IWH
Halle Institute for Economic Research
Member of the Leibniz Association

IWH Technical Reports

No. 1

April 2016

IWH R&D Micro Database

Part 1:
Data, Data Origin and Data Quality

Revision and Expansion of the Version from 2013-10-14

Mirko Titze, Matthias Brachert, Alexander Giebler, Wilfried Ehrenfeld

Authors:
Mirko Titze, Tel +49 345 7753 861, mirko.titze@iwh-halle.de
Matthias Brachert
Alexander Giebler
Wilfried Ehrenfeld

Contact:
Cornelia Lang
Head of the IWH Research Data Centre
Tel + 49 345 77 53 802
Fax + 49 345 77 53 820
cornelia.lang@iwh-halle.de

Issuer:
Halle Institute for Economic Research (IWH) –
Member of the Leibniz Association

Executive Board:
Professor Reint E. Gropp, PhD
Professor Dr Oliver Holtemöller
Dr Tankred Schuhmann

Address:
Kleine Maerkerstrasse 8
D-06108 Halle (Saale)

Postal address:
P.O. Box 11 03 61
D-06017 Halle (Saale)

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

Citation:
Leibniz-Institut für Wirtschaftsforschung Halle (IWH) (Hrsg.): IWH R&D Micro Database. Part 1: Data, Data Origin and Data Quality. Revision and Expansion of the Version from 2013-10-14. IWH Technical Reports 01/2016. Halle (Saale) 2016.

ISSN 2365-9076

IWH R&D Micro Database
Part 1: Data, Data Origin and Data Quality

Revision and Expansion of the Version from 2013-10-14

Mirko Titze, Matthias Brachert, Alexander Giebler, Wilfried Ehrenfeld

Abstract

Almost all industrialized countries have set up support schemes to foster private Research & Development (R&D) activities. However, only little is known on which programs are exactly applied, how much money is spent and whether such programs work in the way they were originally intended. Reliable evaluation studies that allow identifying causal effects can help to clarify if and how these policies work and for which policy schemes the results are promising. This kind of research design, however, places high demands on the quality of data. Against this backdrop, researchers at the Halle Institute for Economic Research (IWH) – Member of the Leibniz Association systematically collected a number of different datasets including information on granted R&D projects in Germany. Using comprehensive and complex procedures, these datasets have been harmonized and linked with datasets containing information on firm-specific characteristics of the recipients (and the non-recipients as well). This technical report provides an overview on the IWH R&D Micro Database, the support schemes included and its variables. Moreover, the harmonization routines are described.

Content

Abstract 1
1 Overview 3
2 Overview of the Funding Programmes included in the Database 5
3 Data Requirements 7
3.1 Information in the Funding Database and its Suitability for Evidence-based Evaluations 7
3.2 Suitable Data Sources for Evidence-based Evaluations 8
3.2.1 Bureau van Dijk (BvD) 8
3.2.2 Employment Statistics from the Federal Agency for Labour (BA) 9
3.2.3 Official Company Data for Germany (AFiD) from the Federal Statistical Office 9
3.2.4 Data from the German Association for Funding Humanities and the Sciences (Stifterverband) 10
3.2.5 Institutions of Publicly Financed Research 10
3.2.6 Other Isolated Economic Indicators 10
4 Harmonising and Linking Data Sets using Record-Linkage-Techniques 11
4.1 Harmonising the IWH R&D Micro Database 11
4.2 The Example of Saxony 12
5 Further Steps in Progress at IWH 14
Literature 16

IWH R&D Micro Database
Part 1: Data, Data Origin and Data Quality

Revision and Expansion of the Version from 2013-10-14

Mirko Titze, Matthias Brachert, Alexander Giebler, Wilfried Ehrenfeld

1 Overview

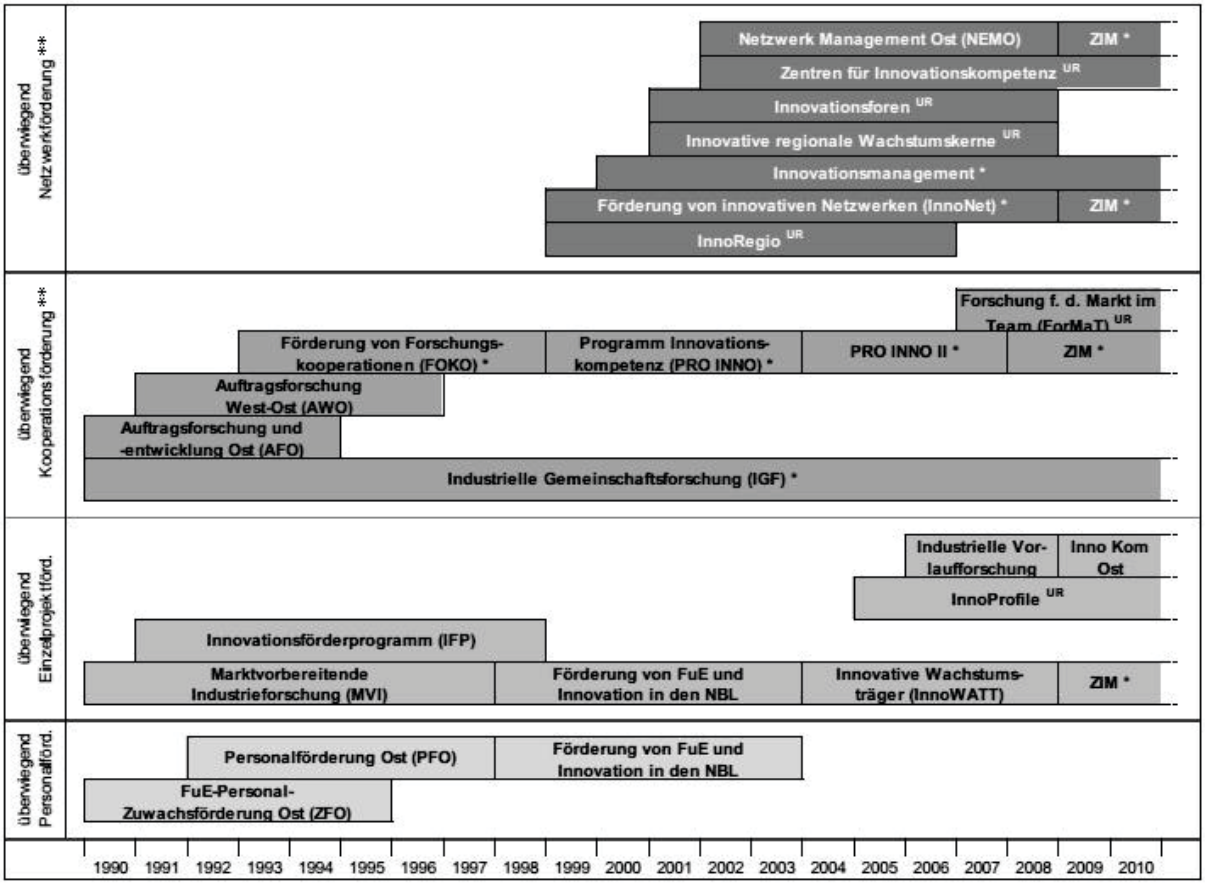
The IWH is studying economic catch-up processes and economic integration in Europe. Economic catch-up and growth processes depend on the efficiency of allocation of production factors and advancing productivity. The IWH is particularly studying how the financial system can assure capital (re)-allocation, structural change, innovation, and advances in productivity and thereby efficient and sustainable real economic development. An important aspect in this context is research into the connection between productivity at the corporate level and innovation. This also encompasses the systematic and comprehensive analysis of the R&D Funding Programmes employed.

Today, almost all industrialized countries employ instruments for funding research and development (R&D) to a greater or lesser extent. The use of R&D funding schemes is usually justified with the assertion that "market failure" prevents the optimal scope of R&D – but that innovation is the driving force of (regional) economic growth today. From a scientific perspective, admittedly, a market failure can only be demonstrated under the most restrictive of conditions, so that for this reason alone, scientific monitoring of state funding schemes is necessary.

Decreasing public budgets demand efficient use of the public funding. Evidence-based evaluations can in this respect provide a valuable impulse for drafting funding guidelines. The goal of such evaluations must be to establish the clearest possible linkage between the funding mechanism and its effectiveness. The keys to evaluation are, along with clearly defined goals in economic policy, also information about the units of analysis (e.g. businesses or regions), and not just about the parties being funded. This procedure places high demands on the quality of the data. Evidence-based evaluations are as a rule not feasible using only the administrative data on funding. Instead, adding further data that do not just described the funded units with adequate precision but also data on the control group, is necessary.

In practice, a number of support schemes are used to fund R&D and innovation; they are also subject to institutional change (target groups, funding quotas, object of funding, etc.) Figure 1 illustrates this for (selected) federal funding schemes that have been applied in East Germany since Reunification.

Figure 1:
Federal funding schemes (BMWI and BMBF***) for R&D and innovation in East Germany (including Berlin) since 1990



Notes: * Federal Programmes. – UR: Initiatives of the programme family "Business Region". – ** Network funding includes generally support of the network management, during the funding of the cooperation with the end of completing research projects. – It does not include programmes for funding new foundings. – *** BMWI: Federal Ministry for Economic Affairs and Energy. BMBF: Federal Ministry for Education and Research.

Source : Günther, Jutta; Nulsch, Nicole; Urban-Thielicke, Dana; Wilde, Katja: 20 Jahre nach dem Mauerfall: Transformation und Erneuerung des ostdeutschen Innovationssystems. Expertenkommission Forschung und Innovation (Hrsg.), Studien zum deutschen Innovationssystem No. 16-2010, Berlin 2010, S. 32.

An important aspect is that it is not just the Federal Government that applies funding schemes, but also the Federal States and the European Union. The applicant is free to decide on one of the many funding institutions. With that however it becomes immediately clear that only a combination of data sets from different funding institutions can provide a comprehensive image of the funding of R&D and innovation in order to avoid hidden treatment (Guerzoni and Raiteri 2015). The **uniqueness** of IWH R&D Micro Database is that it meets this desideratum.

Two aspects are currently the focus of research using R&D micro data at the IWH:

a) Analyses of the extent of innovative projects of various actors (universities and colleges, non-university research institutes, private business) as well as the interconnection of the actors in the framework of funding cooperation and

b) Analyses on the effects of R&D funding (in a further distinction).[1]

The data serve both goal-oriented **economic policy direction** at a regional, national and international level as well as purely academic **use.**[2] The IWH R&D Micro Database makes contributions to the geographic localisation of R&D activities and to the analysis of the effects of R&D funding from operational and regional perspectives. The content of the project will be coordinated by the **Centre for Evidence-based Policy Consulting at the IWH (IWH-CEP).**[3]

2 Overview of the Funding Programmes included in the Database

The **IWH R&D Micro Database** currently encompasses (status: 06.30.2015) eight data sets on various programmes that provide direct R&D project funding[4] from the German Federal Government, the German States and the European Union.

The data come from the project sponsors and/or the responsible Ministries. Access is free with some data sets – others can only be purchased from the ministries.[5]

For the Saxon data for example every publication must be coordinated with the Director of Department 37 (innovation policy, technology funding) in the Saxon State Ministry for Business, Labour and Traffic (German abbr. SMWA), Mr. Christoph Zimmer-Conrad.

The programmes combined in the IWH R&D Micro Database are shown in Table 1. It provides an overview of the most important characteristics recorded for these programmes. The greatest coverage currently exists for the States of Saxony and Saxony-Anhalt. Here there are data for direct R&D project funding from the German Federal Government, the German States and the European Union as well as data on the cooperating partners of the applicant.

---

1 In a broader sense investment funding is understood as a component of funding for innovation. This derives from the fact that the rules e.g. only allow the purchase of new machines and equipment. The literature refers to the so-called product-embodied R&D flows in this context, meaning innovation, that are passed on in the deliver chain (e.g. OECD 1996 and Papaconstantinou et al. 1996).

2 At IWH there are currently two research projects on evidence-based valuation of funding, first *Possible Avenues of Scientific Evaluation of Business Subsidies in the Framework of the Regional Project „Improving Regional Economic Structure" (German abbreviation GRW) in Saxony-Anhalt* and second *Evaluation of the Funding Initiative „Innovative Regional Growth Cores" in the Framework of the BMBF-Innovation Initiative for the New States „Business Region"*.

3 The Centre for Evidence-Based Policy Advice at the IWH (IWH-CEP) was founded in 2014. It is a platform that bundles activities in research, teaching and political consulting and structures them with the objective of providing better foundations for a causal analysis of the economic policy measures in Germany. The IWH-CEP is designed as a service unit and supports the activities in the research groups in which access is provided to a cross-regional network of research and political consulting as well as access to the data sets for causal analyses. Additional information is available under the following link: http://www.iwh-halle.de/d/Research/cep/start.asp.

4 This document construes the term research and development (R&D) broadly. In this broad definition it also includes programmes to promote investment.

5 Data sets accessible for free to all interested parties are e.g. the Funding Catalogue for Funded Federal Projects(www.foerderkatalog.de) and the Cordis-Database for Funded EU-Projects http://cordis.Europe.eu/search/index.cfm?fuseaction=search.simple). All other data are subject to strict confidentiality – publication of these data must be approved by the (State) Ministries.

Table 1:
Datasets[a] of the IWH R&D Micro Database (as of: 30/06/2015)

no.	dataset/funding programme	funding institution	type of project funding	object of funding	in the iwh since	dates cover period	regions included	provision of data by	access	number of projects included
1	funding catalogue	federal government	individual+ association	innovation	2010	1968-2014 (October)	Germany	BMBF, DLR	freely available[c]/exclusive of important characteristics	162 910
2	SAB funding database	free State of Saxony	individual+ association	innovation	2007	1991-2012 (September)	Saxony	SAB, SMWK	exclusive	7 816
3	funding database	State of Saxony-Anhalt	individual+ association	innovation	2012	1998-2011	Saxony-Anhalt	MWW, IB	exclusive	1 238
4	central Innovation Program for SME (ZIM), co-op+solo	federal government	individual+ association	innovation	2012	2004-2012 (co-op), 2009-2012 (solo)	Saxony[b], Saxony-Anhalt	BMWi, AiF, EuroNorm	exclusive	4 899
5	pro Inno + Pro Inno II (precursor of ZIM)	federal government	association	innovation	2013	1999-2004[e] , 2004-2009[f]	Saxony[b], Saxony-Anhalt[b]	BMWi, AiF	exclusive	4 406
6	EU FP 7	EU	association	innovation	2013	2007-2013	EU	SMWK	freely available[d]/exclusive of important characteristics	18 507
7	EU FP 6	EU	association	innovation	2015	2000-2006	EU	SMWK, BMBF	freely available[d]/exclusive of important characteristics	10 107
8	joint mission "Improvement of regional economic structure"	federal+ states	individual	investment	2014	2007-2013	Saxony-Anhalt	MWW, IB	freely available[g]/exclusive of important characteristics	1 654

Abbreviations: AiF – Arbeitsgemeinschaft industrieller Forschungsvereinigungen, AiF Projekt GmbH, project executing organisation for the Federal Ministry for Economic Affairs and Technology (BMWi); BMBF – Federal Ministry of Education and Research; BMWi – Federal Ministry for Economic Affairs and Technology; DLR – project executing organisation at the German Aerospace Center; EU – European Union; EuroNorm – EuroNorm Gesellschaft für Qualitätssicherung und Innovationsmanagement mbH, project executing organisation of the federal government; IB – Investitionsbank Sachsen-Anhalt; MWW – Ministry for Science and Economic Affairs; SAB – Sächsische Aufbaubank - Förderbank; SMWK – Saxon State Ministry for Science and Art.

Comments: [a] The original datasets do not have a panel structure because the observed entity is the project. The characteristics of the project (e.g. participating partners, amount authorised, industry membership, etc.) do not change during the lifetime of the project. However, the datasets can be transferred into a panel structure if the grant recipients (or executing agencies) are recorded one-to-one, e.g. with an identification number. This makes it possible to display a timeline of when a grant recipient applied for and worked on a project. Datasets nos. 2, 4, 5 and 6 have one-to-one identification numbers for actors. However, these are only one-to-one within the datasets, not across datasets (cf. also Section 3). – [b] And co-operation partners outside the state. – [c] http://www.foerderkatalog.de. – [d] http://cordis.europa.eu/fp7/projects_de.html. – [e] Pro Inno. – [f] Pro Inno II. – [g] A list of beneficiaries of public funds is prepared for the state of Saxony-Anhalt for co-financed projects in accordance with agreements on EU structural funds financing.

Source: Own rendering.

3 Data Requirements

3.1 Information in the Funding Database and its Suitability for Evidence-based Evaluations

Analyses of effectiveness should establish a connection between state intervention and the stated objectives. The fundamental evaluation problem consists of answering the question of what would have happened without state intervention. Technically, this means creating a "counter-factual" situation: How, for example, would a certain business have acted if it had not received any funding? The question "What would have happened if a business had not been funded?" can only be answered by the contrast between funded and non-funded businesses.

Basically the evidence-based evaluation procedures can be distinguished in macro- and micro-methods. An analysis of the overall economic effects of funding measures is however difficult, as a rule (Bade et al. 2012), for which reason conventional studies often focus on the micro-level. Analyses of this kind then require isolated economic data (and that not just for the subject group but also for the control group) about:

- the funded economic unit (most of all name, region and industry, size),
- the funding (amount of funding, date of funding, legal regulations) as well as
- economic objectives (e.g. employment, wages/salaries or sales).

The **information** mentioned however is **not to be found** in the list **of funding programmes** given in the IWH R&D Micro Database. Each of the data sets listed in Table 1 contains first a clear encryption of the projects. Based on an identification number one can see which project was carried out in cooperation and which was not. Depending on the programme, the businesses, non-university research institutions and universities are entitled to submit applications. Cooperation's therefore take place in different combinations of all three named institutions, e.g. university-business, business-non-university research institution-university, business-business, etc.

Each of the data sets also contains information about the name of the applicant, his regional characteristics (so-called General Community Identification Number and/or City Name with Postal Code), its industry, the project period (date at the start and end) and the amount approved. The data set for Saxon R&D project funding (No. 2 in Table 1) contains additional characteristics e.g. on project volumes, funding quotas, etc., that are not included in the other data sets.

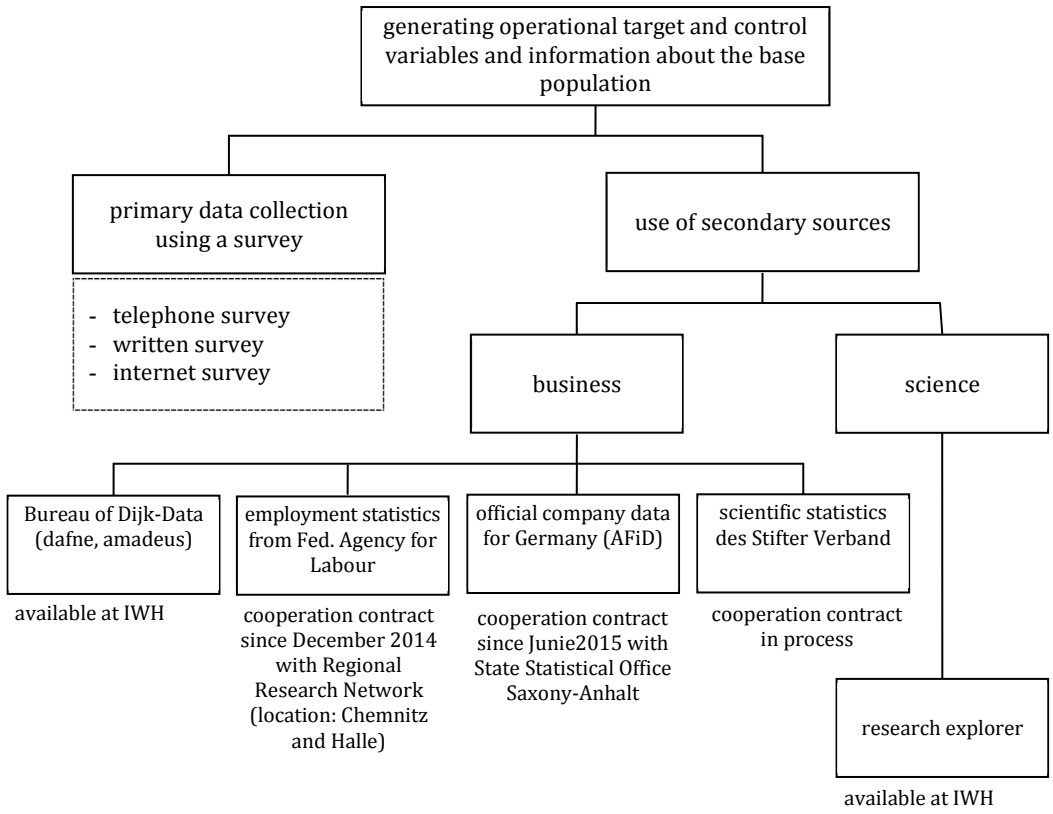
Since the data sets from different funding institutions are supplied, variable names for identical characteristics will be different, for example the names for the amount of funding, the start and end of financing, the recipient of the contribution and the office carrying it out. At the IWH, a concordance of the names of variables was generated that allows one to combine the data sets. Additionally, the industry-specific information relies on different classification systems (e.g. Nice Rev. 1.1 versus Nace Rev. 2) and different levels of detail (sections, departments, groups, and classes), the regional information at different local territories. This information is currently being harmonised at the IWH.

3.2 Suitable Data Sources for Evidence-based Evaluations

In order to still be able to carry out evidence-based analyses, information about the funded units (and the group of unfunded actors) must be added from other data sources. Basically there are two ways to generate missing economic target and control variables: primary data collection using a survey as well as use of secondary sources (Figure 2).

Primary collection brings with it a high time commitment and financial demands– not just with the collecting institution, but also from the actors being surveyed. Participation in such as survey is generally voluntary. High response rates are not assured. Due to these disadvantages the official public statistics office avoids collecting data that have already been recorded in some form a second time. Instead, data from multiple data sources are combined using suitable techniques of data processing (so-called Record-Linkage-Methods).

Figure 2:
Overview of Suitable Data Sources



Source: Based on IWH (2014).

A summary of the advantages and disadvantages of the data sets covered in detail in the following is found in the IWH (2014, Section 6.5).

3.2.1 Bureau van Dijk (BvD)

Business Bureau van Dijk (BvD) in collaboration with Creditreform e. V. offers business information for around 1.3 million German businesses in the so-called dafne-database. This database contains important business characteristics (e.g. legal form, industry key, district code, number of employees, sales, capital, etc.). The data do not come from public statistics, but are instead based in large measure on data that the capital companies are required to publish as

well as self-disclosures from the businesses. The data set evinces some gaps in important key figures such as the number of employees, sales or capital.

Combining these with funding statistics can be undertaken on the basis of the company names and regional characteristics using Record-Linkage-Techniques. BvD Data are already held basically for the period 2003-2012 in the IWH.

3.2.2 Employment Statistics from the Federal Agency for Labour (BA)

Use of employment statistics from the BA brings a series of advantages (Bade and Alma 2010, p. 2 ff.): even and reliable recording, information about the other characteristics (e.g. Income from employment, qualification of the employee, work done) as well as a high degree of coverage. These additional characteristics thus allow an evaluation of the effects of funding on the quantity and quality of the jobs created.

Another advantage consists in the ability to easily link the funding cases with the employment statistics using the business number. The business number is also recorded in the funding statistics. Certain limitations are imposed by the focus on employees who are subject to public insurance obligations. At the same time, the degree of coverage should be around 70-80% of all employed persons. One disadvantage consists in that these statistics do not record any information about employment itself (e.g. sales, product pallet or investments). Access to this data source is only possible through an application submitted to the BA and/or the Federal Ministry for Labour and Social Security. There has been a cooperation agreement between the IWH and the Regional Research Network at IAB (locations Chemnitz and Halle) since December 2014. Data set No. 8 from Table 1 were already linked with business numbers.

3.2.3 Official Company Data for Germany (AFiD) from the Federal Statistical Office

Another source containing information about the base population and suitable economic target and control figures are the official company data (AFiD). These data are provided by the Statistical Offices of the Federal States. They use the business register to combine all micro-data on business and environmental statistics. Additionally, the data are supplied in a panel structure. Integrated panel data are currently available for the areas of agriculture, services, processing companies and energy. It is processing businesses that play a pre-eminent roll in R&D funding. The AFiD-Panel for Industrial Operations and Industrial Management are suited to provide an analysis of exactly these branches. This means a compete record of units with 20 or more employees. The annual reporting group encompasses 68,000 firms.

Compared to the employment statistics from the BA, to this data set has the advantage that it includes other target economic values such as sales and investments along with employment. This allows one to make statements about productivity – i.e. about improving competitiveness in a business.

The data set named here is maintained in the Research Data Centre (German abbr. FDZ) of the State Office for Statistics, Saxony-Anhalt in Halle (Saale). Adding the funding information is possible using Record-Linkage-Techniques and would be carried out by employees at the FDZ. Practically this will be undertaken based on the commercial registry numbers, the tax ID numbers as well as the figures from the Bureau van Dijk. This information is however not included in the administrative data for which reason it is necessary to link with the BvD data set as a first step (see Section 3.2.1). The legal requirement for combining the data proceeds from §13a of the Federal Statistics Act. According to it, the data being added must come from publicly accessible sources.

For the services sector, there is also a panel – admittedly this one depends on random sampling. The random sample includes around 15% of all businesses in the service sector. This data set seems less suitable for the evaluation of funded businesses in the service sector.

Since June 2015 a contract has been in place between the IWH and the State Statistical Office for the use of the AFiD Data. Matching with the Funding Data Sets No. 1 and 8 is currently ongoing – the goal is to improve the matching quality.

3.2.4 Data from the German Association for Funding Humanities and the Sciences (Stifterverband)

All data sets described to date contain no (or little) information about the R&D activities of business. They are collected by Wissenschaftsstatistik GmbH in the German Association for the Funding of Humanities and the Science on commission for the Federal Ministry for Education and Research (BMBF). Essential key figures here are the R&D expenditure as well as R&D personnel. Linking the data sets can be done using the funding identification numbers (in the case of German Federal government funding) and also by the Bureau van Dijk Number. A cooperation contract between the IWH and the Stifterverband is currently being negotiated.

3.2.5 Institutions of Publicly Financed Research

The funding programmes mentioned in Table 1 are only directed at the actors in the economy. Moreover, political leadership places a high premium on actors in publicly financed research being integrated into the innovations process. Entire programme lines aim at networking business with universities and non-university research institutions. The base population of the publicly financed scientific instructions is listed in the so-called Research Explorer. This data source is provided as an online-database under the address http://research-explorer.dfg.de/research_explorer.de.html. This registry encompasses around 23,000 entries on institutions at German universities and non-university research institutions. At universities the level of detail is taken down to the chair level. The information is organised according to geographical, subject-related and structural criteria. The information can be combined with funding statistics based on the name and a regional characteristic using the Record-Linkage-Techniques. Since the data from the Research Explorer are not clearly encrypted, systematic processing is currently being carried out at the IWH.

3.2.6 Other Isolated Economic Indicators

Finally there are databases that provide additional important indicators of the innovation process. These are patent databases (PATSTAT, RegPat, DPMA) as well as publication data (Web of Knowledge Database from the provider Thomson Reuters). The units recorded in the former sources are patents in which essential information about the applicants, inventors, patent classes (fields of technology) as well as patent citations are listed. There is usually a regional characteristic recorded for the applicants and inventors as well. However the inventor and place of registration often diverge in the patent data. A large proportion of inventors are also individuals which makes assignment to an institution (business or scientific institution) nearly impossible, since the affiliation is not stated in the databases. Using Record-Linkage-Techniques at the IWH, the applicants registering European patents (database RegPat) in the Federal Republic of Germany were harmonised and linked with the database from the provider Bureau van Dijk (see also Section 4 in this document).

The information in the publication databases are derived essentially from publications in ranked international professional journals. As a rule the affiliation of each author is given and also includes a regional characteristic. Admittedly the orthography of the author's names and affiliations can differ in some cases drastically so that extensive work is required to harmonise

the data. The Competency Centre for Bibliometry in Bielefeld has corresponding experience in the field of publications data in the Federal Republic Germany.[6]

4 Harmonising and Linking Data Sets using Record-Linkage-Techniques

4.1 Harmonising the IWH R&D Micro Database

The following procedure has proven suitable at IWH for providing data sets for evidence-based evaluations. In a first step the funding data sets named in Table 1 are combined, then the variable names and variable formats (e.g. in date format) are harmonised based on a concordance generated by IWH.

The actors named in the database are harmonised using this method.[7] Since the data sets are provided by different institutions, the names of the actors are not uniformly recorded. There are also occasional typographical errors in recording names. To be distinguished in a technical sense from the slight variations in orthography that arise are clearly different designations for the same institution. An example for this would be the Technical Universities (with the often used abbreviation "TU") like (as a "Classic") the Rhineland-Westphalian Technical University of Aachen (abbreviation: RWTH Aachen).

To this end a systematic harmonisation of the actors using Record-Linkage or Data-Matching Techniques is needed (see for example Christen 2009 as well as Magerman et al. 2006). The term "Record Linkage" designates the combination of information of two data sets of which it is assumed that they refer to the same unit/entity (Herzog et al. 2007, S. 81).

Prior to the actual Record-Linkage-Procedure, a harmonisation of the regional characteristics had proven itself, based on postal codes on the Official Keys to Municipalities (AGS 8). Also to be taken into account are reforms of community territories, e.g. in Saxony 2006, in Saxony-Anhalt 2008, in Mecklenburg-West Pomerania 2011. After updating the regional characteristics it is still a good idea to remove "disruptive" excessive information from the names of the actors (e.g. "Department", "Division", "Faculty", "Chair").

After this there follows a so-called "Pre-cleaning" (c.f. Magerman et al. 2006). This is a correction of characters. To this end the names of the actors are completely converted to capital letters. In the course of this conversion German umlauts or accented characters like "è" are also replaced with their unaccented equivalents. Two blank spaces in a name as well as empty spaces at the beginning or end of a name will be removed. Then parentheses and spelling variations for "and" will be rendered uniform and any expressions in parentheses will be extracted. Following this the legal forms of businesses will be identified (and in saved in a separate variable). This will be done using an identification table, which at this current point in time has more than 600 orthographic variants for different forms of corporate legal organisation. The original spellings of the forms of these companies will then be removed from the names of the businesses. Finally some spellings of frequently used terms that are subject to orthographic variations will be harmonised. Then all blank spaces will be removed from the pre-cleaned expression. These procedures are very-well suited to guarantee classification of slightly variant spellings for the name of the same actor. Since parts of these data were manually recorded or came from scans of

6 http://www.Researchsinfo.de/Bibliometrie/index.php?id=home.

7 Note in this process that a funding process distinguishes between two types of actors: the recipient of the allocation and the administrating office. There former receives the actual allocation, the second actually administrates the project. In most cases both types of actors are identical. In some cases however there are drastic differences. The typical case are the Institutes and Application Centres of the Fraunhofer-Society – the recipient of the allocation is always the central office at the Munich location. The projects are actually worked on at the various locations of the institutes.

documents in paper form, there are isolated spelling errors in the data sets. These include variations in the use of hyphens or blank spaces that prevent direct classification. Delays in this phase can hardly be compensated for even using so-called "fuzzy "classification algorithms. The work performed here is however important and is a good investment in a secure classification.

Using a purely deterministic classification of the original data sets or "fuzzy" methods alone makes it hardly possible in cases of highly variable orthography (e.g. "TU" for "Technical University") to guarantee a sure classification. These cases can be recorded for the purposes of standardisation using automated replacement rules and/or with an additional table for different orthographies at the same institution.

The actual Record-Linkage-Procedure makes use of the Software MergeToolBox which was developed at the German Record Linkage Centre[8] and provided for download for scientific uses (Schnell et al. 2005). Using an algorithm (so-called trigram), similar names of actors in the same region (3 digits of AGS) will be assigned to each other based on a clear identifying characteristic.

The second step includes finally the same procedure for adding secondary data to the funding data sets. Technically speaking, the primary key from the external data source will be assigned to the primary key for the actor generated in the IWH.[9]

4.2 The Example of Saxony

As mentioned at the outset, the greatest coverage in terms of funding programmes at this time is for the States of Saxony and Saxony-Anhalt. This section provides an example of the harmonisation of the data sets in the IWH R&D Funding database based on the example of the Free State of Saxony. From Table 1 , the data sets No. 1 and 2 and 4 to 7 are included in the assessment, and grouped by state programmes (No. 2), Federal programmes (Funding Catalogue as well as ZIM and predecessor programme, i.e. No. 1, 4 and 5) and the E.U. Framework Programme for Research (No. 5 and 6). Also common is a breakdown in to the funding periods of EU. In the concrete case the analysis encompasses the periods of 2000-2006 as well as 2007-2013. The introduction already emphasized that the funding recipients can select from the various funding offers since the programme lines are by and large substitutes.

Beyond a doubt there are marginal differences between them – but for the most part they are identical:

- In all cases this is additional grant funding (so-called non-repayable grants).[10]
- The grant is based on a portion of the funding-eligible costs (in large measure personnel costs for the scientists working on the project. The funding-eligible costs are defined by the guidelines of the EU.
- The same upper limits on funding apply for all three programme lines (broken down by science and business, graduated based on the size classifications of businesses according to the EU guidelines).

8 http://www.record-linkage.de.

9 The procedure described here was used successfully in the BMBF-funded project „Hochschulstrategien für Beiträge zur Regionaldevelopment unter Bedingungen demografischen Wandels". Data from direct R&D Project Funding by the Federal Government (No. 1 in Table 1) were combined (for selected study regions) with patent and bibliometric data. The concrete goal was to measure the extent to which actors who benefitted from R&D-Funding are integrated into the different areas of knowledge production and transfer (Titze et al. accepted for publication).

10 Other types of funding are loans, allowances, suretyships.

From these constellations, a total of seven combinations are possible and these are the three individual forms as well as mixed forms. Table 2 shows the claiming of the different possible combinations by Saxon actors. In total, for the time period 2000-2013, 4,230 actors [received-verb missing in original] approvals. 2,120 (around 50%) of them selected exclusively federal programmes – the largest share of these being the ZIM funding and its predecessor programmes (not explicitly shown in the table). 1,115 (around 26%) of the actors decided on State funding from Saxony exclusively.

Table 2:
Use of funding programmes by Saxon Actors

Period 2000-2013

no.	combination of funding programmes	frequenz	percent
1	only state	1,115	26.4
2	only federal	2,120	50.1
3	only EU	89	2.1
4	state and federal	741	17.5
5	state and EU	8	0.2
6	federal and EU	73	1.7
7	state, federal and EU	84	2.0
	total	4,230	100.0

Source: RohData , IWH R&D Micro Data bank; calculations from the IWH.

Table 2 shows that the evaluation of a single programme without the general embedding in the overall German funding landscape is not possible. If for example the ZIM programme is evaluated, then the control group may not be made up of non-ZIM funded actors. There were in the non-ZIM funded actors group those who selected a substitute for ZIM, for example Saxon State funding. At the same time Table 2 shows however the evaluation designs within the group of R&D funded actors. For example one can investigate the additional effects of Federal government funding on the actors who used State funding (Lines 1 and 4).

Table 3 shows the application patterns over the course of time, Of the 4,230 actors observed, only 1,485 (around 35%) were new to funding in the period 2000-2006; 1,668 (around 40%) were new in the period 2007-2013. Overall it is clear that many actors, (around 60%) remain constantly in some form of R&D-funding.

The way ahead in data processing consists in creating linkage with the data from Bureau van Dijk. In a first step it was possible to find 2,532 of the 4,230 actors (around 60%) in the databases mentioned. Since actors in academia are included among those receiving funds (who are not listed in the data from Bureau van Dijk), it is necessary to make the link with Research Explorer in the second step (c.f. also Section 3.2.5). The actors the remaining would need to be manually researched later. Random samples of the actors not yet found in Bureau van Dijk have indicated that they are listed in the online portal firmenwissen.de with a corresponding BvD-Number.

Matching with the Bureau van Dijk data furthermore creates the link to the patent data, that would doubtless present and interesting result for R&D funding (c.f. Section 3.2.6).

Table 3:
Use of funding programmes by Saxon actors over time

Period 2000-2006 and 2007-2013

2000-2006 \ 2007-2013	no funding	only state	only federal	only EU	state and federal	state and EU	federal and EU	state, federal and EU	total
no funding	0	427	1,044	50	118	3	11	15	1,668
only state	561	127	74	0	81	1	0	6	850
only federal	740	23	336	5	67	1	27	4	1,203
only EU	36	0	5	3	1	0	0	0	45
state and federal	112	36	78	1	152	1	1	15	396
state and EU	2	1	2	1	1	0	0	0	7
federal and EU	7	2	8	1	2	0	9	0	29
state, federal and EU	0	1	3	1	14	1	1	11	32
total	1,458	617	1,550	62	436	7	49	51	4,230

Source: RohData , IWH R&D Micro Data bank; calculations from the IWH.

5 Further Steps in Progress at IWH

In the future, the IWH will focus more directly on systematically and comprehensively expanding the knowledge acquired about the programmes for R&D funding in the Federal Republic of Germany. This will be done essentially according to the following points:

- In the final version, the document "IWH R&D Micro Database" shall consist of three components. The current Part 1: **Data, Data Origins and Data** will be supplemented with two additional parts. Part 2 will include an **overview of the institutional** arrangements for R&D funding policy; Part 3 an **Overview of research questions and suitable methods,** for answering them and analysing the data for R&D project funding.

- Figure 1 shows that **funding** programmes are subject to change. For example, funding quotas, the scope of funding-eligible cots or the scope of funding-eligible branches can change. Another topic for future research is **which R&D funding programmes have or had what significance** (measured by the number of approvals as well as the sums approved). This information is necessary for estimating the degree to which the database covers the actual R&D project. Corresponding overviews are currently– as far as is known– not in existence and must be drafted comprehensively.[11]

[11] Ms Dr Gisela Hillmann, DLR (Gisela.Hillmann@dlr.de), was engaged with the set-up of the database "PROMO", which lists all practiced support schemes (federal, states, EU) in Germany. First meetings with Ms Hillmann already took place at the IWH. As far as is known, data for operational funding for the funding period 2007-2013 are available almost completely. For the periods before, larger gaps are likely to exist because many information were not available anymore.

- The data for the IWH R&D Micro Database shall be regularly updated in the future. Additionally, we shall strive to add to it, mostly in terms of additional state data. In the New German States especially, many instruments of direct R&D project are and were employed by the States.

Literature

Bade, F.-J.; Bornemann, H.; Breuer, A.; Rautenberg, R.: Ansätze für ein besseres Monitoring und eine verbesserte Erfolgskontrolle der Infrastrukturfunding innerhalb der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur". Endbericht. Gutachten im Auftrag des Thüringer Ministerium für Wirtschaft, Arbeit and Technologie. Berlin, Bremen and Dortmund 2012.

Bade, F.-J.; Alm, B.: Evaluierung der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur" (GRW) durch einzelbetriebliche Erfolgskontrolle für den Förderungsperiode 1999-2008 und Schaffung eines Systems für ein gleitendes Monitoring. Dortmund 2010.

Christen, P.: Data Matching – Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Springer: Berlin, Heidelberg 2012.

Guerzoni, M.; Raiteri, E.: Demand-side vs. Supply-side Technology Policies: Hidden Treatment and New Empirical Evidence on the Policy Mix, in: Research Policy, Vol. 44 (3), 2015, 726-747.

Herzog, T. N.; Scheuren, F. J.; Winkler, W. E.: Data Quality and Record Linkage Techniques. Springer Science + Business: New York 2007.

Institut für Wirtschaftsforschung Halle: Möglichkeiten einer wissenschaftlichen Evaluation der gewerblichen Funding im Rahmen der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur" (GRW) in Saxony-Anhalt, Stand 19.02.2014, mimeo.

Magerman, T.; Van Looy, B.; Song, X.: Data production methods for harmonized patent statistics: Patentee name harmonization. 2006

OECD: Technology and industrial performance, OECD: Paris 1996.

Papaconstantinou, G.; Sakurai, N.; Wyckoff, A.: Domestic and international product-embodied R&D diffusion, in: Research Policy, 27, 1996, 301-314.

Schnell, R.; Bachteler, T.; Reiher, J.: MTB: Ein Record-Linkage-Programm für die empirische Sozialforschung. 2005.

Titze, M.; Ehrenfeld, W.; Piontek, M.; Pippel, G.: Netzwerke zwischen Hochschulen und Wirtschaft: Ein Mehrebenenansatz, in: Fritsch, Michael; Pasternack, Peer; Titze, Mirko (Hrsg.): Schrumpfende Regionen – dynamische Hochschulen. Hochschulstrategien im demografischen Wandel. Springer Fachmedien, Wiesbaden 2015, 213-235.

Halle Institute for Economic Research (IWH) –
Member of the Leibniz Association

Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany

P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820
www.iwh-halle.de

ISSN: 2365-9076