Expanding the Research Data Center in Research Data Center Approach

IASSIST 2013

Joerg Heining
RDC-in-RDC: Motivation and Background

- 2011: No remote access systems to confidential data in Germany
  - Legal concerns
  - Problem of admission control
  - Problem of additional knowledge available outside a controlled environment

⇒ **RDC-in-RDC** approach may help to overcome this situation
  - Introduced in 2011
  - First step towards remote access

- Research Data Center of the German Federal Employment Agency (FDZ, located in Nuremberg, Germany)
  - Part of the Institute for Employment Research (IAB)
  - Micro (labor) market data on individuals, households and establishments including social security, other administrative data, and surveys.
**RDC-in-RDC: Idea**

**On-site access** to restricted FDZ data from institutions with the same standards (Guest-RDC) but different location.

- **Guest-RDC** (external site)
- **FDZ in Nuremberg**

**Secure internet connection (Citrix Server and Citrix Access Gateway)**

- Admission control/Identity check
- Applications, contracts, disclosure review, IT

- Server with FDZ data

**Interface to establish a communication link**
External Access Points

Since October 2011:

Hosts:

- RDCs of the Statistical Offices of the Länder (states)
- MiCDA data enclave at Institute for Social Research (ISR), University of Michigan, Ann Arbor, MI, USA
Supervision of External Access Points

- Sites in Germany: Supervised by staff of the Statistical Offices of the Länder
  - Experienced/trained in providing access to restricted data
  - Obliged to German data protection legislation

- Ann Arbor: permanent FDZ on-site staff as supervisor (one person)
  - Enforces German data protection legislation
  - Provides guidance and advice to users

- Local supervisors perform identity check before user gets admission to the safe room with the thin client computers

- Local supervisor establishes a secure communication link to server in Nuremberg
  - Two passwords: One is known to the user, the other one only to the local supervisor
Access Modalities and Funding

- In general: FDZ’s permission for on-site access does not depend on the location
  - Valid use agreement with the FDZ for on-site access is required
  - Additional requirements for on-site access in Ann Arbor (confidentiality agreement, IRB approval/exemption)

- Funding is provided by the German government (Federal Ministry of Education and Research) until 09/30/2013
  - Financial support for sites in Germany and the US
  - Additional staff for FDZ and Federal Employment Agency
  - Permanent on-site FDZ staff in Ann Arbor
Access Days

Nuremberg 2011: 387
Nuremberg: 505
Ann Arbor: 327
Duesseldorf: 154
Berlin: 68
Bremen: 51
Dresden: 47
New Access Points – Coming Soon

University of Applied Labour Studies of the Federal Employment Agency (HdBA)
- Supervision by HdBA staff
- In Operation since April

- Cornell University, Ithaca, NY, USA
- University of California at Berkeley, CA, USA
New Access Points – In The Future

- Harvard University
- University of Essex, Colchester, UK
- Insee, Paris
- More sites in Europe/the US?

Workpackage 4 of the Data without Boundaries Project
Challenges – Organization and Processes

- Cooperation agreements
- Established processes at FDZ and the initial access points within RDC-in-RDC
- Transfer of these processes to new sites
- Maintaining „separation“ between formal FDZ requirements for data access and additional requirements of local hosts
Challenges – Technical Issues

- Implemented system was designed to handle the five initial sites
  - Citrix environment
  - Currently: One terminal server running all applications for users

  ⇒ Re-design of the Citrix environment
  ⇒ Additional terminal servers

- Providing fast and adequate IT support
  - Time zones (minus 6(9) hours)
  - Non-German speaking local supervisors at new sites
Challenges – Supervision/Support of New Access Points

- Common problem to all new sites: no FDZ staff will permanently be present

- A staff member of the local host will act as a supervisor
  - Must be obliged to German data protection legislation
  - FDZ provides training
  - FDZ visits on a regular basis

- FDZ’s rules and standards need to be enforced by a „stranger“
  - Structure of codes
  - Rules on how to behave when working on the server
  - Adding external files (aggregate data, software extensions (ados), etc.) to a project
Challenges – Supervision/Support of New Access Points

- Providing support and guidance on access modalities, data, institutional backgrounds, etc.

- Implementation of a metadata system which also includes information on institutional backgrounds

- Provide more education and training on handling FDZ data via website/metadata system

- Challenges get larger once initial funding has expired - FDZ staff in Ann Arbor has to leave

- Best case scenario: Continuation of funding for permanent on-site FDZ staff in the US – pending.
Conclusion

- RDC-in-RDC works!
  - Access days
  - Interest by other data providers to be come a Guest-RDC
  - RDC-in-RDC is applied by other data providers

- Cheap

- Scalable
  - New Sites
  - Opposite direction (new project with several German data producers, just about to start)

⇒ Step towards remote access - Access is distributed rather than data
⇒ Blueprint for international data sharing
Technical Implementation

Citrix-Thin-Client-Solution:

- Thin client computer at the external site
- "Reduced" version of a computer
- Interface
- Runs only software needed to establish a communication link

Citrix Access Gateway and Citrix Server

- Citrix: Company
- Combination of software (Access Gateway) and server
- Secure Communication link
- International standard (for example Statistics Netherlands, financial institutions, etc.)