-UKDA connection. A MoU can be considered to be a record of an agreement established between two or more organisations. This differs from a legal contract in which, if one partner reneges on the agreement, legal actions can be taken by the other partner. A MoU is used in situations where the legal framework is unclear. This is obvious in the case of the DwB project, because data protection legislation across Europe is generally 'silent' on shared data access to personal data between countries. For this reason a MoU between IAB and UKDA was preferable: from a practical point of view, this is more enforceable. The two organisations could establish and share an access link and, if necessary, could withdraw from the agreement (including the withdrawal of access) without legal repercussions.

Legal actions after breaches of contract would also not be efficient in this setting. No one could be sure that such legal actions would be working in the legal system of the other country. The biggest security for adding control as part of the trust relationship established within the MoU lies in the reputation of the involved institutions. IAB and UKDA are highly respected as experts in treating confidential data in a secure way. If a connection between both parties has to be rescinded because of non security-compliant actions by either party, the reputation of this party is in danger. And this is likely worse than any legal action caused by a breach of contract.

4.3 Overcoming challenges

The process of reaching an Agreement to provide access to data took many months. This section identifies some of the hurdles that were overcome by both organisations, which may serve as useful advice for agreements between organisations that may be arranged in the future.

Throughout the process, project partners understandably enquired as to why making an arrangement took as long as it did (the process began in November 2013 and concluded in February 2015). For example, other partners mentioned that IAB already provides remote access to several institutions in the United States. First, it should be noted that the data made available to researchers based at US institutions are not Secure Use Files. Rather, they are Scientific Use Files, and therefore do not contain the same amount of details as provided in the Secure Use Files. The IAB-UKDA connection represents the first time that IAB Secure Use Files can be accessed from outside Germany. Secondly, an appropriate type of agreement had to be drawn up. Given the lack of European legal framework concerning transnational access to 'personal' data, some considerable care had to be applied to ensure the right agreement was made between the two organisations. For instance, before deciding on having a MoU the possibilities of have a legal binding contract were evaluated. One of the crucial questions that could not be solved under the umbrella of a contract was the following: if a researcher accessing data stored at IAB servers in Germany (having a contract with IAB in Germany) tries actively to undertake forbidden actions like disclosing individuals while he or she is accessing from the Safe Room at UKDA in the UK; which law should be applicable? Should it be the German Law because that is the place where the data is stored and the contact with the researcher was signed; or should it be the UK law because the offence was done by a user sitting in the UK? This could not be solved by the law in place, especially because there is no European law for this case regarding national microdata. In addition, this may imply the specific interpretation of the law in the specific institution. Keeping this in mind, the MoU was a much better way to move forward but discovering so took time. Finally, it should be noted that if the two organisations both 'owned' their data, then the process of agreeing access terms would have been simpler. However, while IAB effectively 'owns' the data provided through its secure access system (IAB data is under the German social security code; individuals were not asked to give their information to researchers; it is done due to freedom of science and to common good), UKDA does not 'own' any of the data it makes available to researchers. Instead, negotiations are made with data owners, such as the Office for National Statistics, for UKDA to hold and provide access to the data on behalf of the data owner. For this reason, UKDA could not simply 'enter' into an agreement to make its confidential data collection available to researchers via IAB. Instead, UKDA had to speak to the data owners that provide the organisation with data so as to find out if any were willing to allow their data to be accessed via IAB. One did, the Understand Society Data Access Committee. However, the UKDA had to demonstrate that access to these confidential data via IAB would remain secure. Much of the work involved in demonstrating 'equivalence' of the services was undertaken to reassure the Data Access Committee on this aspect. Agreement by this committee occurred in mid-2014.

5. ADOPTION OF THE MEMORANDUM OF UNDERSTANDING

Furthermore, the Memorandum of Understanding used between the IAB and UKDA has been a model template for the IAB-GESIS and the IAB-CBS (Statistics Netherlands) connections. This suggests that it can be used more widely, which has always been intended. Acting as a template, this agreement contains the basic information to enable an agreement between two parties to provide access to data. 'Elements' can be added or removed as necessary, reflecting the circumstances of the parties involved. The fact that the GESIS-IAB and the CBS-IAB arrangements have adopted this template provides demonstrable impact of the work invested in this project, and the Memorandum of Understanding therefore constitutes a key DwB output. It should also be noted that the application and adoption of this agreement demonstrates that many RDCs operate in similar ways;thus demonstrating 'equivalence' is a straightforward process. We therefore expect other organisations who wish to provide access points to each others' data to adopt this template and build upon it. The ultimate objective of this template is that it could be used for a transnational secure data access mechanism involving a number of services. For example, if the EuRAN develops into an established service, then such an agreement will require acceptance by three or more services; thus becoming a consortium agreement. Given the similar characteristics shared by these secure services, we believe that the Memorandum of Understanding used for this project could be adopted more widely.

5.1 Lessons for other services wishing to connect

Any two or more organisations wishing to establish a connection will need to consider the issues inherent in these aspects, and agree upon a course of action that is suitable to all parties. This may sound daunting, however, given that in our experience, secure data access services operate in similar ways, the challenge of finalising an agreement is not insurmountable. This experience is shared by other initiatives too. For example in the UK, the Administrative Data Research Network (ADRN) brings together a number of services to provide secure access to administrative data sources. The four centres that make up the network are already established secure data services, and are working together to share standards of operation in a number of operational activities, given similar methods of working and a shared common goal. The UK Research Data Centre Network is an informal group of UK RDCs that provide secure access to confidential sources of data, and include the UK Data Service and Office for National Statistics Virtual Microdata Laboratory. Best-practice operational procedures are discussed and often implemented. These two groups demonstrate that future agreements to combine services and access are possible, because secure access services generally tend to operate in similar ways. The step-by-step process described in the Legal Aspects section provides a basis for achieving understanding and agreement between two organisations who desire to achieve transnational access.

The major barrier to transnational access will be legal, since countries have separate legal systems (which may be opaque: see DwB Deliverable 3.4).More specifically, the main issue is not so much that the countries have separate and different legal systems but that these legal frameworks do not provide any solution for suing someone across borders in case of breach of confidentiality. The concept of the Circle of Trust may allow looking at equivalences in penalties and transferring

responsibilities to the partner. Of course, it is still a long way to build a level of trust that enables transferring such high-level responsibilities.

Language and cultural differences may also be as many barriers but, as demonstrated with this project, such differences can be overcome. Technical arrangements may prohibit transnational access if, for example, the security standards of a given service are so high that they prevent it from hosting an external secure network connection. Even these types of restrictions can be overcome however, and simply requires the will of IT staff in both organisations to come to a mutually acceptable work plan.

Organisational issues may present a hurdle to transnational access if, for example, transnational access implied that a service's established set of routines had to be significantly altered to meet the requirements of a partner organisation, particularly if resources were required. However, as we have established in the 'equivalence' exercise, such services generally operate along similar operational frameworks, and it is unlikely that one service would operate very differently from another.

6. OUTLOOK

While deliverable 4.4 described the way to an agreement as a basis for access connections; deliverable 4.5 will focus on the technical implementation of the IAB-UKDA connection and show the findings of a PoC for a RAN carried out by WP4 partners. Finally deliverable 4.6 will focus on the research projects that will be using the newly-established connection between IAB and UKDA.

7. TEMPLATE OF MOU

The document below is a template Memorandum of Understanding based on the outcomes of the negotiations between IAB and UKDA in order to implement a two-way secure remote connection between both institutions, in order to allow researchers to access remotely each institutions' confidential microdata from a RDC in the other partner's home country.

MEMORANDUM OF UNDERSTANDING

Between

Federal Employment Agency (BA), Regensburger Straße 104, 90478 Nuremberg, represented by its Executive Board,

which in its turn is represented by the Director of the Institute for Employment Research of the Federal Employment Agency (IAB) Prof. Joachim Möller,

and

The UK Data Archive (UKDA)

on behalf of

The University of Essex

Wivenhoe Park

Colchester

Essex CO4 3SQ

hereinafter referred to collectively as the Parties

about

Reciprocal Provision of Safe Room Remote Desktop Access

TABLE OF CONTENTS

INTRODUCTION

This document contains a Memorandum of Understanding (MoU) that sets out the responsibilities of the Parties for a reciprocal arrangement to enable researchers to remotely access confidential/sensitive research data where Party A is responsible for storing the data, and Party B is responsible for securing the access point.

The MoU accepts equivalence between the policies, procedures and guidelines for management of, and access to, confidential/sensitive research data that are in place at IAB and the UKDA. It anticipates the goodwill of both parties for day to day interpretation and smooth operational administration. The Parties have shared knowledge of each other and have mutual trust in their management of data security and standard operating procedures.

Under this (MoU):

IAB is responsible as Party A for those data referenced in Appendix A, Schedule 1 and UKDA is responsible as Party B for those data referenced in Appendix A, Schedule 1 and;

UKDA is responsible as Party A for those data referenced in Appendix B, Schedule 2 and the IAB is responsible as Party B for those data referenced in Appendix B, Schedule 2.

A secured remote desktop connection allows researchers to work via a thin client (access device), that is located in a safe room at Party B; with the research data stored on Party A servers. The dataset never leaves the servers located in the facilities of Party A; only screen updates are transmitted to the thin client in the safe room hosted by Party B (a more detailed description can be found in Appendix C).

This MoU sets out Party A and Party B responsibilities in order to ensure secure access to the confidential/sensitive research data of both Parties. Additional documents (e.g. a non-technical overview of this type of data access system) can be found in the Appendices of this MoU.

THE MEMORANDUM OF UNDERSTANDING

It is hereby agreed between the parties:

Section 1: Definitions

(1) User

A researcher who has been authorised by formal agreement to work with, and access, Party A research data from within a Party B Safe Room.

(2) Secure Remote Desktop Access

Secured encrypted Remote Desktop Connection to a data-processing system held at Party A from an access device at a remote location at Party B. Only mouse and keyboard changes are transferred from B to A and only screen updates are transferred from A to B. The dataset stays at the servers of Party A.

(3) Safe Room

A secured room belonging to Party B used for providing access to confidential/sensitive research data of Party A. Entry is restricted and the setting is regularly monitored by Party B staff. The room is only provided with the access device(s) required to enable Secure Remote Desktop Connections to the distant servers. Only authorised members of staff and users are permitted to access the Safe Room.

(4) Workspace

The workspace is a desk in the safe room where a Thin Client is placed. The workspace is protected and the screen of the access device must not be easily observed by other users or unauthorized people.

(5) Thin Client

A Thin Client works as an access device. It is a computer terminal which only provides an interface (via keyboard, mouse, screen) to the user, and is configured to enable access only to the server of Party A where the data reside (the user will not be able to use the device to access anything else). All operations are carried out on the servers of Party A.

Section 2: Services of Parties

The Parties agree to undertake tasks to enable secured Remote Desktop Access to confidential/sensitive research data, from a Safe Room. The responsibilities are structured by topic. If a topic contains no tasks for the given Party, this is stated by the term “none”. The tasks are described, by Party, below.

Party A

1. Setup/Installation

    1.1. Party A shall provide Party B with a minimum of one Thin Client for the workspace provided in the Safe Room of Party B. The Thin Client will remain the property of Party A.
    1.2. Party A shall enable a Secure Remote Desktop Connection to designated servers located at Party A.

2. Application (User)

    2.1. All tasks concerning User applications to access Party A data are undertaken by Party A. A description of the application process can be found in Appendix D.

3. Account (User)

    3.1. Party A shall regularly provide Party B with information about authorised Users, including their names and institutions and the end dates of their projects. This information is required for access control.
    3.2. Party A shall provide Party B with an authentication code in order to establish the Secure Remote Desktop Connection to Party A servers.
    3.3. Party A shall provide each authorised User with personalized User credentials in order to access the data stored in a project area on the servers of Party A. Party A will make reasonable efforts to ensure that the User does not share their authentication credentials with anyone, including Party B staff.

4. Access control/monitoring (Safe Room)

    4.1. None.

5. Support

    5.1. Party A will take responsibility for user support for the research data, the accompanying documentation, the application process, and for information about statistical disclosure control issues.

6. Statistical disclosure control

    6.1. Dissemination of output files and disclosure control tasks will be undertaken by Party A.

7. Training

    7.1. Party A will inform Party B staff, on a need to know basis, about Party A data security, workflows, organisational structure and instructions for operating the Thin Client belonging to Party A in the Safe Room.
    7.2. Party A will provide training for Users where training is a requirement for access.

8. Maintenance

    8.1. Maintenance and technical service of the Thin Clients, including software, and the repair of the Thin Clients, will be undertaken by Party A or by contractors instructed by Party A as necessary.

9. Safe Room Security

    9.1. None.

Party B

1. Setup/Installation

    1.1. Party B will provide a workspace in a Safe Room.
    1.2. Party B will provide and enable a network point for connecting to the internet from the Safe Room, thus allowing a Secure Remote Desktop Access Connection between the Thin Client and the server hosted by Party A.
    1.3. Party B staff will assist with the establishment of the secure Safe Room and workspace environment from where the Secure Remote Desktop Access Connection is provided to Party A servers.

2. Application (User)

    2.1. None.

3. Account (User).

    3.1. None.

4. Access control/monitoring (Safe Room)

    4.1. Party B will schedule on-site visits to the Safe Room location at Party B for the User.
    4.2. Party B staff will verify the identity of Party A Users upon arrival at the Safe Room.
    4.3. Party B will ensure that only Party A Users have access to the designated Party A workspace.
    4.4. Party B staff will be provided with an authentication code in order to establish the secure Remote Desktop Connection to the Party A servers. This authentication code must be kept confidential by Party B staff and must not be shared with anyone, including the data user.
    4.5. Party B staff will monitor the User to ensure that no attempts to disclose the identity of data subjects occur. 'Attempts' includes but is not limited to: taking pictures, sharing

information with unauthorized persons, use of communication equipment within the safe room, manipulating hardware and software.

4.6. Party B shall notify Party A of any security events with regard to data privacy and data security.

5. Support

5.1. Party B staff will serve as a local contact for researchers interested in conducting research with Party A data at the Safe Room of Party B.

6. Statistical disclosure control

6.1. None.

7. Training

7.1. None.

8. Maintenance

8.1. Party B staff will assist Party A with the maintenance of the Secure Remote Desktop Connection between the Access Device (Thin Client) and the Party A servers.

8.2. Party B will inform Party A of any technical problems which occur.

9. Safe Room Security

9.1. Party B staff, in charge of the operation of the Safe Room, has to be informed about applicable laws and procedures. More information is provided in the Appendix E.

9.2. Party B permits employees of Party A to audit the activities in the Safe Room provided reasonable notice is given to Party B.

9.3. In addition to the employees of Party A, the appointed data privacy officers of Party A, and local authorities of Party A that are required to have right of control by the local law of Party A shall also be permitted to audit the activities in the Safe Room at any time, provided reasonable notice is given to Party B.

Section 3: Period of agreement and modification/termination

This agreement will become effective when signed by both Parties. Either Party has the option to terminate the agreement upon thirty (30) days advance written notice to the other. If terms of the MoU are violated the remote connection can be terminated immediately. Changes and amendments to this agreement shall be valid only if made in writing and agreed by both Parties.

Section 4: Available Research Data and Fees

The Parties' research data, made available under this MoU, are listed in Appendix A Schedule 1 and Appendix B Schedule 2. Any fees that arise are laid down in Appendix F.

Authorized Signatories:

For IAB:

_____ Date: _____

For UKDA:

_____ Date: _____

THIS CONCLUDES THE MEMORANDUM OF UNDERSTANDING BETWEEN THE PARTIES

APPENDIX A: Schedule 1

The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) provides researchers access to microdata in the fields of German social security and employment. It provides data on individuals, households and establishments, as well as data that comprise both establishment and personal information. The FDZ data originate from the notification process of the social security system, from internal procedures of the Federal Employment Agency and surveys which are conducted by the IAB.

IAB is responsible as Party A for, and will make available, those data listed at:

http://fdz.iab.de/en/FDZ_Overview_of_Data.aspx

APPENDIX B: Schedule 2

The UK Data Service in general provides a unified point of access to the extensive range of high quality social and economic data, including UK census data, government funded surveys, longitudinal studies, international macrodata, qualitative data and business microdata.

Under this MoU, the UK Data Service Secure Lab provides secure access to selected datasets that are especially detailed, sensitive or confidential.

UKDA is responsible as Party A for, and will make available, the following data:

SN 6676 Understanding Society

For details see:

http://discover.ukdataservice.ac.uk/catalogue/?sn=6676&type=Data%20catalogue

APPENDIX C: General Description of Safe Room Remote Desktop Services

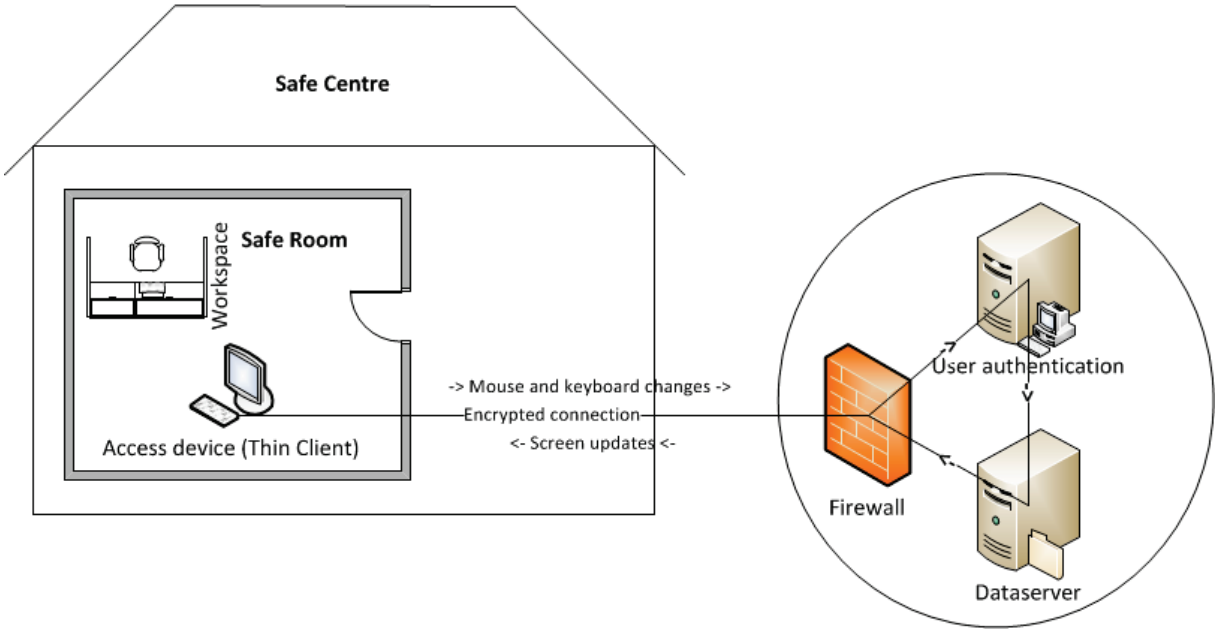
Access facilities (Research Data Centres, Population Centres etc.) exist to provide secure access to confidential microdata for research purposes. The provision of secure access facilities is about ensuring compliance with data and privacy protection regulations. Confidential microdata about individuals (persons, households, institutions), understood as information that can potentially be used to identify an individual, must be strictly regulated and controlled.

A secure method of accessing such data is via a secure remote desktop connection accessed through a 'safe room'. This approach offers a secure environment to access the confidential microdata. Datasets remain on the secure servers of the data provider (in location A), which are then accessed through a secure encrypted internet connection (from location B). Only screen updates, mouse and keyboard changes are transferred through the encrypted internet connection. Therefore no physical transfer of the confidential/sensitive dataset ever occurs: all browsing and analysis of data are undertaken remotely from location B, while the data remain on the secure servers in location A.

The safe room provides additional physical controls (e.g. control of researcher access and monitoring of the room by staff). Essentially, the safe room can be thought of as a physically secured room to protect the access device.

The setting described above requires the collaboration of a minimum of two partners. Party A assumes responsibility for the research data and Party B assumes responsibility for the access point. Responsibilities, tasks and workflows required of each Party are set out in this Memorandum of Understanding.

The research data reside in, and remain at, the secure premises of Party A and a legally binding agreement is in place between Party A and the user. Party B agrees to monitor the safe room setting and fulfil the required arrangements for permitting users to access the data. The arrangement is managed through a trust agreement which is set out in the Memorandum of Understanding (MoU) between the participating organisations.



The figure shows a schematic representation of the Safe Room Remote Desktop Service.

APPENDIX D: Application process

IAB and the UKDA are responsible for equivalent technical environments to give access to microdata for researchers. Both parties apply equivalent practices for user management and the release of outputs. Accordingly, both organisations apply similar and essential operating procedures when administering research project applications for access to their data. These procedures are designed to ensure that projects comply with the legal requirements associated with particular datasets and meet the goals of each organisation.

The processes and accompanying procedures for applications are very similar. A specific difference for UK Data Service is the concept of "fit and proper" researchers as defined in the Statistics and Registration Services Act 2007. Under the Act, a researcher is deemed 'Fit and Proper' when able to demonstrate, to the satisfaction of the National Statistician, that he/she:

- Has the appropriate knowledge and experience necessary for handling potentially disclosive personal information;
- Has provided satisfactory evidence supporting their application that illustrates their professionalism and technical competence to carry out the research proposal;
- Demonstrates a commitment to protecting and maintaining the confidentiality of the data during the creation of outputs and publications that arise during the proposal.

Hence, users of UK Data Service secure lab have to undergo training as part of the application process to enable them to use the confidential data legitimately. This training provides an opportunity for UKDA staff to be assured that the researcher is entitled to use data under the Statistics and Registration Services Act (SRSA). In contrast, IAB requires no such training: instead, a researcher's trustworthiness is deemed appropriate by its affiliation to a recognized research institution.

The following table describes the implementation of the essential stages of the application process at each institution:

UKDA	IAB
1. Applying for data access	
All researchers who request secure access to data from the UK Data Service must first become a registered user of the service, a process under which they must agree to conditions set out in the End User Licence. Then, on requesting access to a controlled study[4], the user will be asked to complete a more rigorous application process. This application process, for access to securely held data comprises three parts: • A **Project Proposal** in which the Principle Investigator (PI)[5], provides information about: the purpose of the project; the datasets applied for; a justification which includes reasons why other less confidential data sources available elsewhere are less suitable for the research; how the research will benefit the public; and details of any data matching requirements. This part of the process is similar to the requirements for users of IAB data under the Social code book (see Appendix E); • A **Researcher Application.** Every user named on the project, including the PI, must provide evidence demonstrating they are a 'fit and proper person', i.e. that they possess the statistical expertise necessary to handle confidential and complex microdata, and that they have professional experience of undertaking such analyses; • A signed **Declaration.** Every user named on the project must read and sign the Declaration, which summarises their responsibilities inherent in applying to, and accessing, confidential microdata.	Potential users of IAB data submit their applications for data access to the Research Data Centre of the BA. All users of data have to submit an application for data access in accordance with applicable German regulation, specifically section 75 Social Code Book X (Appendix E). This application comprises a project proposal which provides information about the purpose of the project, the needed datasets; a justification why especially this dataset is needed to answer the research question and details of any data matching requirements.

[4] Controlled data are data which may be identifiable and thus potentially disclosive.

[5] The Principle Investigator is the leading researcher who has responsibility for the work described in the application.

2. Application Approval Process	
When the UK Data Service receives an application from a PI, staff make a number of checks to ensure that all parts of the application have been completed. This includes validating the research proposal (to ensure the research aims to benefit public understanding) and confirming that the justification of the data chosen is adequate. The entire application (consisting of the research proposal, applications from each of the participating researchers and the Declarations) is sent to the data depositor for their formal approval of both the project and the individual researchers who will work on the project. When the UK Data Service has been notified of this decision, arrangements to set up user accounts and train users are made.	In order to review and approve an application for data access to IAB data the following procedures apply: The applications will be evaluated by either the project review system of the Research Data Centre of the BA or the IAB's formal review process. The IAB will review projects for disclosure risk, feasibility and the requirement for non-public data. The application must explicitly explain why the research project is related to the German social security system or to the field of labour market research. This is an absolute requirement for the approval of the application for data access as stipulated by section 75 Social Code Book X. It is a further requirement that the public interest of the research project is convincingly shown in the application for data access.
3. Formal project agreements	
Each user who intends to use controlled data must read and sign a User Agreement. This sets out the conditions of use and standards with which users must comply when accessing data through the Secure Lab and when requesting statistical outputs. This User Agreement must also be signed by a legal signatory at the user's institution. This is particularly important because it lists the sanctions that can be applied to a user and to their institution, in the event of a breach of the agreement. Once received by the UK Data Service, the User Agreement is only valid when the user accesses the Secure Lab from the named institution. The agreement covers multiple projects so avoids the need for a new Agreement for each. If a researcher moves to a new institution however, they must sign a new agreement which must also be signed by their new institution, before access to the Secure Lab can resume.	After the application has been approved by the German Federal Ministry for Labour and Social Affairs (Bundesministerium für Arbeit und Soziales), the BA represented by the IAB concludes a data access agreement with the data user's home institution. The researcher's institution must agree that researchers on the project will abide by all IAB confidentiality and disclosure analysis procedures and all relevant UKDA requirements. As a part of this agreement all users have to sign the *Guidelines for On-Site Data Use at the FDZ of the BA*, which summarizes the conditions of use in order to protect and maintain the confidentiality of the data during the whole process of data use. With signing the Pledge of Data Secrecy all users have to declare that they were made aware of were made aware of the obligations of data confidentiality according to section 5 of the Federal Data Protection Act (Bundesdatenschutzgesetz, BDSG) and § 78 of German Social Code, Book X - (Sozialgesetzbuch SGB X).

4. User training	
Users who access data through the UK Data Service Secure Lab must attend the *Secure Lab Training* course. Attendance is compulsory because users are made aware of a number of important issues associated with their use of the service, and the handling of confidential or sensitive microdata. Training includes awareness raising of: • the legal framework surrounding researcher access to the data, and the consequences of misuse of their access; • the principles of statistical disclosure control, to enable researchers to prepare statistical outputs that minimise the risk of a breach of confidentiality.	No user training required.

APPENDIX E: Relevant legislative texts

This appendix contains the texts from sections of the Data Protection Acts that apply to secure services at the IAB and the UKDA respectively. Each organisation will make their staff aware of the other's relevant text as appropriate.

For IAB:

Section 78a of the German Social Code, Book X (Sozialgesetzbuch, Zehntes Buch)

Section 78a: Technical and organizational measures

the bodies named in paragraph 35 of the German Social Code, Book I [Public and private bodies] processing personal data either on their own behalf or on behalf of others shall take the technical and organizational measures necessary to ensure the implementation of the provisions of this Act, in particular the requirements set out in the Annex to this Act. Measures shall be required only if the effort involved is reasonable in relation to the desired level of protection.

Annex (to paragraph 78a of the German Social Code, Book X)

Where social data are processed or used automatically, the internal organization of authorities or enterprises are to be arranged in such a way that it meets the specific requirements of data protection. In particular, measures suited to the type of social data or of social data to be protected shall be taken,

1. to prevent unauthorized persons from gaining access to data processing systems with which social data are processed or used (access control),
2. to prevent data processing systems from being used without authorization (access control),
3. to ensure that persons entitled to use a data processing system have access only to the data to which they have a right of access, and that social data cannot be read, copied, modified or removed without authorization in the course of processing or use and after storage (access control),
4. to ensure that social data cannot be read, copied, modified or removed without authorization during electronic transmission or transport, and that it is possible to check and establish to which bodies the transfer of social data by means of data transmission facilities is envisaged (transmission control),
5. to ensure that it is possible to check and establish whether and by whom social data have been input into data processing systems, modified or removed (input control),
6. to ensure that, in the case of commissioned processing of social data, the data are processed strictly in accordance with the instructions of the principal (job control),
7. to ensure that social data are protected from accidental destruction or loss (availability control),
8. to ensure that data collected for different purposes can be processed separately.

One measure in accordance with the second sentence Nos. 2 to 4 is in particular the use of the latest encryption procedures.

Section 75 of the German Social Code, Book X (Sozialgesetzbuch, Zehntes Buch)

Transmission of social data for research and planning

(1) The transmission of social data is permissible as long as it is required for a specific project

1. concerning scientific research in the field of social benefits or employment research
2. concerning planning in the field of social benefits conducted by an official body within the scope of its tasks

and the data subject's interests warranting protection are not impaired or the public interest in the research or planning outweighs to a considerable extent the data subject's interest in privacy. Transmitting data without the data subject's consent is not permissible in so far as it is reasonable to expect the data subject's consent to be sought in accordance with § 67b . Information on family and first names, address, phone number as well as structural measures that are mandatory for the launch of a project by section 1 may be sent for interviews without consent.

(2) The data transmission requires prior authorization by the highest Federal or regional authority responsible for the field from which the data originate. Authorization may only be refused with regard to the observance of social data protection if the requirements of paragraph 1 are not met. The authorization shall stipulate

1. the third party to which the data are transmitted,
2. the type of social data to be transmitted and the group of data subjects concerned,
3. the scientific research or planning for which the transmitted social data may be used and
4. the date until which the transmitted social data may be stored, and, even without specific reference, is subject to the subsequent insertion, change or amendment of a condition.

(3) If the transmission of data to non-official bodies is authorized, the authority granting the authorization shall impose conditions to ensure that the bounds laid down for the authorization by paragraph 1 are taken into consideration and that the data are only stored, changed or used for the purpose for which they were transmitted.

(4) If the third party to which data are transmitted is a non-official body, § 38 of the Federal Data Protection Act applies, subject to the proviso that the monitoring may also take place when the data are not automated or are not processed or used in non-automated files.

For the UKDA

1. UK Data Protection Act 1998

Under Section 33 of the Act, personal data may be used for "research purposes" including statistical or historical purposes provided:

"(a) that the data are not processed to support measures or decisions with respect to particular individuals, and

"(b) that the data are not processed in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject."

Furthermore, the use of personal data which are processed only for research purposes is exempt from Section 7 of the Act (which determines an individual's right to know about what personal data are being processed and for what purpose) provided:

"*(a) they are processed in compliance with the relevant conditions, and*
(b) the results of the research or any resulting statistics are not made available in a form which identifies data subjects or any of them."

2. Section 39(4) of the Statistics and Registration Service Act 2007 (SRSA) applies to researchers and their projects. In the context of the SRSA, a researcher is deemed 'Fit and Proper' when....

The researcher is able to demonstrate, to the satisfaction of the National Statistician, that he/she:

1. Has the appropriate knowledge and experience necessary for handling potentially disclosive personal information;

2. Has provided satisfactory evidence supporting their application that illustrates their professionalism and technical competence to carry out the research proposal;

3. Demonstrates a commitment to protecting and maintaining the confidentiality of the data during the creation of outputs and publications that arise during the proposal.

A research project is deemed suitable when....

The research, in the opinion of the National Statistician, serves one of the following public benefits:

1. Supporting the formulation and development of public policy or public service delivery.

2. Forms part of the programme of research covered by the National Data Strategy or otherwise supported directly or indirectly by the Economic and Social Research Council.

3. Supports an obligation of public law (e.g. Local Development Plans)

4. Explores new statistical methods that can be used to produce statistics that serve the public good.

APPENDIX F: Fees

1. Neither Party A nor Party B will make charges on the other in relation to this service.
2. Neither Party charges researchers for use of their secure service.
3. Neither Party will be responsible for any costs resulting from the need to attend training events that are a prerequisite for access or continued access to data.
4. Neither Party will be responsible for any costs incurred by researchers in travelling between their institution and the safe centre.
5. Both Parties will endeavour to identify award schemes to cover the costs of 3 and 4 above and will support applications for awards to cover these costs where projects are deemed to be appropriate.
6. Both Parties reserve the right to make charges in line with local charging policy.