

# The State of the Art of Remote Access to Confidential Microdata in Europe

David Schiller

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

IASSIST, Cologne  
2013-05-29



**DwB**  
Data without Boundaries

# Outline

- 1 Introduction
- 2 What is Remote Access
- 3 Survey Participants
- 4 Findings of the Survey
- 5 What has been achieved, what is missing?
- 6 Next steps
- 7 Summary and Outlook

## Data without Boundaries (DwB)

- Funded by the European Commission FP7 program
- 4-year project (2011-2015) with 29 partners (NSIs, DAs, Universities) from all over Europe
- Main goals:
  - Legal issues and accreditation
  - Servicing the use of OS data
  - Metadata standards and interoperability
  - Building a single point of access
  - Providing a flexible infrastructure for accessing confidential data
- 12 work packages
- More information: **[www.dwbproject.org](http://www.dwbproject.org)**

## Work Package 4 - Improving Access to OS microdata

### **Ease European research by enhancing data access from a technical point of view**

#### Three tasks (two reports; one pilot):

- 1 A survey about the state of the art of remote access in Europe
- 2 A feasibility study about the transition of separate Remote Access Centres into a Remote Access Network (RAN)
- 3 Pilot (proof of concept) for the implementation of a Remote Access Network

#### Partners:

CBS (Netherlands), CNRS (France), Destatis (Germany), IAB (Germany), INSEE (France), NSD (Norway), ONS (UK), UKDA (UK)

## Work Package 4 - Improving Access to OS microdata

**Ease European research by  
enhancing data access from a technical point of view**

### **Three tasks (two reports; one pilot):**

- 1 A survey about the state of the art of remote access in Europe
- 2 A feasibility study about the transition of separate Remote Access Centres into a Remote Access Network (RAN)
- 3 Pilot (proof of concept) for the implementation of a Remote Access Network

### **Partners:**

CBS (Netherlands), CNRS (France), Destatis (Germany), IAB (Germany), INSEE (France), NSD (Norway), ONS (UK), UKDA (UK)

## Work Package 4 - Improving Access to OS microdata

**Ease European research by  
enhancing data access from a technical point of view**

### **Three tasks (two reports; one pilot):**

- 1** A survey about the state of the art of remote access in Europe
- 2** A feasibility study about the transition of separate Remote Access Centres into a Remote Access Network (RAN)
- 3** Pilot (proof of concept) for the implementation of a Remote Access Network

### **Partners:**

CBS (Netherlands), CNRS (France), Destatis (Germany), IAB (Germany), INSEE (France), NSD (Norway), ONS (UK), UKDA (UK)

## Work Package 4 - Improving Access to OS microdata

**Ease European research by  
enhancing data access from a technical point of view**

### **Three tasks (two reports; one pilot):**

- 1** A survey about the state of the art of remote access in Europe
- 2** A feasibility study about the transition of separate Remote Access Centres into a Remote Access Network (RAN)
- 3** Pilot (proof of concept) for the implementation of a Remote Access Network

### **Partners:**

CBS (Netherlands), CNRS (France), Destatis (Germany), IAB (Germany), INSEE (France), NSD (Norway), ONS (UK), UKDA (UK)

## Work Package 4 - Improving Access to OS microdata

**Ease European research by  
enhancing data access from a technical point of view**

### **Three tasks (two reports; one pilot):**

- 1** A survey about the state of the art of remote access in Europe
- 2** A feasibility study about the transition of separate Remote Access Centres into a Remote Access Network (RAN)
- 3** Pilot (proof of concept) for the implementation of a Remote Access Network

### **Partners:**

CBS (Netherlands), CNRS (France), Destatis (Germany), IAB (Germany), INSEE (France), NSD (Norway), ONS (UK), UKDA (UK)



## Work Package 4 - Improving Access to OS microdata

**Ease European research by  
enhancing data access from a technical point of view**

### **Three tasks (two reports; one pilot):**

- 1** A survey about the state of the art of remote access in Europe
- 2** A feasibility study about the transition of separate Remote Access Centres into a Remote Access Network (RAN)
- 3** Pilot (proof of concept) for the implementation of a Remote Access Network

### **Partners:**

CBS (Netherlands), CNRS (France), Destatis (Germany), IAB (Germany), INSEE (France), NSD (Norway), ONS (UK), UKDA (UK)

## Work Package 4 - Improving Access to OS microdata

**Ease European research by  
enhancing data access from a technical point of view**

**Three tasks (two reports; one pilot):**

- 1 A survey about the state of the art of remote access in Europe**
- 2** A feasibility study about the transition of separate Remote Access Centres into a Remote Access Network (RAN)
- 3** Pilot (proof of concept) for the implementation of a Remote Access Network

**Partners:**

CBS (Netherlands), CNRS (France), Destatis (Germany), IAB (Germany), INSEE (France), NSD (Norway), ONS (UK), UKDA (UK)

## What is Remote Access

**Within the DwB project's WP4 Remote Access (RA) is defined like that:**

- 1 User can access research data remotely
- 2 The network connection is secured
- 3 Data stays (behind firewalls) on the servers of the data provider
- 4 Only a image (live-stream) is forwarded and provided on the desktop
- 5 We talk about access to confidential microdata
- 6 The user can see (live) the data s/he is working with

# What is Remote Access

**Within the DwB project's WP4 Remote Access (RA) is defined like that:**

- 1** User can access research data remotely
- 2 The network connection is secured
- 3 Data stays (behind firewalls) on the servers of the data provider
- 4 Only a image (live-stream) is forwarded and provided on the desktop
- 5 We talk about access to confidential microdata
- 6 The user can see (live) the data s/he is working with

# What is Remote Access

**Within the DwB project's WP4 Remote Access (RA) is defined like that:**

- 1** User can access research data remotely
- 2** The network connection is secured
- 3 Data stays (behind firewalls) on the servers of the data provider
- 4 Only a image (live-stream) is forwarded and provided on the desktop
- 5 We talk about access to confidential microdata
- 6 The user can see (live) the data s/he is working with

# What is Remote Access

**Within the DwB project's WP4 Remote Access (RA) is defined like that:**

- 1** User can access research data remotely
- 2** The network connection is secured
- 3** Data stays (behind firewalls) on the servers of the data provider
- 4** Only a image (live-stream) is forwarded and provided on the desktop
- 5** We talk about access to confidential microdata
- 6** The user can see (live) the data s/he is working with

# What is Remote Access

**Within the DwB project's WP4 Remote Access (RA) is defined like that:**

- 1** User can access research data remotely
- 2** The network connection is secured
- 3** Data stays (behind firewalls) on the servers of the data provider
- 4** Only a image (live-stream) is forwarded and provided on the desktop
- 5** We talk about access to confidential microdata
- 6** The user can see (live) the data s/he is working with

# What is Remote Access

**Within the DwB project's WP4 Remote Access (RA) is defined like that:**

- 1** User can access research data remotely
- 2** The network connection is secured
- 3** Data stays (behind firewalls) on the servers of the data provider
- 4** Only a image (live-stream) is forwarded and provided on the desktop
- 5** We talk about access to confidential microdata
- 6** The user can see (live) the data s/he is working with



# What is Remote Access

**Within the DwB project's WP4 Remote Access (RA) is defined like that:**

- 1** User can access research data remotely
- 2** The network connection is secured
- 3** Data stays (behind firewalls) on the servers of the data provider
- 4** Only a image (live-stream) is forwarded and provided on the desktop
- 5** We talk about access to confidential microdata
- 6** The user can see (live) the data s/he is working with

## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany

## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany

## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany

## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany

## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany

## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany

## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany



## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany

## The survey

- Questionnaire with focus on the technical part of RA solutions prepared by CASD, France
- Only NSIs and CESSDA archives were surveyed
- Conducted November 2011 to February 2012
- Eight RA providers replied (6 NSIs, 1 NSA, 1 Archive):
  - 1 Statistics Denmark
  - 2 Office for National Statistics, UK
  - 3 Statistics Sweden
  - 4 Statistics Slovenia
  - 5 Statistics Netherlands
  - 6 UK Data Archive - Secure Data Service
  - 7 National Institute for Statistics and Economic Studies, France (CASD)
  - 8 Institute for Employment Research, Germany

# Main Findings of the Survey I

- Access points
  - Anywhere
  - Institution
  - Safe room
- International access
  - In general possible; UK only domestically
  - The national law normally says nothing specific about transnational access
- Network and encryption
  - Encrypted Virtual Private Network (VPN) tunnel (only ONS is using a private network)
  - Cryptographic Protocol: Secure Sockets Layer (SSL)

# Main Findings of the Survey I

- Access points
  - Anywhere
  - Institution
  - Safe room
- International access
  - In general possible; UK only domestically
  - The national law normally says nothing specific about transnational access
- Network and encryption
  - Encrypted Virtual Private Network (VPN) tunnel (only ONS is using a private network)
  - Cryptographic Protocol: Secure Sockets Layer (SSL)

# Main Findings of the Survey I

- Access points
  - Anywhere
  - Institution
  - Safe room
- International access
  - In general possible; UK only domestically
  - The national law normally says nothing specific about transnational access
- Network and encryption
  - Encrypted Virtual Private Network (VPN) tunnel (only ONS is using a private network)
  - Cryptographic Protocol: Secure Sockets Layer (SSL)

# Main Findings of the Survey I

- Access points
  - Anywhere
  - Institution
  - Safe room
- International access
  - In general possible; UK only domestically
  - The national law normally says nothing specific about transnational access
- Network and encryption
  - Encrypted Virtual Private Network (VPN) tunnel (only ONS is using a private network)
  - Cryptographic Protocol: Secure Sockets Layer (SSL)

## Main Findings of the Survey II

- User authentication
  - Two-factor authentication
    - Password
    - Smart-card/token with biometrics (fingerprint)
    - Smart-card/token with One-Time-Password (OTP)
    - IP-range and or hardware certificate
- Data storage
  - Different formats: SQL, ASCII, SAS, SPSS, Stata
  - Storage facility: home institution or at least home country
- Available software for user
  - SPSS, Stata, MS word, Latex etc.
  - Sometimes non-standard software against payment
- Fees for using the solution
  - Range from zero to many thousands (different models)

## Main Findings of the Survey II

- User authentication
  - Two-factor authentication
    - Password
    - Smart-card/token with biometrics (fingerprint)
    - Smart-card/token with One-Time-Password (OTP)
    - IP-range and or hardware certificate
  - Data storage
    - Different formats: SQL, ASCII, SAS, SPSS, Stata
    - Storage facility: home institution or at least home country
  - Available software for user
    - SPSS, Stata, MS word, Latex etc.
    - Sometimes non-standard software against payment
  - Fees for using the solution
    - Range from zero to many thousands (different models)



## Main Findings of the Survey II

- User authentication
  - Two-factor authentication
    - Password
    - Smart-card/token with biometrics (fingerprint)
    - Smart-card/token with One-Time-Password (OTP)
    - IP-range and or hardware certificate
- Data storage
  - Different formats: SQL, ASCII, SAS, SPSS, Stata
  - Storage facility: home institution or at least home country
- Available software for user
  - SPSS, Stata, MS word, Latex etc.
  - Sometimes non-standard software against payment
- Fees for using the solution
  - Range from zero to many thousands (different models)

## Main Findings of the Survey II

- User authentication
  - Two-factor authentication
    - Password
    - Smart-card/token with biometrics (fingerprint)
    - Smart-card/token with One-Time-Password (OTP)
    - IP-range and or hardware certificate
- Data storage
  - Different formats: SQL, ASCII, SAS, SPSS, Stata
  - Storage facility: home institution or at least home country
- Available software for user
  - SPSS, Stata, MS word, Latex etc.
  - Sometimes non-standard software against payment
- Fees for using the solution
  - Range from zero to many thousands (different models)

## Main Findings of the Survey II

- User authentication
  - Two-factor authentication
    - Password
    - Smart-card/token with biometrics (fingerprint)
    - Smart-card/token with One-Time-Password (OTP)
    - IP-range and or hardware certificate
- Data storage
  - Different formats: SQL, ASCII, SAS, SPSS, Stata
  - Storage facility: home institution or at least home country
- Available software for user
  - SPSS, Stata, MS word, Latex etc.
  - Sometimes non-standard software against payment
- Fees for using the solution
  - Range from zero to many thousands (different models)

# Want more?

-> [www.dwbproject.org/about/deliverables.html](http://www.dwbproject.org/about/deliverables.html)



Project N°: 262608



ACRONYM: Data without Boundaries

**DELIVERABLE DA.1**

*(Report on the state of the art of current SC in Europe)*

**Work Package 4**

*(Improving Access to OS Microdata)*

REPORTING PERIOD:	From: Month 1	To: Month 18
PROJECT START DATE:	1 <sup>st</sup> May 2011	DURATION: 48 Months
DATE OF ISSUE OF DELIVERABLE:	6 <sup>th</sup> September 2012	
DOCUMENT PREPARED BY:	Partners Names IAB, Destatis, CNRS-REQ, GENES, CNR, NSD, SDN	

Combination of FP 6 & 7 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



## What has been achieved, what is missing?

### ■ Achieved:

- Bringing access points to the user
- Safe connections to one data source
- User support (varies)
- data source specific accreditation and user account

### ■ Missing:

- General accessibility within Europe
- Connecting more than one data source
- Environment for collaborative work
- European (accreditation and) user account

## What has been achieved, what is missing?

### ■ Achieved:

- Bringing access points to the user
- Safe connections to one data source
- User support (varies)
- data source specific accreditation and user account

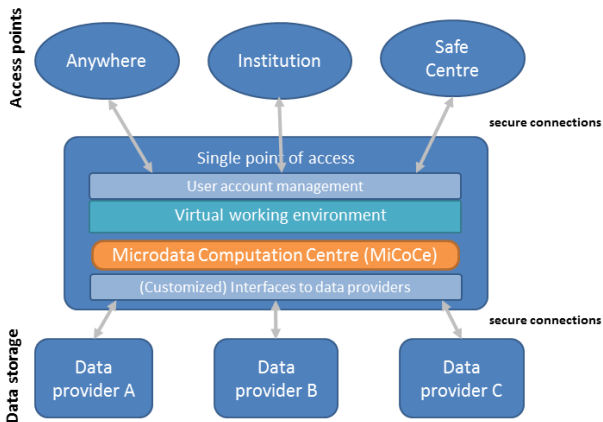
### ■ Missing:

- General accessibility within Europe
- Connecting more than one data source
- Environment for collaborative work
- European (accreditation and) user account

## What has been achieved, what is missing?

- Achieved:
  - Bringing access points to the user
  - Safe connections to one data source
  - User support (varies)
  - data source specific accreditation and user account
  
- Missing:
  - General accessibility within Europe
  - Connecting more than one data source
  - Environment for collaborative work
  - European (accreditation and) user account

# European Remote Access Network (Eu-RAN)





## Summary and Outlook

- Remote access is a promising and secure approach to make confidential data usable for research in the social sciences (of course, not the only one)
- A number of data providers already have running solutions
- The scope is often on national and not European level
- Eu-RAN is discussed with experts all over the world (and the discussion will go on)
- parallel a pilot by CASD, IAB and UKDS proofs parts of the concept
- Results of discussions and of the pilot will improve the concept for Eu-RAN
- Finally a live and running infrastructure will provide services for data providers and user (if we do not run out of time and money)

## Summary and Outlook

- Remote access is a promising and secure approach to make confidential data usable for research in the social sciences (of course, not the only one)
- A number of data providers already have running solutions
- The scope is often on national and not European level
- Eu-RAN is discussed with experts all over the world (and the discussion will go on)
- parallel a pilot by CASD, IAB and UKDS proofs parts of the concept
- Results of discussions and of the pilot will improve the concept for Eu-RAN
- Finally a live and running infrastructure will provide services for data providers and user (if we do not run out of time and money)

## Summary and Outlook

- Remote access is a promising and secure approach to make confidential data usable for research in the social sciences (of course, not the only one)
- A number of data providers already have running solutions
- The scope is often on national and not European level
- Eu-RAN is discussed with experts all over the world (and the discussion will go on)
- parallel a pilot by CASD, IAB and UKDS proofs parts of the concept
- Results of discussions and of the pilot will improve the concept for Eu-RAN
- Finally a live and running infrastructure will provide services for data providers and user (if we do not run out of time and money)

## Summary and Outlook

- Remote access is a promising and secure approach to make confidential data usable for research in the social sciences (of course, not the only one)
- A number of data providers already have running solutions
- The scope is often on national and not European level
- Eu-RAN is discussed with experts all over the world (and the discussion will go on)
- parallel a pilot by CASD, IAB and UKDS proofs parts of the concept
- Results of discussions and of the pilot will improve the concept for Eu-RAN
- Finally a live and running infrastructure will provide services for data providers and user (if we do not run out of time and money)

## Summary and Outlook

- Remote access is a promising and secure approach to make confidential data usable for research in the social sciences (of course, not the only one)
- A number of data providers already have running solutions
- The scope is often on national and not European level
- Eu-RAN is discussed with experts all over the world (and the discussion will go on)
- parallel a pilot by CASD, IAB and UKDS proofs parts of the concept
- Results of discussions and of the pilot will improve the concept for Eu-RAN
- Finally a live and running infrastructure will provide services for data providers and user (if we do not run out of time and money)

## Summary and Outlook

- Remote access is a promising and secure approach to make confidential data usable for research in the social sciences (of course, not the only one)
- A number of data providers already have running solutions
- The scope is often on national and not European level
- Eu-RAN is discussed with experts all over the world (and the discussion will go on)
- parallel a pilot by CASD, IAB and UKDS proofs parts of the concept
- Results of discussions and of the pilot will improve the concept for Eu-RAN
- Finally a live and running infrastructure will provide services for data providers and user (if we do not run out of time and money)

## Summary and Outlook

- Remote access is a promising and secure approach to make confidential data usable for research in the social sciences (of course, not the only one)
- A number of data providers already have running solutions
- The scope is often on national and not European level
- Eu-RAN is discussed with experts all over the world (and the discussion will go on)
- parallel a pilot by CASD, IAB and UKDS proofs parts of the concept
- Results of discussions and of the pilot will improve the concept for Eu-RAN
- Finally a live and running infrastructure will provide services for data providers and user (if we do not run out of time and money)

## Summary and Outlook

- Remote access is a promising and secure approach to make confidential data usable for research in the social sciences (of course, not the only one)
- A number of data providers already have running solutions
- The scope is often on national and not European level
- Eu-RAN is discussed with experts all over the world (and the discussion will go on)
- parallel a pilot by CASD, IAB and UKDS proofs parts of the concept
- Results of discussions and of the pilot will improve the concept for Eu-RAN
- Finally a live and running infrastructure will provide services for data providers and user (if we do not run out of time and money)



# Thank you!

[fdz.iab.de](http://fdz.iab.de)

Contact: [david.schiller@iab.de](mailto:david.schiller@iab.de)



[dwbproject.org](http://dwbproject.org)