Building a Cross Border Data Access System for Improved Scholarship and Policy: The Case of the German IAB Network of RDCs

IASSIST Conference
June 4, 2014

Joerg Heining (IAB), Warren Brown (Cornell), Bill Block (Cornell), Stefan Bender (IAB)
IAB and FDZ

- FDZ: Research Data Center of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB)
- Established in 2004
- Located in Nuremberg, Germany
- Provides access to confidential micro labor market data on individuals/households and establishments
  - Administrative data: social security records, other data stemming from the labor administration (benefit receipt, participation in labor market programs, job search)
  - Surveys
  - Linked data (establishments and individuals, surveys and admin)
- “Weakly anonymized” data are only available on-site
- “Factually anonymized” data (Scientific Use Files) available off-site
IAB’s Network of RDCs

- Since 2012: Several on-site access points in Germany and the US („RDC-in-RDC approach“, Bender and Heining (2011), IASSIST Quarterly, 2011, 35 (3))
- Citrix-thin client technology

- Benefits to scholars, policy makers and data producers-hosts?
Access to a Unique Data Resource for Cutting Edge Research

- IAB data available to U.S. researchers has huge potential payoffs for the scientific community
  - Germany is third largest economy in the world with a workforce of 47 million people and an elaborate system of unemployment insurance and other social benefits
  - Researchers have relatively easy access to detailed longitudinal data (admin data since 1975)
  - These data provide an excellent basis to study the mechanisms governing labor markets and the incentive effects of labor market interventions.
  - IAB surveys may be linked to the administrative records of survey respondents

- Access to IAB data is free and easy
  - Several high ranked peer review journal articles by (US) data users
  - Benefit for (German) policy-makers
  - Value added to the university research centers hosting on-site access
Access to a Unique Data Resource for Cutting Edge Research

[Bar chart showing access to data resource by location and year, with data for Nuremberg, Berlin, Bremen, Dresden, Duesseldorf, Mannheim, Ann Arbor, Cornell, Berkeley, Harvard, and RDC-in-RDC total. The chart indicates access until 04/2014.]
Training and Education

- Average approval time for student projects is less than two months
  \(\Rightarrow\) students may use IAB data for class assignments as well as longer term projects such as master’s theses and doctoral dissertations

- Group approvals so that students may use restricted data for class (projects still take place at IAB on-site access points, but can be built into the regular classroom curriculum)

- Students and scholars learn how to work with administrative data and deal with problems

- Good and transferable experiences/skills when they want to work with comparable data from different data producers/populations centers (for example, data from the U.S. Census Bureau, Longitudinal Employer-Household Dynamics (LEHD) data, etc.) for future research
Realizing Positive Externalities in Data Sharing

- Staff of the hosts act as supervisor of IAB’s access points in the US
- Host staff learns about access rules, legal institutional backgrounds, available data and procedures at IAB
- Knowledge transfer between international experts on data access and dissemination
- Joint funding raising
- Basis for the further development of transnational data access, dissemination strategies and international data sharing

⇒ Scientific Use File hosting project between IAB and CISER
Off-Site Remote Access to IAB Data

- Joint project between IAB and CISER
- IAB also prepares and disseminates so-called Scientific Use Files (SUF) for off-site access
- Security Problem:
  - Approved users download SUF data from secure server to install locally
  - Potential abuse of illegally keeping local copy of the data after expiration of the use agreement
- Problem Solved:
  - IAB SUF data will be stored in the Cornell Restricted Access Data Center (CRADC)
  - Users connect to CRADC after concluding use agreement with IAB
  - Advantage: No local copies of SUF data
IAB@Cornell: http://ciser.cornell.edu/IAB.shtm

The Cornell Institute for Social and Economic Research, CISER is delighted to announce a collaborative effort with the Institute for Employment Research (IAB) of the German Federal Employment Agency enabling approved researchers to securely access confidential administrative microdata on labor markets. These resources are available to any researcher approved by IAB, no affiliation with Cornell is required.

Access will be available either from your campus office or in person at the IAB Research Data Center at CISER depending on the level of restriction.

COMING SOON: Access to less restricted Scientific Use Files will be available from your campus office. This is a unique resource available only through Cornell for any approved researcher.

Access to more restricted data on servers in Germany is through the IAB Research Data Center at CISER.

The data files contain detailed information on employment, unemployment benefit receipts, participation in labor market programs and registered job search, and a large number of socio-economic characteristics.

The following scheme depicts the different data sources and how they are combined in the available data products:
Log In Screen for CRADC Users

CISER Secure Data Services

You are attempting to access a computer system with restricted-access data operated by the Cornell Institute for Social and Economic Research (CISER) as a part of the Cornell Restricted Access Data Center (CRADC). If you do not have appropriate permissions you should not proceed. Unauthorized use of these data is subject to penalties imposed by Cornell University and by the providers of the data.

OK
Windows Environment for Access to Files
## Statistical Analysis Software on CRADC

<table>
<thead>
<tr>
<th>ArcGIS</th>
<th>JMP</th>
<th>RStudio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortran</td>
<td>Matlab</td>
<td>SAS</td>
</tr>
<tr>
<td>Gauss</td>
<td>Math Kernel</td>
<td>SAS DQ</td>
</tr>
<tr>
<td>IBMSPSS</td>
<td>MPLUS</td>
<td>Stata MP</td>
</tr>
<tr>
<td>IMSL</td>
<td>R</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

- IAB’s network of RDCs has been a true story of success so far
- Cheap and scalable solution for international data sharing
- Benefits to scholars and policy makers
- Positive externalities to hosts
- Benefits to IAB
  - Larger network of academic experts on labor market research
  - Larger network of data producers and data disseminators
- IAB will expand its network of RDC’s in the future.
- Ideal conception for the future: reciprocal RDC-in-RDC
Data available at FDZ

- Micro labor market data on individuals/households and establishments

Administrative Data

Social Security Notifications

Process-generated data of the BA

Data available at the FDZ

Surveys
Types of Data available at FDZ

- **complete microdata**
  - delete direct identifier
  - confidential microdata
  - on-site analysis potential
  - microdata

- **confidential microdata**
  - anonymisation method
  - de-facto anonymised microdata
  - off-site analysis potential

- **de-facto anonymised microdata**
  - stronger anonymisation method
  - fully anonymised microdata

Available at FDZ