Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)

FDZ-Methodenreport

08/2012
EN

Methodological aspects of labour market data

Technical and organisational measures for remote access to the micro data of the Research Data Centre of the Federal Employment Agency

Joerg Heining,
Stefan Bender

Bundesagentur für Arbeit

Technical and organisational measures for remote access to the micro data of the Research Data Centre of the Federal Employment Agency

Joerg Heining, Stefan Bender

Die FDZ-Methodenreporte befassen sich mit den methodischen Aspekten der Daten des FDZ und helfen somit Nutzerinnen und Nutzern bei der Analyse der Daten. Nutzerinnen und Nutzer können hierzu in dieser Reihe zitationsfähig publizieren und stellen sich der öffentlichen Diskussion.

FDZ-Methodenreporte (FDZ method reports) deal with the methodical aspects of FDZ data and thus help users in the analysis of data. In addition, through this series users can publicise their results in a manner which is citable thus presenting them for public discussion.

Contents

Zusammenfassung

Das Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg bietet einen Fernzugriff auf datenschutzrechtlich sensible Mikrodaten an. Datennutzerinnen und -nutzer können von den Forschungsdatenzentren der Statistischen Ämter der Länder an den Standorten Berlin, Bremen, Düsseldorf, Dresden sowie an der Hochschule der Bundesagentur für Arbeit in Mannheim auf die Daten des FDZ zugreifen. Zusätzlich besteht diese Möglichkeit auch am Institute for Social Research (ISR) der University of Michigan in Ann Arbor, MI, USA. Dieser Methodenreport beschreibt die für diesen Fernzugriff notwendigen technischen und organisatorischen Maßnahmen zur Einhaltung der datenschutzrechtlichen Vorgaben.

Abstract

The Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB) in Nuremberg provides remote access to confidential micro data for the first time. Data users can access FDZ data from the research data centres of the statistical offices of the Länder at their sites in Berlin, Bremen, Düsseldorf and Dresden, and at the University of Applied Labour Studies of the German Federal Employment Agency in Mannheim. Additionally, this possibility also exists at the Institute for Social Research (ISR) of the University of Michigan in Ann Arbor, MI, USA. This report describes the technical and organisational measures necessary for ensuring data confidentiality.

Keywords: remote access, data protection

The authors like to thank David Schiller and Peter Jacobebbinghaus.

Bundesministerium für Bildung und Forschung

DWB Data without Boundaries

SEVENTH FRAMEWORK PROGRAMME

The projects underlying this Methodenreport received financial support from the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) under the funding reference number 01UW1002, and funds for the project “Data without Boundaries” in the context of the Seventh Framework Programme (FP7/2007-2013) of the European Union under the funding reference number 262608.

1 Introduction

The Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB) in Nuremberg provides remote access to confidential micro data. Since October 2011 data users have been able to access FDZ data from the research data centres of the statistical offices of the Länder at their sites in Berlin, Bremen, Düsseldorf and Dresden, and at the University of Applied Labour Studies of the German Federal Employment Agency in Mannheim. Additionally, this possibility also exists at the Institute for Social Research (ISR) of the University of Michigan in Ann Arbor, MI, USA. There are plans to expand this possibility to further locations both in other European countries (in the context of the Data without Boundaries (DwB) project, www.dwbproject.org, which is supported by the European Union) and in North America.

In the past, in order to access the FDZ micro data in the context of research visits, data users had to travel to Nuremberg. Within the framework of the "Research-Data-Centre in Research-Data-Centre" project (PFiF), which is funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung - BMBF) and assessed by the German Data Forum (Rat für Sozial- und Wirtschaftsdaten - RatSWD), the possibility of remote access was implemented. The basic idea here is that researchers can access the FDZ data via a secure Internet connection (so-called remote access), with the data themselves remaining in Nuremberg. After concluding a data use agreement with the FDZ, researchers can log onto a server in the separate guest network of the FDZ using a so-called thin client computer from currently six locations. The actual data processing is performed on this server, the thin client computers serve only to create the connection. The connection between the thin client computer and the server in Nuremberg is encrypted using the Access Gateway software developed by Citrix.

This Methodenreport describes the technical implementation and the organisational and technical measures that are necessary to ensure compliance with the data protection regulations for such remote access. First, the basic idea of such remote access is outlined before the formal criteria, i.e. application procedure for data access and data use agreement, are examined. In Section 4 the technical implementation is explained, while Section 5 describes the technical and organisational measures in detail. Section 6 closes by looking at future developments.

2 Basic idea

The basic idea in implementing remote access is to permit access to the FDZ micro data from the premises of another RDC (different institution and location). The data are accessed in a similar way to the on-site use at the FDZ of the BA at IAB that has been possible to date. The only difference is that the guest researcher's room where the researcher sits at the keyboard and computer screen is not at the FDZ of the BA at IAB in Nuremberg but at a different RDC (guest-RDC). The data are accessed from dedicated workstations at the guest-RDC.

For this, the same security criteria must be fulfilled at the guest-RDC as apply at the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). The data are accessed via a secure data line. What is important here is that the data are neither processed nor stored at the guest-RDC. This is done, as previously, on a secure server in Nuremberg. All that is done at the guest-RDC is establishing the connection to the data processing systems in Nuremberg and ensuring admission control (see Figure 1: Basic idea). The processing of requests for data access and the data use agreements, the administration of user details, the administration of the FDZ guest network and output control remain the responsibility of the FDZ of the BA at IAB in Nuremberg.

There are differences between Germany and the USA with regard to the supervision concept of the sites. In Germany a clearly defined small group of staff members at the particular site is responsible for monitoring data access. These members of staff do not gain access to the FDZ data themselves. They only perform the technical and organisational measures necessary to ensure data confidentiality. At the site in Ann Arbor, on the other hand, a member of staff from the FDZ of the BA currently works on-site.

With regard to the level of anonymisation of the data there is no difference between on-site use in Nuremberg or at one of the external sites. Both in Nuremberg and at the sites of the statistical offices of the Länder and the University of Applied Labour Studies of the German Federal Employment Agency, authorised data users can access weakly anonymous data. This does not apply at the Ann Arbor site, however. There the datasets are rendered anonymous by the FDZ staff so that no weakly anonymous data are transmitted to the USA in the context of this remote access.

One key element of ensuring data confidentiality at the FDZ of the BA is that only absolutely anonymous results and files leave the separate area of the FDZ guest network. By checking the output files created by the user, the staff of the FDZ of the BA ensures that no details (e.g. establishments, households, individuals) can be identified. These principles also apply for data use at the external sites. Also when data are accessed outside of the Nuremberg site, the data user receives only absolutely anonymous output from the FDZ of the BA. The technical implementation (see Sections 4 and 5) means that it is not possible for the data user to remove data or output files him- or herself at the external site.

Besides the data use agreements concluded with external data users, contractual agreements between the FDZ of the BA and the guest-RDC or the institutions of the guest-RDC regulate the rights and obligations of these external sites.

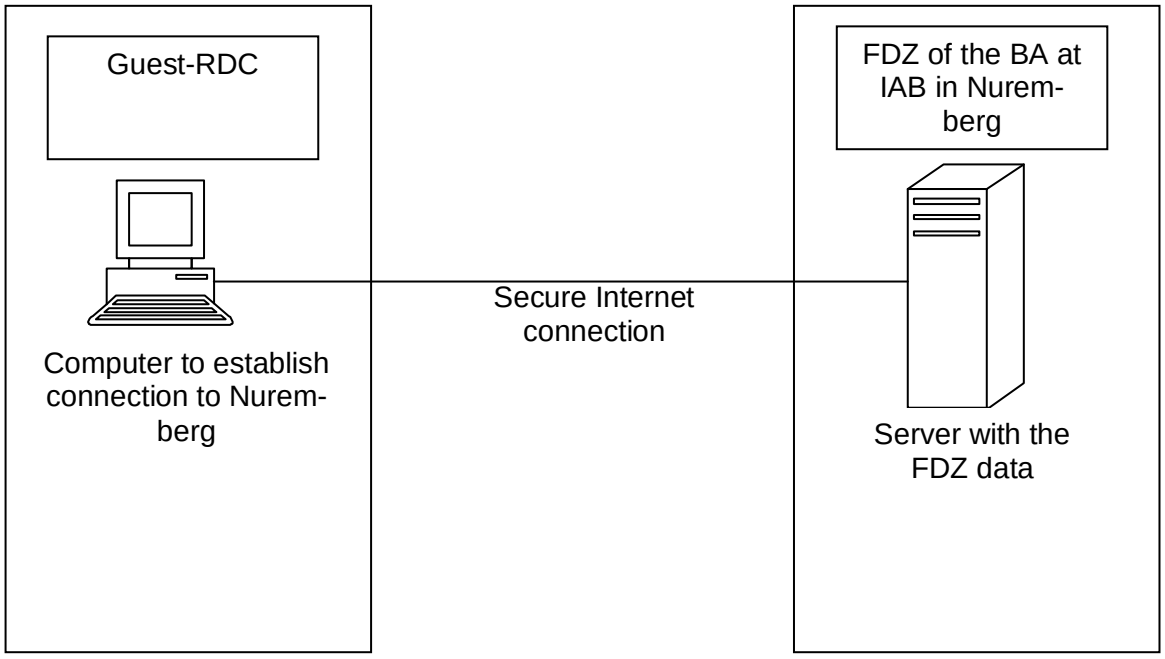

Figure 1: Basic idea

3 Application for data access and data use agreement

At the Research Data Centre (FDZ) of the Federal Employment Agency (BA) all external researchers continue to submit an application for data access in accordance with § 75 of Social Code Book Ten (SGB X). After this has been checked by the staff of the FDZ in Nuremberg it is passed on to the legal department of the IAB and from there to the Federal Ministry of Labour and Social Affairs (Bundesministerium für Arbeit und Soziales – BMAS). After the research project has been approved by the BMAS, the FDZ of the BA concludes a data use agreement with the data user's institution, in which the data user undertakes to comply with the data protection regulations recorded in the agreement and to bear the consequences stipulated in the data use agreement and in German law if the agreement is breached. Furthermore, all data users who do not have an employment relationship in the public service are additionally sworn to observe the German data protection legislation.

In addition to the title and a description of the project, the duration and the data requirements, the data use agreement explicitly defines rules of conduct for working with the data of the IAB / of the FDZ BA at IAB. These rules of conduct apply at all of the sites named above.

The data use agreements concluded between the FDZ and the external data user are "irrespective of location" and permit access to data both in Nuremberg and at one of the sites named above. However, it must be taken into account that there are additional access requirements for data access in Ann Arbor besides a reviewed request, approval by the Federal Ministry of Labour and Social Affairs and the existence of a valid data use agreement, such as the approval of the project by an Institutional Review Board (this is equivalent to an ethics commission). This is valid legal practice in the USA with regard to access to and the

use of micro data at the individual level. If a data user does not meet these requirements, data access is not possible in the USA. However, this does not affect data access at one of the German sites or in Nuremberg.

4 Technical implementation

4.1 Basic concept

The technical implementation of the remote data access occurs via so-called thin client computers (see below). At each of the sites named above there is at least one of these devices which allow the user at an external site to log onto the FDZ guest network in Nuremberg via a secure Internet connection. The data are not processed on the thin client computer at the external site but on a server in the separate FDZ guest network.

At all of the sites the thin client computers are situated behind the firewall of the particular site. The servers in Nuremberg are also protected by a corresponding firewall.

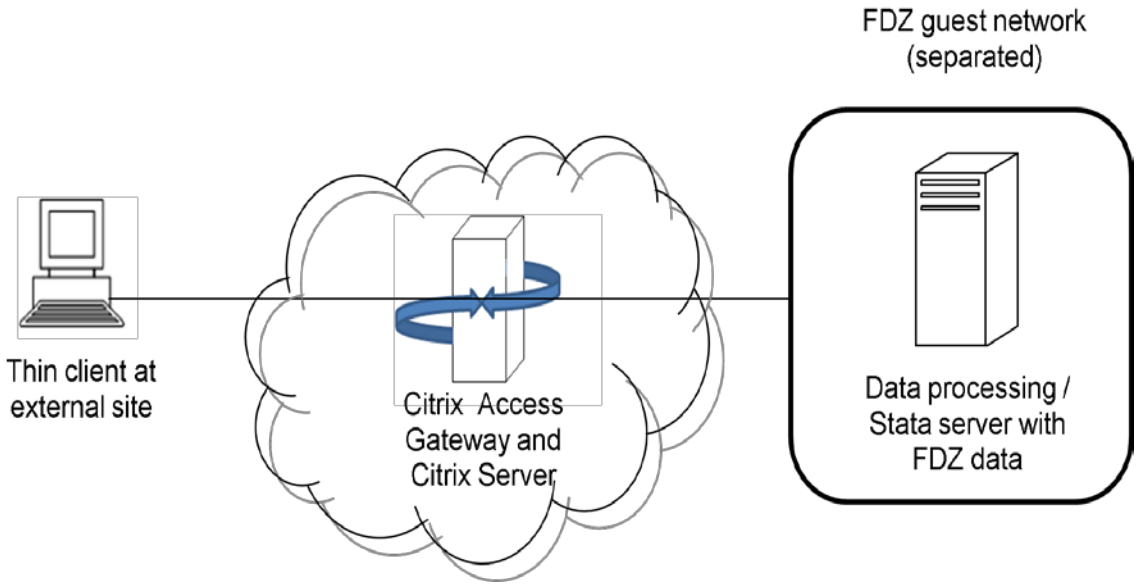

Figure 2: Technical implementation of remote access to the FDZ data (very simplified and schematic representation)

4.2 Thin client

A thin client computer is a computer that, unlike conventional computers, is equipped with only a small amount of hardware. A thin client is merely the user interface to a server. The actual communication, data processing and the storage of data are performed on the server.

The thin clients of this solution have no connection options at all for removable media (USB stick, external hard disc etc.) or for external peripheral devices (printers, external optical

drives etc.). It is therefore not possible for an external data user to install software or files on the thin client or to remove them.

Furthermore, the software available to an external data user on the thin client computer is very limited. Only Windows Internet Explorer is available, which the data user needs in order to establish a connection to the server in Nuremberg. On opening the Internet Explorer the user reaches a pre-set web page from which the connection is established. At no time can the data user switch to a different web page or change the pre-set page. No other software is available to the data user on the thin client.

The thin client computers were basically configured in such a way that as long as there is no connection to the server, the data user on the thin client can only open Windows Internet Explorer. Other functions are either not available or can be invoked by the user but any modifications carried out have no effect.

4.3 Citrix Access Gateway and Citrix server

The thin client computer at the external site is connected with a server in the FDZ guest network via the Access Gateway software and a corresponding Citrix server which was procured by the Federal Employment Agency. With this an encrypted connection is established between the thin client and the server in the FDZ guest network via a public network, which makes reliable data transmission and secure remote access possible. The contents of communication between the FDZ server and the thin client are encrypted and the participating communication partners, FDZ server and thin client, have to authenticate themselves to each other. The (temporary) connection established in this way meets the highest security standards. Communication between the server and the thin client can therefore not be read by third parties, i.e. by potential data intruders. Citrix is used throughout the world, also, for example, by other RDCs abroad (e.g. in the Netherlands or Denmark) as well as by banks.

In order to establish a secure connection between the thin client computer and the server in the FDZ guest network, the data user at the external site opens Windows Internet Explorer on the thin client. The pre-set Internet address leads the user to the Citrix input mask in which he then enters his personalised user name and the password for the project.

The connection set-up is additionally protected by a further password. This is known only to the member of staff supervising the data access on-site and is not passed on to the external data user (see Section 4.2). The connection password is changed at regular intervals.

4.4 FDZ guest network

The actual processing of data is performed in the guest network of the FDZ of the BA. The FDZ guest network is a separate network consisting of several servers. Apart from a few ports that are activated for maintenance work by the IT systems department of the BA, there is no connection between the BA Intranet or the outside world (Internet) and the FDZ guest network. Access to the FDZ guest network from outside is only possible via the thin clients at

the sites mentioned above. These sites have to authenticate themselves when logging onto the FDZ guest network.

The guest network was set up by the experts of the IT systems department of the BA and the IT Services and Information Management department of the IAB. It meets the valid IT security standards of the BA. The legal regulations in accordance with § 78a Social Code Book X (SGB X) and the Annex to § 78a SGB X with regard to access, input and dissemination control etc. are observed in the FDZ guest network (see Section 4).

5 Organisational and technical measures

After describing the basic idea and the technical implementation of this remote access, the paper now turns to the organisational implementation. The description is orientated towards the technical and organisational measures for access to personal data which are demanded in the Annex to § 78a SGB X.

5.1 Admission control

As explained above, the actual processing of the data is performed on the servers of the FDZ guest network. The data are also stored on these servers. The servers are housed in the BA computer centres. Here the strictest security standards apply with regard to structural and organisational measures in order to prevent unauthorised persons from gaining admission to the server room.

The FDZ regularly provides the external sites in Germany with lists of the names of the data users that have concluded a valid data use agreement for on-site use with the FDZ of the BA at IAB. For this a signed data use agreement of the FDZ must exist which explicitly includes data access in the context of research visits to an external site.

For a research visit, the data user arranges an appointment with the members of staff responsible at the particular site. On the date arranged the staff of the RDC of the statistical office or the University of Applied Labour Studies check the data user's identity by examining a valid identification document (passport or ID card) and comparing it with the list of names transmitted by the FDZ.

At the Ann Arbor site a data user also only gains admission to the thin client computers after arranging an appointment with the member of staff of the FDZ of the BA who is on-site. The FDZ staff member verifies the data user's identity by means of a valid identification document. It is not necessary to transmit lists of names of authorised data users to the member of staff in Ann Arbor as this person has access to the user database of the FDZ of the BA in which this information is stored.

At all of the sites external data users can only gain admission to the thin client computers during certain opening hours. At each site, admission is additionally secured, for example by a gate. The rooms with the thin client computers are always kept locked and are only opened for an authorised user to access the data. In these rooms the thin client computers are addi-

tionally protected against theft by means of Kensington locks. In principle, the thin client computers may be used by the data user only when the supervisor responsible for the site is present.

5.2 Access control

Irrespective of which site is used, a project directory is created on the servers in the FDZ guest network for every data user with a valid data use agreement entitling them to on-site access. The data specified in the agreement are filed in this directory. The data user can only access this project directory if he logs onto the FDZ guest network using his user name and personal password. It is not possible for the data user to access other project or data directories.

Both user name and password are project-specific and person-specific and known only to the data user. The data user is strictly prohibited by contract from disclosing his user name and password. Both the user ID and the password are activated only for the contractually agreed project duration. After the end of the project it is no longer possible for the user to access the project directory.

In the case of the external sites, access to the data processing systems, i.e. to the servers in the FDZ guest network, is additionally protected. First, the staff at the external sites does not know either the data user's user name or password. The access details are passed on to the data user from the FDZ of the Federal Employment Agency by telephone after verification of his identity by the on-site supervisors.

Besides that, the connection between the thin client computer and the server in Nuremberg is protected by an additional password. This password is known only to the on-site supervisors, but not to the user. The connection password is changed at regular intervals. When the thin client computer is inactive, the link to the server in Nuremberg is disconnected and can only be re-activated by entering the connection password again.

This remote access is therefore secured in two respects: first, by means of project-specific and person-specific user IDs and passwords, and second, as a result of the connection password. Both of the elements are essential for access to data at one of the external sites.

5.3 Data access control

See Section 4.2. The servers of the FDZ guest network in Nuremberg are not connected with the Internet. No access to these servers is possible via the Internet apart from using the specially configured thin client computers, the Citrix Access Gateway and the Citrix server.

5.4 Dissemination control

The thin client computers do not permit any removable media or peripheral devices to be connected. It is therefore not possible for the data user at the site to remove file extracts or data extracts via the thin clients. Using the Citrix Access Gateway solution a secure connection is established between the thin client and the server in the FDZ guest network in Nurem-

berg which is not open to inspection by third parties. In this respect, no data can be removed or disclosed in this way, either.

The data user receives only log files and output files that have been checked by staff of the FDZ of the BA and are absolutely anonymous. It is therefore not possible to identify individual cases with these log files and output files. The criteria applied for this by the FDZ are described in Hochfellner et al. (2012).

Furthermore, the on-site supervisors also ensure that data users at the external site make no written copies from the screen of the thin client computer. The data users are prohibited from communicating with anyone outside the rooms containing the thin client computers while they are there. The same applies for the use of mobile phones, cameras, mobile computers etc. Compliance with these regulations is also monitored and ensured by the on-site supervisors. The same rules therefore apply as at the Nuremberg site.

5.5 Input control

Every case of access to the FDZ guest network is recorded. If any irregularities are ascertained for a user, his or her access authorisation can be withdrawn immediately and the penalties stipulated in the contract implemented.

5.6 Separate processing of social data collected for different purposes

If a data user applies for access to several data files provided by the FDZ for a research project, a separate directory is created in the FDZ guest network for each data product. The data user can access each of these directories associated with his project, but it is not possible for him to shift data files or extracts of data files between the directories. This prevents any increase in the risk of individual entities being re-identified as a result of the merging of different data products. This measure also guarantees the separate processing of social data collected for different purposes.

6 Increasing the sites

The FDZ of the BA is planning to increase the number of external sites in the USA in the near future. Following the example of the sites set up so far, access to FDZ data is also to be provided in future at Cornell University in Ithaca, NY and the University of California at Berkeley. In a subsequent step, additional sites are to be set up in Europe, too, in the context of the Data without Boundaries project.

At the two new sites in the USA, too, the same rules and requirements apply for data access applications. However, the data users may have to meet additional requirements, as they do at the University of Michigan, in order to take into account the data protection regulations valid in the USA.

With regard to the technical implementation there are no differences to the existing sites. In contrast to the site in Ann Arbor, there will not be any employees of the FDZ of the BA permanently on-site at the University of California and Cornell University. Staff members of the

particular guest-RDC will be responsible for admission control and for checking the data user's identity at these sites. These RDC employees are to be sworn to observe German data protection legislation like members of the public service in Germany. Furthermore, the RDC staff responsible at Berkeley and Cornell will be trained by the FDZ with regard to the data protection regulations in Germany and the procedures at the FDZ. In addition, members of staff of the FDZ of the BA at IAB will visit the two sites regularly to monitor and ensure compliance with all the technical and organisational measures.

As is the case at the external sites in Germany, the FDZ will transmit lists of the names of authorised data users to the people responsible at the particular site. This ensures that only authorised data users gain access to the thin client computers in Berkeley and Cornell, too. Otherwise, the technical implementation and the organisational and technical measures aimed at ensuring data confidentiality will be identical to those at the present external sites.

At the European sites under consideration, too, the technical and organisational implementation will essentially be equivalent to the measures at the current sites. It is not yet foreseeable whether additional requirements will be prescribed with regard to the application procedure for data access or the supervision concepts at the sites.

References

Bender, Stefan; Heining, Jörg (2011): The Research-Data-Centre in Research-Data-Centre approach: A first step towards decentralised international data sharing. In: IASSIST Quarterly, Vol. 35, No. 3, S. 10-16.

Hochfellner, Daniela; Müller, Dana; Schmucker, Alexandra; Roß, Elisabeth (2012): Datenschutz am Forschungsdatenzentrum. (FDZ Methodenreport, 06/2012), Nürnberg, 27 S.

Imprint

FDZ-Methodenreport 08/2012

Publisher
The Research Data Centre (FDZ)
of the Federal Employment Agency
in the Institute for Employment Research
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff
Stefan Bender, Dagmar Theune

Technical production
Dagmar Theune

Download
http://doku.iab.de/fdz/reporte/2012/MR_08-12_EN.pdf

Internet
http://fdz.iab.de/

Corresponding author:

Dr. Joerg Heining,
Research Data Centre (FDZ)
Regensburger Str. 104
D - 90478 Nuremberg
Phone: +49 (0)911-179-5392
Email: joerg.heining@iab.de