FDZ

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

FDZ·Methodenreport

06/2014

EN

Methodological aspects of labour market data

The MPI-IC-IAB-Inventor Data 2002 (MIID 2002): Record-Linkage of Patent Register Data with Labor Market Biography Data of the IAB

Matthias Dorner,
Stefan Bender,
Dietmar Harhoff,
Karin Hoisl,
Patrycja Scioch

Bundesagentur für Arbeit

The MPI-IC-IAB-Inventor Data 2002 (MIID 2002): Record-Linkage of Patent Register Data with Labor Market Biography Data of the IAB.

Dorner, Matthias[a]; Bender, Stefan[a]; Harhoff, Dietmar[b c]; Hoisl, Karin[b c] and Patrycja Scioch

[a] Institute for Employment Research (IAB), Nuremberg

[b] Max Planck Institute for Innovation and Competition (MPI-IC) - Munich Center for Innovation and Entrepreneurship Research (MCIER), Munich

[c] Ludwig-Maximilians-Universität München (LMU), Munich

Abstract

This report describes the generation of a linked employer-inventor data set, the MPI-IC-IAB-Inventor Data 2002 (MIID 2002), using methods of record linkage. The MIID 2002 combines patent register data comprising inventors residing in Germany and who are listed on patent filings with the DPMA in 2002, with administrative labor market biography data on employees. Labor market biographies originate from social security records and are provided by the Institute for Employment Research (IAB). Our matched data comprises 46,180 unique employee-inventor pairs who were involved in the filing of 42,435 patents with the German Patent and Trademark Office (DPMA) in 2002. With its rich scope of variables combining individual and establishment characteristics with patent and inventor related information, the MIID 2002 provides a novel register based data set for research on inventors and their patenting activities in the context of the labor market.

Zusammenfassung

Der Methodenreport beschreibt die Erstellung eines verknüpften Erfinder-Betriebs-Datensatzes, den MPI-IC-IAB Erfinder Daten 2002 (MIID 2002), unter Verwendung von Methoden des Record Linkage. Der Datensatz verknüpft Patent-Register-Daten zu Patentanmeldungen von in Deutschland wohnhafter Erfindern beim Deutschen Patent- und Markenamt (DPMA) im Jahr 2002 mit Erwerbsbiografiedaten zu sozialversicherungspflichtig Beschäftigten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Insgesamt umfasst der Datensatz 46,180 eindeutig identifizierte Beschäftigten-Erfinder Paare und die von diesen Personen im Jahr 2002 angemeldeten 42,435 Patente. Durch das breite Spektrum an Variablen, das eine Vielzahl von Individual- und Betriebsmerkmalen mit Informationen zu Patenten und Erfindern umfasst, bieten die MIID 2002 Daten eine neuartige, auf administrativen Daten beruhende Datenbasis zur Erforschung von Erfindern und deren Patentaktivitäten im Kontext des Arbeitsmarktes.

Keywords: Linked-Employer-Inventor Data, Record Linkage, Inventors, Patent register data.

Acknowledgments

The authors acknowledge excellent research assistance of Stefan Seth (IAB-FDZ) in the initial preparation of the patent register data using Perl programming to split patent records. Dietmar Harhoff acknowledges financial support from Deutsche Forschungsgemeinschaft (DFG) under the auspices of its Collaborative Research Program SFB/TR 15. Karin Hoisl acknowledges financial support from the Hans Sauer Foundation.

Contents

List of Tables

List of figures

1. Introduction

This report describes the data sources and the generation of a novel linked employer-inventor micro data set, the MPI-IC-IAB-Inventor Data 2002 (MIID 2002). It documents the data generation process related to the current research based on this data, which is jointly carried out by the Institute for Employment Research (IAB) and the Max Planck Institute for Innovation and Competition (MPI-IC) (see Dorner et al. 2014).

The MIID 2002 combines two administrative data sources: patent register data including information on inventors documented on patent filings with the German Patent and Trademark Office (DPMA) in 2002 and labor market biography data of employees. The latter data was provided by the IAB and originate from administrative procedures and social security records. Both data sources contain population data in their scientific field and are frequently used for empirical research.

According to the PatVal-EU survey[1] about 90 percent of inventors surveyed in Germany report themselves as regular employees (Gambardella et al. 2005). Therfore, these individuals should also be recorded in the employment register of the social security system, which represents the matching population of inventors in our project. The connection between individuals represented in both data sources was established using methods of record linkage. Preprocessed name and address data of inventors and corresponding information available for record linking at the IAB served as matching keys.

For 46,180 inventors listed on 42,435 patents, individuals from the IAB administrative data could be matched using both deterministic and distanced based methods of record linkage. Using the number of inventors who are reported with an address in Germany as the basis, we realized a matching rate of 77 percent. This rate compares well with linkage results of similar projects that matched inventors with census micro data in Sweden (Jung and Ejermo 2014) and Finland (Väänänen 2010; Toivanen and Väänänen 2012). Given the high coverage of the initial inventor population, we consider the MIID 2002 a highly representative data basis for studying patenting activities of inventors in the context of regular employment and industrial innovation in Germany.

The added value is generated by the availability of novel variables with thematic relevance for innovation and labor market research. The MIID 2002 enriches patent register data on inventors with accurate longitudinal information about employment biographies, socio-

[1] The PatVal-EU project is an EU wide survey of inventors listed on 10,000 patents filed at the European Patent Office. Further details and results are available in Gambardella et al. (2005) and Guiri et al. (2007).

demographic and job related variables, as well as establishment characteristics, corresponding to the employment episodes of inventors. Vice versa, the empirical analyses of inventors' employment careers based on IAB labor market biographies comprises in the MIID 2002 benefits in at least two important dimensions. Patent register data provides a novel identifier at the level of inventor teams who are jointly listed patent applications. Such information about relationships between workers, either within or across establishments, and related team outcomes are not documented in administrative data of the IAB. Furthermore, counts of patent filings and the corresponding citation metrics represent an additional individual level outcome for empirical analyses. This outcome may be related to and explain the evolution of individual labor market careers of inventors. The scope of variables available in the MIID 2002 is unique for analyses on inventors and stands out against large patent register data sets on inventors. It therefore provides a comprehensive data basis for empirical research at the intersection of labor market and innovation.

Research directions include the analysis of socio-demographic profiles of inventors (for Swedish inventors see Jung and Ejermo 2014). Inventor mobility is another area of research which particularly benefits from the availability of detailed labor market biography data on inventors. The data will be used to investigate the patterns and determinants of inventor mobility across industries, types of establishments and regional labor markets. Moreover, productivity measures derived from patent activities give the opportunity to analyze productivity effects of inventor mobility based on inventive output prior and after a move (e.g., Hoisl 2007; Shalem and Trajtenberg 2009), accounting for the establishment context. Further research is also planned in the field of inventor networks and their evolution.

The remainder of this report is structured as follows: first, the patent register data including information on inventors, which serve as sampling frame for the data generation, are described. Subsequently, we outline the IAB employment biography data which are used as reference data base in the record linkage. Section three elaborates on the record linkage procedures applied in order to link inventors and employees from the available data bases. Section four provides descriptive results of the matched data, focusing both on patent variables and inventor characteristics obtained from the IAB labor market biography data. The report concludes with a summary.

2. Data sources

2.1 Patent Register Data

Our project utilizes of patent register data obtained from the PU Database[2] of the German Patent and Trademark Office ('Deutsches Patent und Markenamt', DPMA) as its sampling frame. These data are made available by the DPMA to researchers and to commercial data providers on an annual basis, initially in ASCII, nowadays in XML format. From all patents filed with the DPMA in 2002 we extracted 50,707 patents (N = 133,959 inventor-patent records) reporting at least one inventor with a German address. These entries represent about 80 percent of all patents filed with the DPMA in 2002 (N= 63,444) according to its annual report (DPMA 2004). The patents excluded from our data basis are filed solely by inventors reported with a non-German or missing residential address. Table 1 summarizes patent and inventor records in our raw patent data sample prior to any deduplication of records.

Table 1: Patent and inventor counts for 2002

Number of Patents	Number of Inventor-Patent records	Average number of inventor records per patent (residential address in Germany only)	Std. Dev.	Min.	Max.
50,707	133,959	2.639	2.144	1	42

Source: PU database 2002 from the DPMA, Authors' own calculations.

We complement the German patent register data at the patent level (using a correspondence table of patent IDs) with information extracted from the PATSTAT database (Version 10/2009) which depicts a "snapshot of the EPO master documentation database (DOCDB) with worldwide coverage [...] including bibliographic data, citations and family links."[3]

The combined patent register data contains the following information:

– Patent application number representing a unique identifier for patent filings. This application number is assigned by the DPMA.

[2] The official forms for patent applications to the DPMA are presented in the Appendix (A1, A2). These materials give an overview on the scope of patent documents. An english documentation with guidlines for patent applicants at the DPMA is provided under http://www.dpma.de/docs/service/formulare_eng/patent_eng/p2791_1.pdf, accessed June 21, 2014.

[3] See http://www.epo.org/searching/subscription/raw/product-14-24_de.html, accessed March 28, 2013.

- Priority date[4] and publication date of the patent application.
- Publication authority of patent (e.g., German Patent- and Trademark Office, European Patent Office).
- Publication kind of patent.
- Inventor names including name and surname (some of the names also include academic titles).
- Residential address of inventors including zip code and place of residence, i.e., city name.
- Patent applicant as stated on the patent.
- Number of citations (PATSTAT Version 10/2009) the patent applications have received from subsequent patents (forward citations) within 3/4/5/10 years after publication of the search report, corrected for equivalents (Harhoff 2009);
- IPC classes, to which the patent was assigned to by the patent office (DPMA).
- Assignment of the IPC classes to 34 technological classes and aggregation into 6 main areas (Schmoch 2008).

Citation data forms a frequently used proxy for the value of a patent (Harhoff et al. 1999). In general, citations reflect both aspects of technical importance as well as economic relevance. Patent offices classify patents according to the "International Patent Classification" (IPC). The IPC contains about 70,000 entries represented by a 7-digit alphanumeric classification. Each patent application is assigned by the patent office to one or more classes corresponding to the technical aspects of the invention (for instance, F03D 1/02 denotes wind motors with a plurality of rotors, typically used in the production of electrical power from wind). A technological classification employing the IPC system is typically based on the information contained in the patent specification as well as the examples, figures and claims attached to the application document (OECD 1994).

Several proposals exist that convert the over 70,000 7-digit IPC classes into a more aggregate classification that is suitable for statistical analyses. In this study, we use the updated classification proposed by the German Fraunhofer Institute for Systems and Innovation Research (ISI) and the French Intellectual Property Institute (INPI) to form largely homogeneous

[4] „The Paris Convention for the Protection of Industrial Property provides that once you file an application in one country party to the Convention, you are entitled to claim priority for a period of twelve months and the filing date of that first application is considered the "priority date." Therefore, when you apply for protection in other member countries (of the Paris Convention) during those twelve months, the filing date of your first application is considered to have "priority" over other applications filed after that date."
See: http://www.wipo.int/sme/en/faq/pat_faqs_q9.html, accessed: April 17, 2014.

technology groups. This classification aggregates the IPC classes to 34 technological fields, for example, telecommunications, information technology, and biotechnology (Schmoch 2008). Depending on the type of invention, patents are allocated to one or several of these 34 classes. To allow an unambiguous assignment of the patent applications to only one of the 34 technology fields, a second variable was generated that translated multiple assignments into fractional weights. For instance, a patent that had been classified into H01P, H01Q (both telecommunications), and H01B (= electrical machinery, apparatus, energy) was assigned to “telecommunications” with a fractional weight of two thirds and to “electrical machinery, apparatus, energy” with a weight of one third. Since the largest fractional weight is in telecommunications, the patent application is assigned to this class. In case a patent received identical fractional weights, one of the respective classes was selected at random.

2.2 IAB Labor Market Biography Data

The best way to observe an inventor’s working career in German administrative data is the use of the Integrated Employment Biographies (IEB) of the Institute for Employment Research (IAB). It is the only micro data set in Germany which comprises a rich set of historic administrative records from different data bases of the social security system that allow to track individual labor market biographies over time. Anonymized individual and establishment level identifiers, both derived from the social security data are included. Confidential information such as the social security number, individual names or address data are separated from the IEB and stored in a secured data base which are not accessible in IAB research data. Each data source of the IEB records the labor market status of an individual in the social security system. Table 2 provides an overview on these data bases comprised in the IEB.

The IEB consolidate individual labor market biographies as recorded in social security registers and administrative data of the Federal Employment Agency (BA) on the basis of an anonymized system-independent individual identifier that is generated at the IAB and is linkable to the social security number. Records from each data source described in Table 2 are measured exactly to the day as spell data[5], including a date of beginning and an ending date. Due to the standardized administrative procedures used for the data collection and due to the relevance of the recorded employment and unemployment periods for individual claims

---

[5] Differently from episodes of non-employment in the IEB, employment spells or episodes are recorded with a maximum duration of one year since employers are obligated to report respectively renew notifications of their employees to the social security authorities on an annual basis. If changes in the employment relationship between employer and employee occur during the year or if the employee separates with the establishment, a new notification has to be filed.

from the social security system (e.g., unemployment benefits, contributions determine future pension claims), the IEB data is considered as highly reliable data.

Table 2: The data sources of the Integrated Employment Biographies (IEB)

Data source	Description	Validity
Employment (Employment History File, BeH)	Employment subject to social security.	As of 1975
	Marginal part-time employment.	As of 1999
Unemployment Benefits (Benefit Recipient History, LeH and Benefit Recipient History SGB II, (X)LHG)	Receipt of unemployment benefits in accordance with Social Code Book III.	As of 1975
	Receipt of unemployment benefits in accordance with Social Code Book II.	As of 2005
Job search (Job-seeking history, (X)ASU**)**	Registered job search with the Federal Employment Agency as a jobseeker in accordance with Social Code Book III.	As of 2000
	Registered job search with the Federal Employment Agency as a jobseeker in accordance with Social Code Book II.	As of 2005
Training measures (Participation-in-measures History File)	Planned or actual participation in an employment or training measure.	As of 2000

Source: IEB, Authors' own depiction.

With regard to the analysis of inventor's labor market biographies the most important features provided by the IAB data are outlined in the following paragraph. Besides the variable reporting the employment status[6] of inventors after a match (e.g., employment, unemployment beneficiaries, job search), employment episodes are the core data source of the IEB. These episodes contain a variety of variables which are particularly informative for the analysis of inventors. These variables include socio demographic characteristics such as gender, year of birth, citizenship or the educational background. Additionally, job characteristics such as occupational codes or gross daily wages are recorded in detail and in a time consistent classification. Location information such as district codes is recorded for the place of residence (since 1999) and the place of work. Another important feature of the administrative

[6] Each employment status is represented by a variety of variables which distinguishes between different types of employment or non-employment, including also further characteristics of the respective status such as different benefits. For an overview on these variables please refer to the SIAB data documentation (e.g., Dorner et al. 2010; Vom Berge et al. 2013).

data of the IAB is that it contains multilevel data, i.e., beyond the level of the individual, rich information on the establishment context of employees is available. The Research Data Center (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) (henceforth: FDZ) supplies the so called Establishment-History-Panel (BHP) (Spengler 2008; Gruhl et al. 2012) which aggregates all individual employment records on the basis of the unique establishment number provided from the annual notifications of employers to the social security system. The BHP includes a record for each establishment that contains at least one employee on the reference date of June 30 in each year as of 1975. Besides structural characteristics of the establishment such as the number of employees, gender, age, educational, occupational or wage structure, that are readily available from the BHP, additionally measures on labor turnover in establishments (inflows and outflows) and information about the appearance respectively disappearance of the establishment identifier can be merged, too. The latter variables classify (dis-) appearance of establishments on the basis of labor turnover measures (Hethey-Maier and Schmieder 2013). Locational information covering both the place of residence and work location are also included in the IAB data and enable research on spatial mobility of inventors and analysis of the embeddedness of inventors in regional systems of innovation.

The employment data included in the IEB do not cover the full working population in Germany. Exceptions are groups of individuals who are exempt from social security contributions and the notification system such as civil servants ("Beamte") or self-employed persons. The share of these subgroups of the total workforce in Germany is about 15 percent (Herberger and Becker 1983). With regard to the targeted population of inventors, these individuals are systematically omitted in the administrative data of the IAB and do not enter the population used for the subsequent record linkage. Surveys of inventors such as the German subsample of the PatVal-EU survey (Gambardella et al. 2005) show that the groups of workers not covered in our social security data comprise about 10 percent of the inventor population. Additionally, civil servants are concentrated in sectors that are seldom involved in patenting activities (e.g., public administration) and account only for a relatively small share of patents. Regular professors however are typically employed as civil servants in Germany and therefore are not covered by our data. In general, our employment data are representative for the core population of inventors working as regular employees in Germany. We therefore claim representativeness only for the system of industrial innovation in Germany, as opposed to the academic system of innovation for which our data is systematically underestimating the number of inventors. For a more detailed description of the data sources of the IEB please refer to the documentation of the Sample of Integrated Labor Market Biographies which represents a 2 percent random subsample of the IEB (see Dorner et al. 2010; Vom Berge et al. 2013).

3. Preparation and Record Linkage of the Datasets

3.1 Record Linkage of patent register and administrative data of the IAB

In general the term record linkage refers to the process of "bringing together information from two records that are believed to relate to the same entity" (Herzog et al. 2007, p. 81). Record linkage thus covers both, duplicate detection within a single file as well as combining multiple data sets from different sources. The latter enables analysts to enrich an existing data base with additional information using other data sets. Therefore record linkage projects establishing links between data are frequent in commercial and academic research.

If exact identifiers (e.g., social security number, business register number) of the units of observation or other clean keys are available, relatively simple routines are applicable to merge records of different data sources. In practice however, data bases rarely have common exact identifiers, particularly when working with micro data at the level of individuals. As matching keys such as names or addresses are error-prone, i.e., subject to typographical errors, misspellings or censoring, more complex procedures[7] from the set of record linkage methods are required.

The aim of the record linkage project described in this report is to enrich data on inventors, extracted from patent documents comprised in the PU database of the German Patent and Trademark Office (DPMA), with labor market biography data of the IAB. Towards this objective, all inventors who have filed at least one patent with the DPMA in 2002 and who are recorded with a valid German address in the patent register represent the sampling population for the record linkage. Patent offices do not assign inventors a unique identifier in their registers and therefore an inventor must be defined based on individual characteristics such as unique combinations of first name and surname.

In absence of a common identifier allowing for a direct link between individuals in both data sources other merging keys are required to establish a match at the individual level. For this purpose, only a small set of potential key variables are collected in both administrative data bases. For instance, the name of the inventor and the residential address are required when filing a patent[8], even though the patent applicant is usually the employing company or organ-

[7] For seminal paper on a theory of record linkage see Fellegi and Sunter (1969).

[8] The inventor(s) (given name, family name, address) shall be named by the applicant without request within a period of 15 months from the filing or priority date (*Erfinderbenennung* according to § 37 PatG). The inventor may

ization. Inventor names and addresses may be linked with corresponding information of employees. Confidential address and name data on employees from social security registers at the Federal Employment Agency are not included in the IEB but are available for record linking. Access to these data is granted only based on a reasoned project application and when the data linkage is done in a secured technical environment assuring data protection of the confidential information. For this purpose, the German Record Linkage Center[9] FDZ was established in 2011. Additionally, confidential data such as names or addresses may only be used for the record linkage itself and has to be deleted after determining the matches. This assures that the actual research is solely based on anonymized data such as the IEB. The address data available at the IAB for record linking was provided by the IAB department 'IT Services and Information Management' (ITM). In the subsequent sections we describe the process of record linkage starting with the pre-processing of the data bases.

3.2 Preprocessing

In the preprocessing steps of a record linkage project, matching keys are prepared in order to enable satisfactory quality of the matches before performing the actual linking. Preprocessing techniques involve the splitting-up of compound identifiers (parsing), the transformation of identifier values into a standard representation (standardization, normalization), and standard data cleaning procedures, such as checks for plausibility and consistency (see Schnell et al. 2003; Herzog et al. 2007).

The major task in our case was to extract the individual information on inventors in a structured format from the patent documents of the PU data. In the original data, information on all inventors involved in each patent filing is stored in a single string variable. We had to split the long string into substrings, each representing an inventor entity. This step was performed utilizing a Perl program and regular expression programming. Additionally, attributes of the inventors as recorded on the patent document (e.g., surname, first name, middle name, city, postal code) were assigned to each inventor record in the restructured data. This step involved a data reshape of all information from a wide format into a long data format, organized as inventor-patent records. Data on the patent itself (e.g., patent ID assigned by the DPMA), the inventors (e.g., name, potential academic title, address) and the patent applicant (either company or individuals with corresponding identifier and address) were maintained with the reshaped information pertaining to the inventors. Table 3 exemplifies the data structure after

---

request that the naming will neither be published nor entered in the Register. This request shall be filed, if possible together with the naming of the inventor on one and the same document. The inventor(s) must however be made known to the DPMA.

[9] For further information on the German RLC see http://www.record-linkage.de; accessed: August 28th 2013.

this initial data cleansing step for five hypothetical patents and ten inventor-patent records.

Table 3: Structure of the data and potential problems

_n	Patent ID	Surname	First name	Title 1	Title 2	Postal Code	Residential location name	Country code	Problem description
1	527	Mayer	Frank Dr.			69115	Heidelberg	DE	Title in first name
2	527	Schmidt	Annemarie	Prof.		69120		DE	City name is missing
3	527	Groß		Dipl.-Phys.	Dipl.-Biol.	69121	Dossenheim	DE	First name is missing
4	527	Müller	Hans Peter	Dipl.-Ing.		69115	Heidelberg	DE	2 names in first name; umlaut (ü) in name
5	53	M E I E R	U L R I C H			D-68169	MANNHEIM	DE	D in postal code; everything in capital letters; additional blanks
6	53	Muster	Roland	Dr.			Barcelona	ES	Not Germany
7	53	Schubert	J. E.	Dr.		D-67063	Ludwigshafen	DE	First name abbreviated; D in postal code
8	107	Schubert	Jörg Erich	Dr.	Ludwigshafen	67063		DE	Same individual as above?; umlaut (ö); names in first name; city in second title
9	128	Van de Velde	Edward	Prof. Dr.		81477	Munich	DE	City name in English; name affix
10	486	Gonzalez-Gomez	Prof. Dr. DR. h.c.	Pablo		01069	Dresden-Mitte	DE	Urban district in city; first name and title reversed; not common German name

Source: Authors' own depiction.

Each inventor listed on a patent that was filed in the year 2002 is stored in a single record and name, first name, titles and the components of address are separately stored as corresponding variables. The last column which is not part of the actual data set lists potential problems that may emerge because of unclean data or the availability of information on the

inventor. Therefore in some cases the Perl routines did not work properly separating the available information into distinct name and address variables. These problems including possible solutions to fix them will be discussed below.
Before any further processing steps were performed, inventors with a country code other than “DE” (representing Germany as country of residence) were dropped (e.g., record 6, ID 53 in Table 3), since the address data on employees from the social security system solely includes addresses in Germany. After the deletion, 63,257 inventors (each combination of name and surname represents one individual) recorded in the patent data were retained. The remaining population of deduplicated inventors and addresses of individuals recorded in the address history file for 2002[10] represent the sample used for further data cleansing and subsequent record linkage procedures. The address data available from administrative records on individuals in the IEB (names, cities or postal codes) are structured similarly to the records listed in Table 3 and are also subject to the same data quality issues (abbreviations, umlaute, special characters etc.) with regard to the required preprocessing tasks.

Using these string variables available from both data bases as matching keys, the record linkage is prone to errors. Generally, these problems concern standardization and parsing. Standardization includes the removal of name suffixes and local district add-ons to city names, e.g. “Mitte” in “Dresden-Mitte”, name affixes or titles such as “van de” in “van de Velde” (records 9 and 10), or the substitution of umlauts and special characters such as “ß” e.g., in German names like “Groß” and “Müller” (records 3 and 4). The most common standardization routines that were applied are listed in Table 4.
Parsing divides information into components, e.g., separating academic titles from names (record 1 in Table 3) or dividing a first name that consist of two parts into a first name and a middle name (records 4 and 8 in Table 3). Parsing also covers the correction of falsely assigned information, e.g., in record 7 (Table 3) where the variable city name is recorded in the title variable and in the last line, where first name and title are interchanged. After performing the standardization and parsing routines, the strings are far better comparable and the realization of true matches is more likely than before. By applying the same standardization and parsing procedures as described in Table 4 on both data sets, some observations proved to be corporate applicants or duplicates and thus were consolidated or excluded. Standardization and the correction of typographical errors or abbreviations also reduced the number of

[10] In cases of invalid address information for 2002, valid information stemming from the previous year was used.

unique addresses[11]. Eventually, a total number of 59,694 unique inventor's records who were involved in patent applications in 2002 remain in our final inventor sample.

Table 4: Standardization of addresses

Variable	Preprocessing step
first name, middle name, surname, city	substitution of ß and umlauts like ä, ü, ö substitution of accents transformation of all letters to upper case removal of blanks (leading, trailing, and other) removal of separators (. ; : ; - ' / () &) removal of titles and name affixes removal of all (except for A-Z) special characters removal of district add-ons to city names
postal code	substitution of missing codes with blank removal of all (except for 0-9) special characters

Source: Authors' own depiction.

3.3 Record Linkage

The actual record linkage procedures were performed using the software "Merge Toolbox" (MTB) (Schnell et al. 2004). By using string-comparison-functions the similarity of the matching keys of inventors and individuals recorded in the social security data is computed. The toolbox provides a set of string functions that enable the user to choose appropriate matching procedures given the underlying data.

The first step was an exact record linkage on all key variables (deterministic matching), in our case the inventor's first name, middle name, surname, name of city and postal code.

In subsequent steps, this deterministic approach of record linkage was gradually relaxed towards string comparisons of names and addresses which differ due to typographical errors or missing information in one key variable[12]. In such a case, a set of keys would again be matched exactly, with the exception of a single variable that would be either omitted or compared by string comparison functions, such as the Jaro- or the Jaro-Winkler metric (Jaro 1989, Winkler 1990). These metrics calculate a measure that would be used to evaluate the quality of the match under a given configuration of the MTB. The Jaro (1989) string comparison returns a metric that gives values of partial disagreements between different strings. It is

[11] The IAB data contains multiple addresses for individuals who moved their residence during the year of 2002. Additionally, minor typographical changes of the address information can result in multiple records. The number of address records was reduced from the initial volume by deleting duplicates and records with missing information either on the name or address of the employee.

[12] For a summary on the record linkage methodology and an overview of applications see Winkler (1995).

based on the number of insertions and deletions of characters weighted with the length of the strings until two strings correspond with each other perfectly, e.g. "Belrin" and "Berlin". This computed score always ranges from 0 to 1, with the value of 1 being an exact match. As an extension to the Jaro metric, Winkler (1990) suggested to put more weight on the initials of names since data quality of subsequent characters deteriorates monotonically towards the end of a string. The outlined stepwise record linkage procedure will be discussed in detail below with an explanation of the specific metrics that were applied.

The postal code (and further substrings of the five digit postal code) is considered as reliable information and was used as blocking variable in all steps. Blocking is a common approach in record linkage projects to reduce the search space for each record of the source file and achieve a significant reduction of the computing time, given the assumption that true matches only occur within a value of the blocking variable (see Herzog et al. 2007).

Table 5: Record linkage procedure

Steps	Blocking variable	Exact	Omitted	String comparator function (besides exact matching)
1	postal code (blocking on various digit levels)	first name middle name surname name of city		
2		first name middle name surname		name of city: Jaro
3		first name surname name of city	middle name	
4		first name surname	name of city	middle name: Jaro
5		surname	name of city middle name	first name: Jaro
6		first name	name of city middle name	surname: Jaro-Winkler

Source: Authors' own depiction.

As described above, the first steps of the record linkage compare the full set of variables exactly. This is the most restrictive way to link the data but the matches obtained are optimal by definition. Step two relaxes this restrictive linkage procedure by comparing the name of the city not exactly as before, but using the Jaro metric in a distance based matching approach. City name was chosen as the first variable that is not considered exactly since it is not as important as the name of an individual and thus little differences in the spelling do not reduce the quality of a match considerably, particularly since we still require the postal code to match exactly. In each of the subsequently listed steps in Table 5, the MTB created an output file including all matched pairs between inventors and employees and a score indicating the quality of the match. Using this measure, a threshold specific to each step was defined, which eventually classifies a record pair as a match. Afterwards the quality of the matches was verified manually, based on the similarity score. A small number of matches were reclassified from a non-match to a match respectively from a false positive match to a non-match. Finally an inventor ID was generated that allows matching individuals, i.e., inventors and employees, in both data bases.

3.4 Results of the record linkage

The number of matches after each linkage step is listed in Table 6. The numbers represent matched records after the manual verification of the linkage and account for (potential) corrections.

Table 6: Results of the record linkage

Total number of unique inventors before linkage		**59,694**	
Procedure		**Found**	**Percent**
1	All exact (incl. all exact with reversed first and middle name)	43,182	72.34
2	All exact, city name Jaro	2,312	3.87
3	All exact, middle name omitted	1,488	2.49
4	All exact, city name omitted, middle name Jaro	115	0.19
5	All exact, first name Jaro	651	1.09
6	All exact, surname Jaro-Winkler, city name + middle name omitted	211	0.35
Total number of inventors found in employee data (after verification)		47,959	80.33
Consolidated duplicates in verification step		198	0.32
Total number of matched inventors after verification		47,761	80.01
Unsolved multiple assignments (1:m inventor <> individual in IAB data)		1.581	2.65
Unique matches (1:1 inventor <> individual in IAB data)		**46,180**	**77.36**

Source: Authors' own depiction.

Each inventor who could be disambiguated on the basis of the name information from the patent data was assigned a unique person identifier. Duplicates occur when an inventor is linked to the same individual in several steps, e.g., due to appearance in both databases with two different addresses. The duplicates were consolidated manually.

In total 47,761 inventors originating from the processed patent data were found in the IAB data and classified as matches given the name and address information. When using the deduplicated number of inventors recorded with a German residential address as the reference, the realized matching rate is about 80 percent. The corresponding number of patents is 43,076 (85.0 percent). Additionally 1,581 inventors (2.64 percent of the initial population) could not be disambiguated and remain with links to multiple individuals recorded in the IEB. Reducing the sample by these ambiguous matched inventors yields a matching rate of 77.36 percent and a sample comprising 46,180 inventors (42,435 patents).

In a similar record linkage project, Jung and Ejermo (2014) realized a matching rate of 78.9 percent when matching inventors in Sweden listed on a patent filed with the European Patent office against individuals recorded in Swedish social security data. Väänänen (2010), respectively Toivanen and Vänäänen (2012) who employ a similar data basis in their research, achieved a matching rate of 73 percent between Finish inventors recorded in the NBER Patent data[13], when matching them to linked employer-employee register data of Statistics Finland.

[13] For a description of the NBER Patent Data please refer to Hall et al. (2001) or Jaffe and Trajtenberg (2002).

4. Description of MPI-IC-IAB-Inventor Data 2002 (MIID 2002)

The following chapter provides descriptive statistics on the MPI-IC-IAB-Inventor Data 2002 (MIID 2002) obtained from the record linkage. First, the analysis elaborates on the patent data included in the MIID 2002. In the second part of the descriptive analysis, inventor characteristics obtained from the IAB labor market biography data are outlined.

4.1 Patent data

The descriptive analysis uses an extract of the matched inventor-employee population which excludes patents without a single matched inventor (i.e., fully incomplete matched patents) and patents with multiple inventor-employee matches who could not be disambiguated. After consolidation, the data comprises a total number of 40,268 patents (94.9 percent) filed in 2002. At the individual level, these patents are filed by 44,206 inventors (95.7 percent), each with a unique combination of surname and first name.

Complete and partial matches of patents

At the level of patents, a subsample of 30,215 patents (75 percent) has been matched fully, i.e. all inventors listed on the patent are unambiguously matched with an employee in the IEB. Additionally, there are partial matches in which the number of matched inventors does not correspond with the original number of contributors. Since we are only able to match German social security data, inventors with a foreign residential address are systematically omitted. However, for 75 percent of the patents including both inventors with foreign and German address, we were able to match the full number of the German inventors (1,719 out of 2,308 patents). Patents with a contribution of individuals not covered by the social security data (e.g., civil servants, self-employed) may also yield partial matches. Differently from inventors with a non-German address however, we are not able to distinguish between reasons why a match could not be established.

Inventors per patent

The size of inventor teams involved in patent filings recorded in the MIID 2002 is summarized in Table 7.

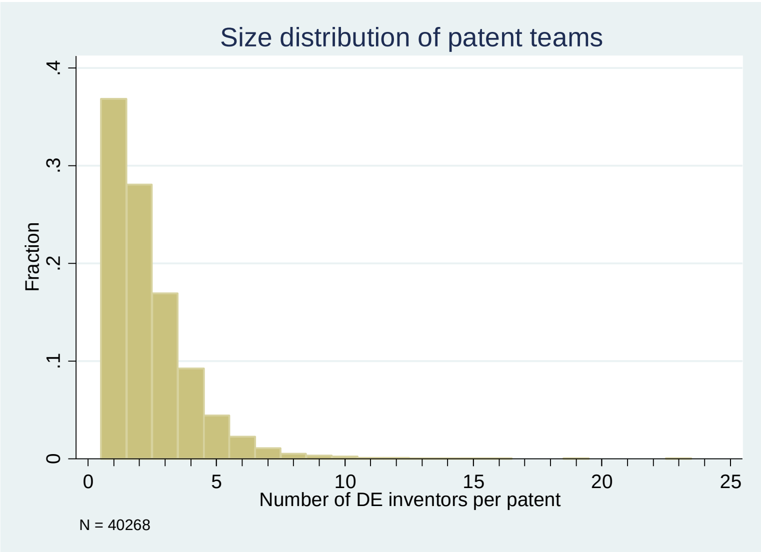
Table 7: The size of inventor teams

Variable	Obs.	Mean	Std. Dev.	Min	Max
Inventors (all patents)	40,268	2.351	1.583	1	23
Inventors (completely matched patents only)	30,215	2.029	1.330	1	16

Source: MIID 2002. Authors' own calculations.

Figure 1 depicts the detailed distribution of the team sizes. According to the distribution of team sizes, about 36 percent of all patents in our sample are filed by a single inventor. The remaining team patents comprise up to a maximum number of 23 inventors. On average, patents in our matched sample are filed by 2.3 inventors, as opposed to 2.6 inventors and found in the raw data prior to the record linkage (see Table 1).

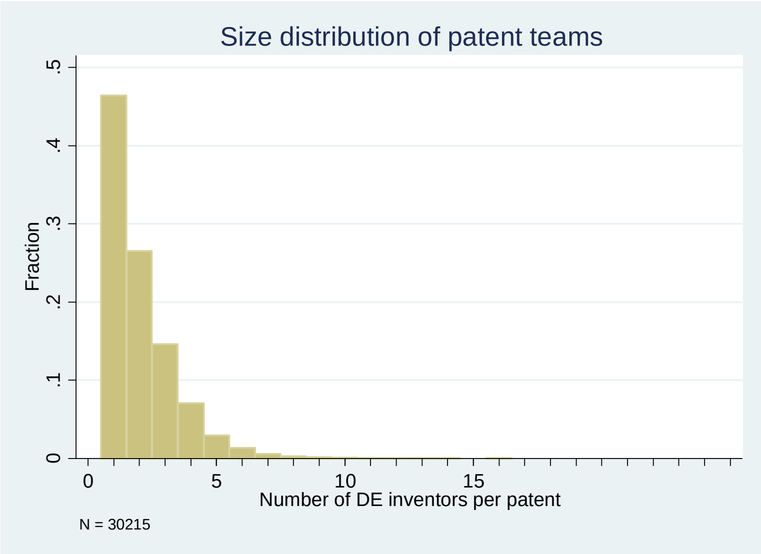
Figure 1: Distribution of inventor team size – all patents



Source: MIID 2002. Authors' own calculations.

If we consider only the 30,215 patents in which all involved inventors are matched with employees (completely matched patents), the share of single inventor patents increases to roughly 46 percent (see Figure 2). This is due to the fact that with increasing size of the inventor team, the likelihood for a non-matched contributor is also increasing. As a result, the average size of patent teams decreases to 2.02 inventors and the maximum number of inventors listed on a completely matched patent decreases from 23 to 16. In total, 16,191 patents (54 percent) are filed by more than one inventor.

Figure 2: Distribution of inventor team size– completely matched patents



Source: MIID 2002. Authors' own calculations.

Patent citations

The number of citations of patents is often used as a measure of a patent's technological relevance or even as an approximation of the commercial value of a patent (e.g., Griliches 1990; Harhoff et al. 1999).

From the sample comprised of all patents, 31.4 percent (12,660 patents) received at least one citation from another subsequent patent. The corresponding forward citation measure for the sub set of completely match patents is 32.0 percent (9,687 patents). Tables 8 and 9 report citation statistics of the patents included in the MIID 2002, using different time windows after the patent filing in which forward citations may occur.

The full set of measures indicates that with regard to citations both samples of patents – the subsample of completely matched patents and the full set of patents – are basically consistent without any significant variation. The temporal citation patterns recorded at three, four and ten years after the application are also very similar. The propensity of a patent to be cited evolves over time after its filing from about 20 percent, three years after the filing, up to roughly 30 percent after ten years. The citation counts exhibit that on average a patent is cited by 0.75 other patents. The maximum number of citations and potentially most valuable patent in the MIID 2002 sample of 2002 accounts for 57 citations within a ten year period after its filing.

Table 8: Descriptive statistics on forward citations of all patents

Variable	Obs.	Mean	Std. Dev.	Min	Max
Number of citations total	40,268	.741	1.666	0	57
Number of citations 3 years after patent filing	40,268	.336	.895	0	27
Number of citations 4 years after patent filing	40,268	.494	1.188	0	28
Number of citations 10 years after patent filing	40,268	.740	1.666	0	57
Dummy: citation of patent	40,268	.314	.464	0	1
Dummy: citation of patent within 3 yrs	40,268	.196	.397	0	1
Dummy: citation of patent within 4 yrs	40,268	.251	.434	0	1
Dummy: citation of patent within 10 yrs	40,268	.313	.464	0	1

Source: MIID 2002. Authors' own calculations.

Table 9: Descriptive statistics on forward citations of completely matched patents

Citations/ Patents	Obs.	Mean	Std. Dev.	Min	Max
Number of citations total	30,215	.756	1.696	0	57
Number of citations 3 years after patent filing	30,215	.339	.900	0	27
Number of citations 4 years after patent filing	30,215	.501	1.198	0	28
Number of citations 10 years after patent filing	30,215	.755	1.696	0	57
Dummy: citation of patent	30,215	.321	.467	0	1
Dummy: citation of patent within 3 yrs	30,215	.199	.399	0	1
Dummy: citation of patent within 4 yrs	30,215	.256	.436	0	1
Dummy: citation of patent within 10 yrs	30,215	.320	.466	0	1

Source: MIID 2002. Authors' own calculations.

Technological field of patents

Examiners at the patent offices classify the technological focus of a patent in high detail using International Patent Classification (IPC) codes (see). Schmoch (2008) proposes a categorization of IPC codes into 34 technological fields which are frequently used for statistical analysis of patent data. These technological fields can be further aggregated into five broader technological areas. The distribution of patents in the MIID 2002 over these technological areas is presented in Table 10. The largest share of patents is assigned to technologies comprised by the field of “Mechanical Engineering” (41 percent). The fields of “Electrical Engineering” and “Chemistry” are equally important accounting for about 19 percent of the patents. “Instruments” account for a share of 12 percent while other fields represent the rest (7 percent). The technological focus of the patents included in the MIID 2002 matches quite well with the specialization of German manufacturing sector and its characteristic (export-) products (e.g., machinery, motor vehicles, automotive parts, chemistry).

Table 10: Distribution of patents classified in main technological areas

Main technological areas	Freq.	Percent
Electrical Engineering	7,647	18.99
Instrument	5,206	12.93
Chemistry	7,664	19.03
Mechanical Engineering	16,749	41.59
Other Fields	2,912	7.23
NA	90	0.22
Total	**40,268**	**100.00**

Source: MIID 2002. Authors’ own calculations.

Productivity of inventors

The MIID 2002 is structured by inventors within patents and therefore allows for further descriptive analysis at the individual level. Across all patents, the data includes 44,206 unique inventors. While socio demographic characteristics on these inventors are only available from the IAB data[14], the patent data allows computing the inventive productivity of inventors in 2002. Inventive productivity of inventors can be calculated from patent counts per unique inventor and citation counts for the patents. These indicators represent informative non-labor market outcomes and may be related to the evolution of individual employment biographies.

Table 11: Descriptive statistics of patenting activities per inventor in 2002

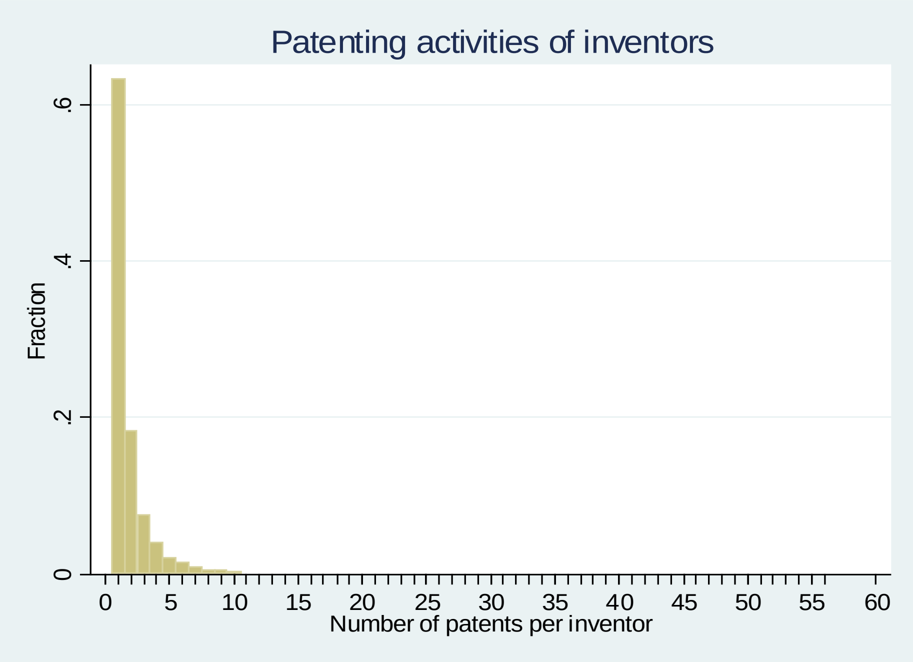
Variable	Obs.	Mean	Std. Dev.	Min	Max
Number of patents per inventor in 2002	44,206	1.9108	2.054	1	56

Source: MIID 2002. Authors’ own calculations.

[14] For a detailed descriptive analysis of the socio demographic characteristics of inventors included in the MIID 2002, see Dorner et al. (2014).

As can be seen in Table 11, each inventor comprised in the MIID 2002 contributed on average to about 1.9 patent filings in 2002. The most productive inventor was involved in 56 patents[15]. Productive inventors with equally high numbers of patents however are an exception as the skewed distribution of patent productivity (Figure 3) depicts.

Figure 3: Distribution of patenting activities per inventor in 2002



Source: MIID 2002. Authors' own calculations.

The majority of inventors, about 63 percent, are only involved in the filing of a single patent in the focal year 2002. With regard to the number of patents, it makes a difference whether inventors patent on their own, or whether a patent was developed from joint inventive activities with co-inventors. Since the patent register data does not report information about actual individual contributions of inventors to patents, we assume an equal share for each inventor who was involved in a patent filing. Table 12 provides descriptive statistics on the fractional counts of patenting activities of inventors.

Table 12: Fractions of inventor's contributions to patents in 2002

Variable	Obs.	Mean	Std. Dev.	Min	Max
Inventor's fractional counts of patent contributions	44,206	.832	1.009	.0435	39

Source: MIID 2002. Authors' own calculations.

[15] This figure needs to be interpreted with caution since frequent combinations of names may yield an inflation of individual patent portfolios.

Compared to the statistics above, we see a decline of the measures, since co-patenting is very common. On average, the inventors comprised in our sample contribute to 0.83 fractionalized patents in 2002. The maximum individual contribution amounts to 39 inventor-patent shares. At the bottom of the distribution, the lowest individual contribution represents only 0.04 percent. This fraction results from a collaborative patent of the inventor shared with 22 co-inventors in the largest patent of the data. Instead of counting patent filings at the individual level, the quality of patents measured by citations of individual patents by other patents can be calculated from the MIID 2002. Both indicators are assumed to be a good proxy for the individual value of an inventor for the employing firm or on the labor market in general. As before, we document citation counts and fractional citation counts on the inventor level.

Table 13: Forward citations for patents filed in 2002 per inventor

Variable	Obs.	Mean	Std.Dev.	Min	Max
Citations of patents per inventor	44,206	.6139	1.065	0	28
Fractional counts of citations of patents per inventor	44,206	.411	.736	0	16.233

Source: MIID 2002. Authors' own calculations.

Inventors realize on average 0.61 citations for their patents filed in 2002 (see Table 13). The number of citations evolves over time and reaches the latter value of about 0.6 at about ten years after the patent filing. Fractional counts account for co-inventors who were involved in the filing of the cited patents. The average fractional counts for citations amount 0.41 per inventor for the whole time period covered by our data.

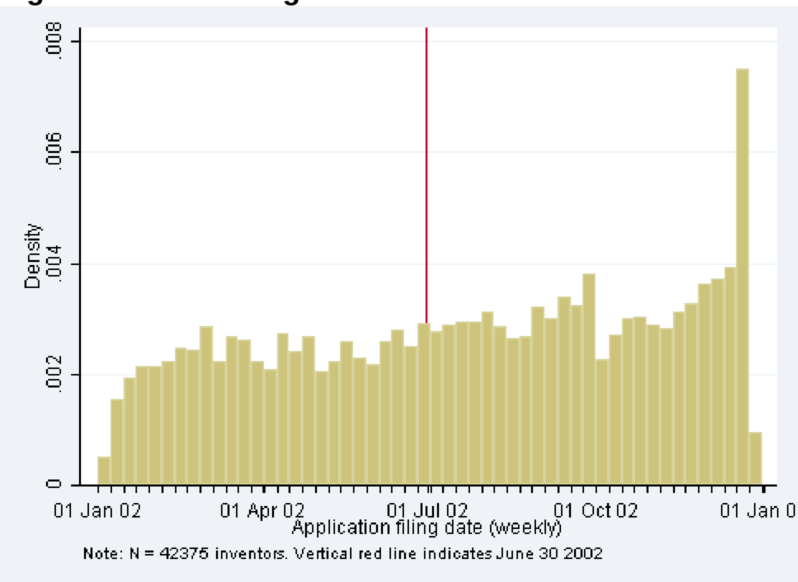
The presented patent and citation counts of inventors refer only to patents filed in 2002. Both measures however may be extended to other years in which inventors are recorded as active inventors in patent register data such as PATSTAT.

4.2 Labor market biography data

The second part of the descriptive analysis provides an overview of inventor characteristics originating from the IAB labor market biography data included in the MIID 2002. An advantage of the MIID 2002 is that high quality register information on socio demographic and job characteristics of inventors become available which are not included in patent register data.

The sampling frame for the subsequent analysis at the individual level is restricted to unique inventors from the matched population in the MIID 2002 who are reported with an episode in the IEB, overlapping any of the patent filing dates recorded in the patent register data for the focal year 2002. This condition is fulfilled for 42,375 inventors, representing 91.7 percent of the unique individuals in the matched sample. Since inventors may file several patents during an annual employment episode with a firm, or inventors may file multiple patents with different employers within a year, the last patent registered in 2002 was chosen as the focal date for the descriptive analysis. Figure 4 displays the distribution of the application dates of the focal patents used for the descriptions.

Figure 4: Patent filing dates



Source: MIID 2002. Authors' own depiction.

Additionally, the subsequent analysis was restricted to the main occupation[16] in order to avoid double counts of parallel episodes at the individual level (e.g., simultaneous employ-

[16] The main occupation is defined on the basis of a sorting of spells which assigns the highest priorty to a regular employment spell with the highest wages. Further information on the underlying algorithm used to define the main occupation is provided by Vom Berge et al. (2013).

ment and registered job search). The remaining population of inventors not included in the descriptive analysis does not exhibit a data spell overlapping a patent in 2002 or their labor market biography does not include any valid IEB episode in 2002. The descriptive analysis of the MIID 2002 presented below is complemented with inventor statistics reported in the German subsample of the PatVal-EU Survey (Gambardella 2005; Guiri et al. 2007) and with general characteristics of employees in Germany recorded in the SIAB[17] data on June 30 2002 (Dorner et al. 2011; Vom Berge 2013).

Employment status of inventors

Based on the preprocessed MIID 2002 data, Table 14 reports the employment status of inventors at the time of the patent filing with the DPMA in 2002.

Table 14: Employment status of inventors in 2002

Employment status of inventors	MIID 2002	
	Freq.	**Percent**
Employees (BeH)	41,862	98.79
thereof: Regular Employees	*40,654*	*97.11*
Other employees (e.g., trainees, retirees, marginal employees)	*1,208*	*2.89*
Beneficiaries (LeH)	376	0.89
Jobseeker (ASU)	81	0.19
Others (in particular measures)	56	0.13
Total	42,375	100

Source: MIID 2002. Authors' own calculations.

The distribution shows that the large majority of 98.79 percent of the inventors in the MIID 2002 are recorded as employees at the time of the focal patent in 2002. Analyzing the employment status in detail reveals that only few inventors (2.89 percent) are reported with a special type of employment status such as trainees, retirees or marginal employees. The remaining small fraction of inventors who are not registered employed (about 1.1 percent) are recorded mostly as beneficiaries from the unemployment insurance system.

Gender

According to the MIID 2002, only percent 5.2 percent of the inventors are women (Table 15).

---

[17] The Sample of Integrated Labor Market Biographies (SIAB) is a random 2 percent sample of the IEB and it is representative for the population of employees in Germany on June 30 2002 (Dorner et al. 2011; Vom Berge et al. 2013).

Although the share of women among inventors is about three times larger than the corresponding figure in the PatVal-EU study, the underrepresentation of women in patenting is still striking. Ejermo and Jung (2014) show that share of Swedish women increased substantially from less than 2.4 percent in 1985 to just over 9 percent in 2007. In recent years there has been increasing interest in the question why women patent less than men (e.g., Hunt et al. 2013).

Table 15: Inventor Gender

Gender	MIID 2002		PatVal-EU	SIAB
	Freq.	Percent	Percent	Percent
Male	40,178	94.82	98.36	52.83
Female	2,197	5.18	1.64	47.17
Total	**42,375**	**100**	**100**	**100**

Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013), PatVal-EU (Gambardella et al. 2005). Authors' own calculations.

Age

In the MIID 2002 inventors are on average 41.4 years old, while employees recorded in the SIAB are on average aged 39.9 years in 2002. Based on a matched sample of Swedish inventors who filed patents with the European Patent Office, Ejermo and Jung (2014) report an average age of about 44 years in their data ranging from 1985-2006. Over time however, they show in their Swedish data, that inventors tend to get younger at their first patent filing.

Table 16: Age distribution of inventors

Age groups	MIID 2002		PatVal-EU	SIAB
	Freq.	Percent	Percent	Percent
<= 30 yrs.	3,546	8.37	2.72	24.34
31-40 yrs.	18,576	43.84	31.40	28.23
41-50 yrs.	12,385	29.23	27.20	25.41
51-60 yrs.	6,358	15	32.00	17.23
61-70 yrs.	1,485	3.5	6.29	4.36
>= 71 yrs.	25	0.06	0.39	0.43
Total	**42,375**	**100**	**100**	**100**

Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013), PatVal-EU (Gambardella et al. 2005). Authors' own calculations.

Since our data includes only the cohort of inventors in 2002, a comparative statistic with a time trend cannot be provided. A comparison on the basis of six age groups of inventors with the PatVal-EU survey and the representative sample of employees recorded in the SIAB is presented in Table 16. The analysis shows that all groups of inventors younger than 50 years of age tend to be overrepresented in the MIID 2002 while the share of older inventors is lower than in the PatVal-EU study. A reason for the difference is most likely related to the coverage of the IAB employment data. For instance patenting professors employed as civil servants, who also tend to be overrepresented in the older age groups, are not covered at all. Relative to the SIAB the age distribution of inventors illustrates that inventors with an age younger than 30 years are seldom, but the group of 31-40 years old inventors is highly overrepresented compared to the age distribution in the SIAB.

Nationality

The MIID 2002 sample does only include inventors who are reported with a residential address in Germany according to the patent document. Foreign inventors residing in Germany are of general interest as they might serve as bridges of socio cultural capital or they might be important agents for the introduction of foreign products in Germany. Analyzing the nationality of inventors as it is reported by employers in the IEB, we find that the share of foreign inventors amounts to 3.6 percent. This figure is less than half of the share of foreign workers in the full population according to the SIAB (7.5 percent).

Education

Education is at the core of the discussion on personal characteristics of individuals when policy measures are targeted on an increasing propensity to engage in inventive activities (Väänänen 2010). For inventors in the MIID 2002, information on educational attainment of inventors is reported in detail in the labor market biography data of the IAB. For all 41,862 employed inventors six education categories at the corresponding patent filing dates are displayed in Table 17. The tabular analysis displays that inventors on average have received higher education than the population average of employees as found in the SIAB. According to the IEB the majority of inventors are reported graduates from either a university of applied sciences (‘Fachhochschule’) or a regular university. University graduates also represent the largest group among the matched individuals (about 47 %). Almost 70 percent of the individuals in the MIID 2002 hold an academic degree obtained from a technical college or a university. The finding that highly skilled employees dominate the population of inventors compares quite well with other inventor data sets. The corresponding statistic in the PatVal-EU study, i.e. the share of inventors with tertiary education is 84 percent (Gambardella et al. 2005), while in the full population of employees only about 8 percent are reported to hold an academic degree. Besides the majority of highly skilled inventors there is also a notably sized population of about 24 percent of the individuals in the MIID 2002 who graduated from the German vocational system ('Duales System'). Only few inventors are reported as low skilled, i.e., without secondary schooling or unknown education. In the latter cases, erroneous information might at least explain some of these cases with unknown education. The number of inventors with missing information is relatively low[18].

Table 17: Education and training

Education and training	MIID 2002		SIAB
	Freq.	Percent	Percent
Secondary / intermediate school leaving certificate without completed vocational training	329	0.79	15.63
Secondary / intermediate school leaving certificate with completed vocational training	8,234	19.67	53.4
Upper secondary school leaving certificate without completed vocational training	472	1.13	1.84
Upper secondary school leaving certificate with completed vocational training	1,860	4.44	3.68
Completion of a university of applied sciences	10,200	24.37	2.98
College / university degree	19,466	46.5	5.04
NA	1,301	3.11	17.43
Total	**41,862**	**100**	**100**

Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013). Authors' own calculations.

Occupation

The availability of detailed occupational codes from the employment register is a major advantage of the MIID 2002. Occupational information is reported in detail according to the occupational classification "Klassifikation der Berufe 1988" (KldB 88). Electrical, motor and chemical engineers are the top three ranked occupations among inventors accounting jointly for about 42 percent of all inventor occupations. Blossfeld (1987) used the detailed KldB 88 classification to derive a more general typology of 12 occupational fields. Table 18 compares the distribution of inventors in the MIID at the time of patent filings in 2002, with the population of all employees in 2002 recorded in the SIAB. The comparison of the occupational distributions shows that inventors are highly overrepresented in technical and engineering jobs (76.90 percent vs. 6.94 percent). Another important group of inventors is working in management positions at the time of patent applications (8.38 percent). These three groups jointly represent about 89 percent of the MIID 2002 records, but only about 9 percent of employees in the SIAB. Among the inventors who are not reported with occupations in these three

[18] Note that this analysis does only use the cross section of the SIAB data and routines to correct for potential inconsistencies along the individual employment biography (e.g. Fitzenberger et al. 2006) are not applied.

fields, only a small minority are reported to work explicitly in occupations with low skill profiles (e.g., simple manual occupations, simple services, simple commercial and administrative occupations).

Table 18: Occupational fields

Occupational fields	MIID 2002		SIAB	
	Freq.	Percent	Percent	Median wage (€)
Agricultural occupations	24	0.06	1.59	44.6
Simple manual occupations	461	1.10	12.06	70.4
Skilled manual occupations	1,064	2.54	14.01	71.6
Technicians	8,366	19.98	4.28	103.9
Engineers	23,828	56.92	2.66	139.0
Simple service	172	0.41	16.45	33.6
Qualified service	254	0.61	5.66	47.4
Semi-professions	88	0.21	7.00	71.3
Professions	1,257	3.00	1.79	105.1
Simple commercial and administrative occupations	316	0.75	10.02	41.9
Qualified commercial and administrative occupations	2,388	5.70	20.78	76.8
Manager	3,507	8.38	2.59	122.5
NA	137	0.33	1.11	21.8
Total / Overall median wage	**41,862**	**100**	**100**	**64.91**

Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013). Authors' own calculations.

Wage

The availability of detailed wage information is a particular advantage of administrative data such as IAB employment biography data. The match between inventors and individuals in the IEB provides a link to precise and very detailed wage information from the social security system. When comparing the wage levels of the occupational fields listed in Table 18, it is evident that inventors work significantly more often in jobs which tend to be paid above average. Engineers, the dominating occupational group of inventors, receive the highest wage rewards in their jobs (139.0 €). Management jobs (122.5 €) and technical occupations (103.9 €) are ranked runner up respectively fourth in the hierarchy of occupational fields. When focusing on individual wages, the average wage of inventors is 137.19 Euro compared to 65.63 Euros in the full population of employees on June 30 2002 (Table 19).

Table 19: Gross daily wage and top coding of wages in 2002

	Obs.	Mean	Std. Dev.	Median	Perc. top coded
MIID 2002: Gross daily wage	41,862	137.19	23.23	147.95	67.25
SIAB: Gross daily wage	613,608	65.63	42.12	64.91	6.55

Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013). Authors' own calculations.

A shortcoming of social security information on wages is that wages exceeding the social security contribution assessment ceiling (in 2002: 147 Euros in East Germany; 123 Euros in West Germany) are only reported with the value of this threshold. In the case of inventors, about 67 percent of the recorded wages are right censored. The median wage of inventors in the MIID 2002 is also located just at the value of the contribution ceiling. In the comparison group of all employees recorded in the SIAB, wages only seldom exceed the contribution ceiling (6.55 percent).

Work and residential locations

Both residential and work locations of inventors are recorded in the MIID 2002 data with the accuracy of the level of municipalities as the smallest administrative units in the data. For brevity, the distribution of inventors' locations at the level of federal states ('Bundesländer') is presented in Table 20 and compared with the corresponding figures of the SIAB.

Table 20: Inventors' locations at the level of German federal states

	Place of residence			Place of Work		
	MIID 2002		SIAB	MIID 2002		SIAB
	Freq.	Percent	Percent	Freq.	Percent	Percent
Schleswig-Holstein	709	1.69	3.28	574	1.37	3.04
Hamburg	522	1.25	2.05	808	1.93	2.67
Lower Saxony	2,994	7.15	9.40	2,742	6.55	9.00
Bremen	103	0.25	0.77	188	0.45	1.02
North Rhine-Westphalia	7,626	18.22	21.5	7,705	18.41	21.86
Hesse	3,438	8.21	7.44	3,601	8.6	7.84
Rhineland-Palatinate	1,999	4.78	4.91	2,005	4.79	4.49
Baden-Wuerttemberg	10,271	24.54	13.56	10,353	24.73	14.05
Bavaria	9,803	23.42	15.82	9,668	23.09	16.04
Saarland	285	0.68	1.20	268	0.64	1.30
Berlin	1,124	2.69	3.5	1,224	2.92	3.75
Brandenburg	477	1.14	3.02	336	0.8	2.62
Mecklenburg-Western Pomerania	125	0.3	2.06	11*	*	1.91
Saxony	1,331	3.18	5.16	1,294	3.09	5.04
Saxony-Anhalt	302	0.72	2.97	257	0.61	2.72
Thuringia	723	1.73	2.92	725	1.73	2.66
NA	30	0.07	0.43	<3	*	0.01
Total	**41,862**	**100**	**100**	**41,862**	**100**	**100**

Note: "*" and "<3" indicate anonymized data cells.
Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013). Authors' own calculations.

The states of Baden-Wuerttemberg and Bavaria in southern Germany are ranked on top with regard to the share of full inventor population in the MIID 2002. While both states are highly overrepresented in their shares, compared to the fraction of general employment, the largest state labor market in Germany, North Rhine-Westphalia, is underrepresented with regard to its inventor share. The remaining federal states are represented below their employment proportion in the MIID.

Establishment level information

The Establishment History Panel (BHP) was merged with the individual employment biography of inventors in the MIID 2002. This link provides access to rich and high quality information about an inventor's employer and the structure of coworkers at the plant level for each year. The connection between linked employer-employee biography data available at the IAB-FDZ and inventor data is unique for Germany. The subsequent descriptive analysis outlines some establishment characteristics of inventors' employers at the time of the patent filings in the MIID. Again, the descriptive findings are presented in comparison to the population of employees in the SIAB data on June 30 2002 and their establishments.

Age and size of establishments

The age of an establishment is an informative indicator for inventors' establishments since many young firms are linked to innovative business ideas or novel products and services. As filing patents involve direct costs and being inventive as an organization requires R&D expenditures, older establishments and mostly larger firms are another group of organizations where inventors are assumed be frequently employed.

To assess the age of an establishment, the BHP includes the date respectively year when an establishment for the first time reports at least one employee to the social security administration. If the inventor is the only employee, the first year of an inventor's employment according to the biography data represents the year of foundation. Due the left censoring of the social security data in 1975, firms which existed before that year are reported with a foundation year in 1975. The share of inventors working for young firms with less than three years of age is only 7.48 percent of inventors. In its magnitude, this finding corresponds well to the overall distribution of employees on young firms in the SIAB (8.84 percent).

Table 21: Age of establishment (left censored in 1975)

Variable	Obs	Mean	Std. Dev.	Min	Median	Max
MIID 2002: Age of establishment (yrs.)	41,734	19.05868	10.1777	0	27	27
SIAB: Age of establishment (yrs.)	613,538	16.5504	9.966858	0	16	27

Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013). Authors' own calculations.

The descriptive statistics on establishment age are presented in Table 21. They illustrate that older firms dominate among the employers of inventors in the MIID 2002. A mean of 19 and a median age of 27 years, i.e., the maximum age in the left censored distribution, indicate that inventors in the sample are highly concentrated on older firms.

While small establishments' growth dynamics may draw to some extent from inventive activities, older establishments are the ones with significantly more employees and resources. Information on the establishment size (number of employees per establishment) is available for all employees in the SIAB data. For descriptive purposes, the size distribution measured on the basis of seven establishment size categories is depicted in Table 22.

Table 22: Inventors' establishments' size classes

Establishment size classes	MIID 2002		SIAB
	Freq.	Percent	Percent
1-19	1,970	4.71	29.37
20-49	1,619	3.87	14.46
50-99	2,155	5.15	11.53
100-499	9,153	21.86	24.49
500-999	5,049	12.06	7.67
1000-1999	6,335	15.13	5.22
2000+	15,453	36.91	7.24
NA	128	0.31	0.01
Total	**41,862**	**100**	**100**

Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013). Authors' own calculations.

The analysis of firm size confirms the basic relationship hypothesized already from the age distribution of inventors' establishments. Large establishments of 500 and more employees are highly overrepresented in the MIID 2002, while small firms are, compared to the SIAB, less likely to employ inventors.

Industries

Merging the BHP data with individual employment biographies enables to report industry affiliation of inventors' employers in detail. In 2002 the NACE Rev.1/ WZ 93 industry classification can be analyzed up to the granularity of the five digit level. For brevity, Table 23 displays the distribution of inventors over the NACE Rev. 1 industry sections.

77.27 percent of all inventors are found to be employed in an establishment operating in the manufacturing sector. Although inventors are highly overrepresented in manufacturing, the focus of the German economy on industrial production and its innovative reputation is well represented in these figures. Runner up industry in terms of patenting and equally represented to its share among all workers is "[K] Real estate, renting and business activities". Other larger industries in the service sectors such as "[N] Health and social work" are represented disproportionally among inventors in the MIID 2002.

Table 23: Inventors' industries (NACE Rev. 1 / WZ 93, Sections)

NACE Rev. 1 / WZ 93, Sections	MIID 2002		SIAB
	Freq.	Percent	Percent
[A] Agriculture, hunting and forestry	56	0.13	1.27
[B] Fishing	.	.	0.01
[C] Mining and quarrying	44	0.11	0.43
[D] Manufacturing	32,345	77.27	24.38
[E] Electricity, gas and water supply	152	0.36	0.82
[F] Construction	339	0.81	6.44
[G] Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household	1,526	3.65	16.15
[H] Hotels and restaurants	14	0.03	3.51
[I] Transport, storage and communication	136	0.32	5.45
[J] Financial intermediation	27	0.06	3.55
[K] Real estate, renting and business activities	5,246	12.53	12.48
[L] Public administration and defence; compulsory social security	96	0.23	5.71
[M] Education	930	2.22	3.71
[N] Health and social work	419	1.00	10.86
[O] Other community, social and personal service activities	400	0.96	4.90
[P] Private households with employed persons	<3	*	0.19
[Q] Extra-territorial organizations and bodies	.	.	0.08
NA	13*	*	0.06
Total	**41,862**	**100**	**100**

Note: "*" and "<3" indicate anonymized data cells.
Source: MIID 2002, SIAB 1975-2010 (Vom Berge et al. 2013). Authors' own calculations.

The subsequent tables display the most important industries of inventors at the more detailed level of 2 digit industries. A distinction is made between inventors who are employed in sub industries of the manufacturing and the service sector.

Table 24 starts with inventors employed in the manufacturing sector and it presents the most important (sub-) industries in this sector with at least 1000 inventors in 2002. The eight industries, for which this condition is met, jointly represent more than 90 percent of the inventor population in manufacturing and 70 percent of all inventors in the MIID 2002. The top ranked industries (machinery, motor vehicles, electrical machinery, chemicals) correspond well with the core of the German industry profile and also fit well with the technological focus of the patents outlined above.

Table 24: Inventors' industries (2-dig. level) in manufacturing ([D])

NACE Rev. 1 / WZ 93, Groups (2-dig. Level)	MIID 2002	
	Freq.	Percent
29 Manufac. of machinery and equipment n.e.c.	6,161	19.05
34 Manufac. of motor vehicles, trailers and semi-trailers	5,595	17.3
31 Manufac. of electrical machinery and apparatus n.e.c.	5,037	15.57
24 Manufac. of chemicals and chemical products	5,019	15.52
33 Manufac. of medical, precision and optical instruments	2,749	8.5
32 Manufac. of radio, television and communication equipment	2,246	6.94
28 Manufac. of fabricated metal products	1,510	4.67
25 Manufac. of rubber and plastic products	1,074	3.32
Others (excl. NA)	2,954	9.13
Total	**32,345**	**100**

Source: MIID 2002. Authors' own calculations.

In services, delineated by in the sectors [G] - [Q], the number of inventors exceeds 500 in only four industries (78.09 percent), with "Research and Development" being the most important sub industry (32.47 percent). Ranked fourth in Table 25, the sub sector "Education" represents the most important sector directly related to public patenting activities and comprises inventors employed by universities. Since professors are usually employed as civil servants, their inventive activities at universities and technical colleges are systematically omitted and reduce the number of inventors covered in this particular field.

Table 25: Inventors' industries (2-dig. level) in services ([G]-[Q])

NACE Rev. 1 / WZ 93, Groups (2-dig. Level)	MIID 2002	
	Freq.	Percent
73 Research and development	2,856	32.47
74 Other business activities	1,971	22.41
51 Wholesale trade and commission trade	1,112	12.64
80 Education	930	10.57
Others (excl. NA)	1,927	21.91
Total	**8,796**	**100**

Source: MIID 2002. Authors' own calculations.

5. Summary

This report provides an extensive overview on the data sources and processing steps employed for the generation of the MPI-IC-IAB Inventor Data 2002 (MIID 2002) and it shows a set of descriptive results on the scope and potential of this novel linked employer-inventor data set.

The MIID 2002 combines patent register data of the German Patent and Trademark Office (DPMA) on inventors residing in Germany with labor market biography data of the Institute for Employment Research (IAB), originating from social security records. Both data bases were matched at the individual level of inventors and employees, using methods of record linkage. During the process of linkage, standardized individual names and addresses served as matching keys. Using the preprocessed inventor data as the basis, we realized a matching rate of about 80 percent between inventors and employees. After excluding ambiguous inventor-employee matches, the MIID 2002 includes high quality micro data on 46,180 inventor-employee pairs, related to 42,435 patent filings in 2002. The realized matching rate at the inventor level of 77 percent compares well with similar data linkage projects performed in Sweden (Jung and Ejermo 2014) and Finland (Väänänen 2010; Toivanen and Väänänen 2012).

The added value of the data linkage is the availability of novel variables for empirical research on inventors and their careers. Comprehensive labor market biography data of the IAB may be enhanced with new individual productivity measures derived from patenting activities of an inventor as recorded in patent register data. Moreover information about teams of employees working in inventive teams is documented on patents. Vice versa, patent register data, which is generally limited in its scope of individual level variables, benefits from the availability of a rich set of employee and employer characteristics recorded in labor market biography data of the IAB. To our best knowledge the MIID 2002 represents the largest micro data base currently used for research on patenting activities of inventors in the context of the labor market and the national system of innovation of Germany.

The authors of this report are currently engaged in research on inventor mobility using the MIID 2002 data (Dorner et al. 2014). Further research exploiting the longitudinal linked inventor-employer structure of the data is planned.

An extension of the inventor population to further years, including a record linkage with individuals recorded in IAB employment biography data is currently in progress. An extensive assessment whether the resulting data may be provided as anonymized research data according to the data privacy law is intended in the future.

References

Blossfeld, H.-P. (1987): Labor Market Entry and the Sexual Segregation of Careers in the Federal Republic of Germany. The American Journal of Sociology, 93(1): 89-118.

Dorner, M.; Heining, J.; Jacobebbinghaus, P.; Seth S. (2010): The Sample of Integrated Labor Market Biographies. Schmollers Jahrbuch – Journal of Applied Social Science Studies, 130(4): 599-608.

Dorner, M.; Bender, S.; Harhoff, D.; Hoisl, K. (2014): The Patterns and Determinants of Inventor Mobility – Evidence from the Employment Biographies of Inventors in Germany. mimeo.

Deutsches Patent- und Markenamt (DPMA) (Ed.) (2004): Jahresbericht 2003 - Deutsches Patent- und Markenamt. München.

European Patent Office (Ed.) (2012): Data Catalog for the EPO Worldwide Patent Statistical Database, München.

Fellegi, I.P.; Sunter, A.B. (1969): A Theory for Record Linkage. Journal of the American Statistical Association, 64(328): 1183-1210.

Fitzenberger, B.; Osikominu, A.; Völter, R. (2006): Imputation rules to improve the education variable in the IAB employment subsample. Schmollers Jahrbuch – Journal of Applied Social Science Studies, 126(3): 405-436.

Gambardella, A.; Giuri, P.; Mariani, M. (2005). The Value of European Patents: Evidence from a survey of European Inventors – Final report of the PatVal EU Project. http://www.alfonsogambardella.it/PATVALFinalReport.pdf, accessed: December 12, 2013.

Giuri, P. et al. (2007): Inventors and invention processes in Europe: Results from the PatVal-EU survey. Research Policy, 36(8): 1107–1127.

Griliches, Z. (1990): Patent Statistics as Economic Indicators: A Survey. Journal of Economic Literature, 28(4): 1661-1707.

Gruhl, A.; Schmucker, A.; Seth, S. (2012): The Establishment History Panel (BHP). Handbook Version 2.1.1. FDZ-Datenreport, 04/2012, Nuremberg.

Hall, B.; Jaffe, A.; Trajtenberg, M. (2001): The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper 8498.

Harhoff, D.; Narin, F.; Scherer, F.M.; Vopel, K. (1999): Citation Frequency and the Value of Patented Inventions. The Review of Economics and Statistics, 81(3): 511-515.

Harhoff, D. (2009): Patent citation Project. http://www.en.inno-tec.bwl.uni-muenchen.de/research/proj/patent-cit-project/index.html, accessed: August 29, 2012.

Herberger L.; Becker B. (1983): Sozialversicherungspflichtig Beschäftigte in der Beschäftigtenstatistik und im Mikrozensus. Wirtschaft und Statistik, vol. 4, 290–304.

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

Hethey-Maier, T.; Schmieder, J. F. (2013): Does the use of worker flows improve the analysis of establishment turnover? - Evidence from German administrative data. Schmollers Jahrbuch – Journal of Applied Social Science Studies, 133 (4): 477-510.

Hoisl, K. (2007): Tracing mobile inventors - The causality between inventor mobility and inventor productivity. Research Policy, 36 (5): 619-636.

Hunt, J., Garant, J.-P.; Herman, H.; Munroe, D. J. (2013): Why are women underrepresented amongst patentees?, Research Policy, 42(4): 831-843.

Jaffe A.; Trajtenberg, M. (Eds.) (2002): Patents, Citations & Innovations. A Window On The Knowledge Economy. The MIT Press, London/Cambridge.

Jaro, M.A. (1989): Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida. Journal of the American Statistical Association, 84(406): 414-420.

Jung, T.; Ejermo, O. (2014): Demographic patterns and trends in patenting: Gender, age, and education of inventors. Technological Forecasting and Social Change. Online first, http://dx.doi.org/10.1016/j.techfore.2013.08.023.

Organisation for Economic Co-Operation and Development (OECD) (Ed.) (1994): Using Patent Data as Science and Technological Indicators. Patent Manual 1994. Paris.

Schmoch, U. (2008): Concept of a Technology Classification for Country Comparisons, Final Report to the World Intellectual Property Organisation (WIPO). Fraunhofer Institute for Systems and Innovation Research, Karlsruhe.

Schnell, R.; Bachteler, T.; Bender, S. (2004): A Toolbox for Record Linkage. Austrian Journal of Statistic, 33(1/2): 125-133.

Schnell, R.; Bender S.; Bachteler, T. (2003): Record Linkage Using Error-prone Strings. Proceedings of the Section on Survey Research Methods, p. 3713–3717.

Scioch, P. (2010): The Impact of Cleansing Procedures and Coding Decisions for Overlaps on Estimation Results - Evidence from German Administrative Data. Schmollers Jahrbuch – Journal of Applied Social Science Studies, 130(4): 485-512.

Spengler, A. (2008): The Employment History Panel. Schmollers Jahrbuch – Journal of Applied Social Science Studies, 128(3): 501-509.

Trajtenberg, M.; Shalem, R. (2009): Software Patents, Inventors and Mobility. http://ssrn.com/abstract=1547276, accessed: April 15, 2013.

Toivanen, O.; Väänänen, L. (2012): Returns to Inventors. The Review of Economics and Statistics, 94(4): 1173-1190.

Väänänen, L. (2010): Human Capital and Incentives in the Creation of Inventions – A Study of Finnish Inventors. PhD Thesis Aalto University of Helsinki. http://hsepubl.lib.hse.fi/pdf/diss/a368.pdf http://epub.lib.aalto.fi/pdf/diss/a368.pdf, accessed: June 2, 2014.

Vom Berge, P.; König, M.; Seth, S. (2013): Sample of Integrated Labor Market Biographies (SIAB) 1975-2010. FDZ-Datenreport, 01/2013, Nuremberg, 65 p.

Winkler, W. E. (1990): String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. Proceedings of the Section on Survey Research Methods (American Statistical Association): 354–359.

Winkler, W. E. (1995): Matching and Record Linkage, In: Cox, B.G. et al. (Eds.): Business Survey Methods. John Wiley & Sons Inc. Hoboken, NJ.

Appendix

A1: Patent application form of the DPMA (3 pages)

Deutsches
Patent- und Markenamt

An das
Deutsche Patent- und Markenamt
80297 München

(1) Vordruck nicht für PCT-Verfahren verwenden, siehe Seite 4	**Sendungen des Deutschen Patent- und Markenamts sind zu richten an:** **Name, Vorname / Firma** **Straße, Hausnummer / ggf. Postfach** **Postleitzahl** **Ort**	**Antrag auf Erteilung eines Patents** **1** **Datum** TT MM JJJJ ☐ **TELEFAX** vorab am TT MM JJJJ
(2)	Zeichen des Anmelders/Vertreters (max. 20 Stellen)	Telefon des Anmelders/Vertreters
(3)	Der Empfänger in Feld (1) ist der ☐ Anmelder ☐ Zustellungsbevollmächtigte ☐ Vertreter	ggf. Nr. der Allgemeinen Vollmacht
(4) nur auszufüllen, wenn abweichend von Feld (1)	**Anmelder** **Name, Vorname / Firma lt. Handelsregister** **Straße, Hausnummer (kein Postfach!)** **Postleitzahl** **Ort** **Land** (falls nicht Deutschland)	
Handelsregisternummer nur bei Firmen anzugeben	☐ Der Anmelder ist eingetragen im Handelsregister Nr. beim Amtsgericht	
	Vertreter **Name, Vorname / Bezeichnung** **Straße, Hausnummer** **Postleitzahl** **Ort**	

P2007 10.11 2

(5) soweit bekannt	Anmelder-Nr. / Vertreter-Nr. / Zustelladressen-Nr.
(6) siehe Seite 4 IPC-Vorschlag ist unbedingt anzugeben, sofern bekannt	**Bezeichnung der Erfindung** — **IPC-Vorschlag des Anmelders** /
(7) siehe Erläuterung und Kostenhinweise auf Seite 4	**Sonstige Anträge** ☐ Die Anmeldung ist **Zusatz** zur Patentanmeldung (zum Patent) ➔ Aktenzeichen der Hauptanmeldung (des Hauptpatents) ☐ **Prüfungsantrag** - Prüfung der Anmeldung **mit** Ermittlung der öffentlichen Druckschriften (§ 44 Patentgesetz) ☐ **Rechercheantrag** - Ermittlung der öffentlichen Druckschriften **ohne** Prüfung (§ 43 Patentgesetz) ☐ **Aussetzung** des Erteilungsbeschlusses auf ______ Monate (§ 49 Absatz 2 Patentgesetz) *(Max. 15 Monate ab Anmelde- oder Prioritätstag)*
(8)	**Erklärungen** ☐ Teilung ☐ **Ausscheidung** aus der Patentanmeldung ➔ Aktenzeichen der Stammanmeldung ☐ an **Lizenzvergabe** interessiert (unverbindlich) ☐ **Nachanmeldung im Ausland** beabsichtigt (unverbindlich)
(9) siehe auch Seite 4	☐ Inländische **Priorität** (Datum, Aktenzeichen der Voranmeldung) ☐ Ausländische **Priorität** (Datum, Land, Aktenzeichen der Voranmeldung; vollständige **Abschrift(en) der ausländischen Voranmeldung(en)** beifügen)

P2007 10.11 3

(10) Erläuterung und Kostenhinweise siehe Seite 4	**Gebührenzahlung** in Höhe von ____________ EUR ☐ **Einzugsermächtigung** Vordruck (A 9507) *ist beigefügt* ☐ **Überweisung** *(nach Erhalt der Empfangsbescheinigung)* **Wird die Anmeldegebühr nicht innerhalb von 3 Monaten nach dem Tag des Eingangs der Anmeldung gezahlt, so gilt die Anmeldung als zurückgenommen! Bitte beachten Sie die Hinweise zur Änderung der Anmeldegebühr zum 1. Oktober 2009 auf Seite 4!**
(11) siehe auch Seite 4	**Anlagen** 1. ________ Vertretervollmacht 2. ________ Erfinderbenennung (P 2792) 3. ________ Zusammenfassung (ggf. mit Zeichnung Fig. ________) 4. ________ Seite(n) Beschreibung (ggf. mit Bezugszeichenliste) 5. ________ Seite(n) Patentansprüche ________ Anzahl Patentansprüche 6. ________ Blatt Zeichnungen 7. ________ Abschrift(en) der Voranmeldung(en) 8. ________ Zitierte Nichtpatentliteratur 9. ________ Anzahl Datenträger ☐ für Sequenzprotokoll nach § 11 Absatz 2 Patentverordnung ☐ für umfangreiche Anmeldungsunterlagen nach § 6 Absatz 1 Satz 2 Patentverordnung 10. ________ Angabe des geographischen Herkunftsortes gemäß § 34a Patentgesetz 11. ________
	(12) **Unterschrift(en)** (13) **Funktion des Unterzeichners** **Bitte beachten Sie die Hinweise auf der nächsten Seite**

Source: DPMA (http://www.dpma.de/docs/service/formulare/patent/p2007.pdf)

A2: Naming of inventor form (2 pages)

Deutsches
Patent- und Markenamt

Erfinderbenennung

P 2792 10.11 1

Die Erfinderbenennung muss auch erfolgen, wenn der Anmelder selbst der Erfinder ist. Ist der Anmelder Miterfinder, so ist er auch mitzubenennen.

Amtliches Aktenzeichen *(wenn bereits bekannt)*

Platz für **Zeichen des Anmelders/Vertreters**

Bezeichnung der Erfindung *(bitte vollständig)*

Erfinder (1)

Vor- und Zuname

Straße, Hausnummer

Postleitzahl Ort

Erfinder (2)

Vor- und Zuname

Straße, Hausnummer

Postleitzahl Ort

Erfinder (3)

Vor- und Zuname

Straße, Hausnummer

Postleitzahl Ort

P2792 10.11 2

Erfinder (4)

Vor- und Zuname

Straße, Hausnummer

Postleitzahl Ort

Achtung: bei mehr als vier Erfindern bitte gesondertes Blatt benutzen!

Das Recht auf das Patent ist **auf den Anmelder übergegangen durch**:
(z.B. Erfinder ist/sind der/die Anmelder, Inanspruchnahme aufgrund §§ 6 u. 7 ArbnErfG, Kaufvertrag mit Angabe des Datums, Erbschaft usw.)

Es wird versichert, dass nach Wissen des/der Unterzeichner/s weitere Personen an der Erfindung nicht beteiligt sind.

______________________, den ________

Eigenhändige Unterschrift des Anmelders oder der Anmelder bzw. des Vertreters.
Bei Firmen genaue, eingetragene Firmenbezeichnung angeben.

Antrag auf Nichtnennung als Erfinder

Nur von denjenigen oben genannten Erfindern auszufüllen, die nach außen hin nicht bekanntgegeben werden wollen (§ 63 Abs. 1 S. 3 PatG).
Der Antrag kann jederzeit widerrufen werden. Ein Verzicht des Erfinders auf Nennung ist ohne rechtliche Wirksamkeit (§ 63 Abs. 1 S. 4 u. 5 PatG).

☐ Es wird beantragt, den bzw. die Unterzeichner dieses Antrags in der oben angegebenen Patentanmeldung als Erfinder nicht öffentlich bekanntzugeben. Die Einsicht in die obige Erfinderbenennung wird nur bei Glaubhaftmachung eines berechtigten Interesses gewährt.

______________________, den ________

Eigenhändige Unterschrift des Erfinders oder der Erfinder

Source: DPMA (http://www.dpma.de/docs/service/formulare/patent/p2792.pdf)

Imprint

FDZ-Methodenreport 06/2014 Englisch

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, Heiner Frank

Technical production
Heiner Frank

Download
http://doku.iab.de/fdz/reporte/2014/MR_06-14_EN.pdf

Internet
http://fdz.iab.de/

Corresponding author:

Matthias Dorner
Institute for Employment Research (IAB)
of the Federal Employment Agency (BA)
Regensburger Str. 104
90478 Nuremberg
phone: +49(0)911 / 179-8895
mailto: Matthias.Dorner@IAB.de