CESPRI

Centro di Ricerca sui Processi di Innovazione e Internazionalizzazione
Università Commerciale "Luigi Bocconi"
Via R. Sarfatti, 25 – 20136 Milano
Tel. 02 58363395/7 – fax 02 58363399
http://www.cespri.unibocconi.it

Francesco Lissoni, Bulat Sanditov, Gianluca Tarasconi

The Keins Database on Academic Inventors: Methodology and Contents

WP n. 181 **September 2006**

Stampato in proprio da:
Università Commerciale Luigi Bocconi – CESPRI Via Sarfatti, 25 20136 Milano
Anno: 2006

The Keins Database on Academic Inventors: Methodology and Contents

Francesco Lissoni[♣♠], Bulat Sanditov[♣♦], Gianluca Tarasconi[♣]

[♣]CESPRI-Università Bocconi, Milan; [♦]MERIT-Universiteit Maastricht;
[♠]Università degli studi di Brescia

corr.author: francesco.lissoni@unibocconi.it

Abstract:

The paper describes the methodogy used to build a database on academic inventors from France, Italy, and Sweden (1978-2004), which was delivered to the European Commission as part of the KEINS project (Knowledge-Based Entrepreneurship: Innovation, Networks and Systems), and will provide the basis for future publications. It provides an overview of the database contents, as well as information on access rules and on related datasets by CESPRI-Università Bocconi. The database is the result of joint efforts by CESPRI-Università Bocconi (Milan, IT), BETA – Universitè "Louis Pasteur" (Strasbourg, FR), IMIT-Chalmers University (Gotheborg, SE), Umea Universitet (SE), and Università degli studi di Brescia (IT).

Keywords: University patents; European universities; Patent database.

JEL Codes: C81; I23; O34

1. Introduction

The KEINS database on academic inventors contains detailed information on university professors from France, Italy, and Sweden, who appear as designated inventors on one or more patent application registered at the European Patent Office (EPO), 1978-2004. Produced for the EU-sponsored project on *Knowledge-based Entrepreneurship: Innovation, Networks and Systems*, it will be made available to all interested researchers, starting June 2007, through the CESPRI website[1]. Besides CESPRI, KEINS partners contributing to the database have been: BETA (Universitè "Louis Pasteur", Strasbourg) and IMIT-Chalmers University (Gotheborg). Umea Universitet and Università degli studi di Brescia have also contributed with data and by undertaking data-cleaning tasks.

The KEINS database originates from the EP-INV database produced by CESPRI-Università Bocconi, which contains all EPO applications, reclassified by applicant and inventor; and from three lists of university professors of all ranks (from assistant to full professors), one for each of the above mentioned countries (PROFLISTs). Academic inventors have been identified by matching names+surnames of inventors in the EP-INV database with those in the PROFLISTs, and by checking by e-mail and phone the identity of the matches, in order to exclude homonyms.

Thanks to this methodology, the KEINS database differs from other data collections on university patents, in that it includes not only any patent owned by universities, but also all patents that originate from university scientists and are owned by business companies, public research organizations, and the scientists themselves. Therefore, it allows to measure more accurately the contribution of university to technology transfer *via* patented inventions.

Users of the KEINS database need to achieve some understanding of the complex methodology followed to identify inventors in the EP-INV database, and of the subsequent matching procedure; and if they wish to extend the geographical coverage of the database beyond the three original countries, they may find it useful to employ the SQL, SAS, and Access software tools that we developed over time. This is what the present paper is about, alongside with summary information on the database contents. For a richer analysis of academic inventors and patents in France, Italy, and Sweden based upon the KEINS database, see Lissoni et al. (2006).

* *Acknowledgements*: Setting up the KEINS database has required the contribution of many people. Besides the authors of the present paper one should mention: Patrick Llerena, Julien Penin, and Muge Ozman (BETA-Université "L.Pasteur") for the French section; Maureen McKelvey (Chalmers Univ.), Ingrid Schild and Cecilia Yttergren (Umea Univ.) for the Swedish section; Christian "Troubleshooter" Catalini (CESPRI-Bocconi), who has fine-tuned the Access©-based checking tools; GianPaolo Ziletti, Samuela Bellini, Antonio DellaMalva, Riccardo Cropelli, Roberto Giusto, and Massimiliano Perini, who have kindly shared data collected for their dissertation works either at the university of Brescia or Bocconi university. The original data on patents and inventors at the basis of the KEINS database come from the EP-CESPRI database, which is also the result of a collective effort: Stefano Breschi, Fabio Montobbio, and Lorenzo Cassi are among those who have most contributed to it.

[1] http://www.cespri.unibocconi.it

In section 2 we present the contents of the EP-INV database, and the methodology behind it. In particular, we discuss the logic and results of the *Massacrator*© routine developed by one of us [2].

In section 3 we present the contents of the three national PROFLISTs, with special emphasis on the Swedish one, which did not derive from any well-structured administrative database (such as the French and the Italian one), but had to be assembled from various sources for the specific purposes of the KEINS project.

In sections 4 and 5 we present the procedure followed for matching EP-INV and the PROFLISTs, and for checking the results, respectively.

In section 6 we conclude by provide summary statistics on the contents of the resulting KEINS database.

A number of Appendixes contain useful information on classification systems and software routines we employed, and on access and dissemination rules concerning the KEINS database.

2. The EP-INV database

The EP-INV dataset is part of the broader EP-CESPRI database, which provides information on patents applied for at the European Patent Office (EPO), from 1978 to January 2005. The EP-CESPRI database is based upon applications published on a regular basis by the *Espacenet Bulletin* and is updated yearly; presently, it contains about 1,500,000 patent applications. Data fall into five broad categories:

1. *Patent data*: publication number (from now on: PUNR), title, abstract, priority dates, application year, technological class (IPC 12-digit)[3], "granted" dummy, and equivalences.
2. *Applicant data*: unique Cespri code (from now on: CODFIRM), name(s)[4], address, city, province, region, and country[5]; applicant companies active in selected industries are also classified according to their economic activity (SIC code), and the unique Dun&Bradstreet code[6]. Overall, the database contains information on about 144,000 companies (applicant individuals excluded).

[2] The original core of the EP-INV database was built with *ad hoc* criteria for Italian inventors back in 2001. Data from that pioneer database were first published in Balconi et al. (2004). The *Massacrator*© routine builds upon that experience, and it is general enough to be applied to inventors from all countries. Gianluca Tarasconi is the author of *Massacrator*©.

[3] IPC stands for International Patent Classification, an international classification system produced by WIPO (World Intellectual Property Organization; http://www.wipo.org) and adopted by EPO for examination purposes. Economists and sociologists of technical change have developed a number of IPC-based synthetic classifications for research purposes, a very popular one being the OST-INPI/FhG-ISI technology nomenclature, a joint product of the Observatoire de Sciences et Technologies (FR), the INPI (Institute Nationale Proprieté Industrielle, FR), and the Fraunhofer Institute for Systems of Innovation Research (DE). EP-CESPRI data also come reclassified in this way (for reference OST, 2004, p.513)

[4] Patents can be applied for jointly by one or more firms or individuals.

[5] Province, region and country stand for a bottom up classification of administrative levels of the examined nation; for example, UK has counties and regions and state (England, Scotland, N. Ireland), France has counties and regions only, India has county and state.

[6] The selected industries are Computers (SIC 4 digit: 3571, 3572, 3575); Instruments (SIC 4 digit: 3826, 3827); Pharmaceuticals (SIC 4 digit: 2834); Plastics (SIC 4 digit: 2821); Semiconductors (SIC 4 digit: 3674); Telecoms (SIC 4 digit:3661, 3663). See: Breschi et al. (2004)

3. *Inventor data*: unique Cespri code (from now on: CODINV), name, surname, address, city, province, region, country, co-inventors' CODINV codes[7]. Inventors listed originally in the database were about 1,790,000 inventors, but the number is foreseen to decline over time, due to successive rounds of data cleaning (see below in this section).
4. *Applicant's parent company data*: group name, domestic parents, proprietor changes[8]
5. *Citations*: citations to patent (EPO; USPTO) and non patent literature, per citing patent.

Information contained in the EP-INV database coincide with that listed at point 3. above, but can be easily connected with information listed at points 1., 2., and 5.; moreover, information listed at points 1., 2., 4., and 5. have been used to create the EP-INV database.

Appendix 1 contains a summary description of the structure of the whole EP-CESPRI database, and a detailed description of the structure and contents of the EP-INV database, including the complete list of variable names and their explanation.

The creation of the EP-INV database followed three steps:

FIRST, the standardization of names and addresses (in order to assign unique CODINV codes to all inventors with the same names, surnames, and address);

SECOND, the calculation of "similarity scores" for pairs of inventors with the same name and surname, but different addresses;

THIRD, the identification (by country) of a threshold value over which two inventors in a pair are considered the same individual, and assigned the same unique code CODINV.

In what follows we describe each step in detail.

FIRST: Standardization of inventors' names and addresses

Original EPO data on inventors, from the Espacenet Bulletin, come in a text string, which is processed in three steps.

1. *Parsing*. The original text string is parsed into several fields: joint "name+surname", address, city, province, region, state (for US inventors), country, zip code, and a residual field.
2. *Cleaning of address data*. Parsed data are cleaned by: shifting information contained in wrong fields (like zip code, county…); fixing mistakes in zip codes, according to national post office tables; standardizing city names or parts of names (e.g.: "Saint" is turned into "St.")

[7] Patents are very often signed by more than one inventor. Each inventor's record contains info (CODINV codes) on all the co-inventors, i.e. all the individuals comprised in the database who have been listed at least on one patent alongside with the inventor.

[8] Only for companies that belong to industries listed in footnote 4 above

3. *Cleaning of names*. The “name+surname” field“ is parsed once more into the following fields: first name, second name, extension (e.g. Jr, Sr, III), surname, and academic title (e.g. Dr., Prof, Ing….). This operation was based on four steps:
   - inventors with the same address and equal first name, surname, extension and initial of third name are corrected for third name (e.g.: “Rossi Giovanni Paolo” is turned into “Rossi Giovanni P.”);
   - inventors with same address, same city, and full name different for less than 3 characters or less than 10% of the total characters are given the same name (based on the name of inventor with higher number of patents);
   - inventors with the same full name, same address and city different for less than 3 characters or 10% of total chars in the name of the city are given the same city (based on the city of inventor with higher number of patents);
   - inventors with the same full name, same city and address different for less than 3 chars or 10% of total chars in the name of the address are given the same address (based on the address of inventor with higher number of patents).

All of the above steps are performed recursively, and each inventor is assigned each inventor a CODINV unique code.

SECOND: Computation of similarity scores

All inventors *with the same name and surname*, but different CODINV (that is, different address, city, province, region or nation, are compared in *pairs*, through the Massacrator© SQL routine[9]. Massacrator© compares biographical information on each inventor in the pair, as well as on the technological contents (IPC code) and applicant of each inventor’s patents. It also creates and exploits information on relationships between the inventors in the pair, such as the existence of citations running from one inventor’s patents to the other’s, the existence of a common co-inventor, or the existence of a social tie, through chains of co-inventorship, up to three degrees of separation (figure 1).

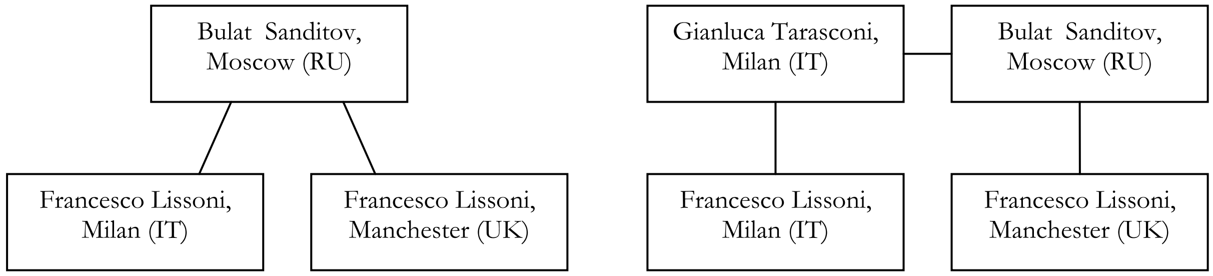
Similarities in the biographical information, or in the technological contents of the patents of two inventors in a pair suggest that two inventors may indeed be the same person.

The existence of any kind of relationship between the two inventors is even a stronger indicator in this direction. If two inventors with the same name and surname, but different CODINV code, have been working with the same people, they are very likely to be indeed the same person, who changed home at some point in time between the priority dates of his/her patents; the same line of reasoning applies to inventors who are connected through a short chain of mutual acquaintances (3 degrees of separation). As

[9] Massacrator© was created by Gianluca Tarasconi

for citation links between the inventors' patents, evidence produced by Breschi and Lissoni (2004) and Singh (2005) suggest that the probability of observing a citation link between two patents increases drastically when the two patents share at least one designated inventor.

Figure 1 – Relational data between inventors with the same name and surname, but different address


Bulat Sanditov, Moscow (RU)
Francesco Lissoni, Milan (IT)
Francesco Lissoni, Manchester (UK)
Gianluca Tarasconi, Milan (IT)
Bulat Sanditov, Moscow (RU)
Francesco Lissoni, Milan (IT)
Francesco Lissoni, Manchester (UK)

a) COMMON CO-INVENTOR **b) 3 degrees OF SEPARATION**

NB.: Ties between two inventors indicate that the latter were designated at least once on the same patent (co-inventionship)

Accordingly, a cumulative "similarity score" was assigned to each pair of inventors with the same name and surname, but different CODINV, based upon scores for individual similarity criteria listed in table 1, to be summed up together. Notice that a negative score was assigned to pairs whose inventors patented at more than 20 years of distance, or when two surnames are very common (in the country to which both inventors come from).

Table 1 – Similarity scores

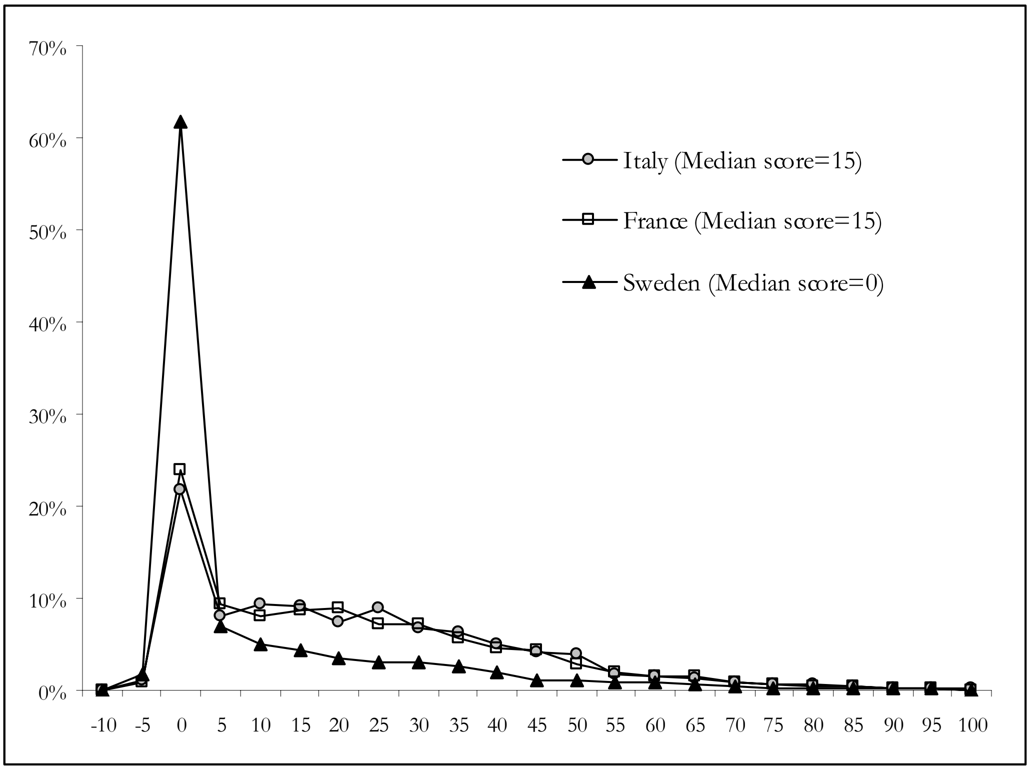
Biographical information	
Same city	+5
Same province	+5
Same region	+5
Same state (US)	+5
Same address [in different cities; it may indicate misspellings in the city field]	+5
Widespread surname	-5
Contents, dates and property of patents	
Same IPC code (4 digits)	+5
Same IPC code (6 digits)	+5
Same IPC code (12 digits)	+10
Priority dates differ for >20 years	-5
Same applicant	+5
Same applicant (the applicant has <50 inventors)	+5
Same group (if Dun&Bradstreet code is available)	+5
Relational data	
Same coinventor	+10
3 degrees of separation	+10
Inventor 1 cites inventor 2	+5
Inventor 1 is cited by inventor 2	+5

THIRD: Use of similarity scores

Intuitively, high similarity scores can be taken as indication of a high probability that the two inventors in the pairs are the same person. Whenever two inventors in a pair are found to be the same, the highest CODINV code is eliminated, and the lowest CODINV code one is assigned to both inventors.

Manual checking of EP-INV records suggest that a large number paired inventors with total score higher than 20 are indeed the same person. But percentages vary across countries, largely because of the different distribution of frequent surnames. Therefore, no automatic re-assignment of CODINV codes has been performed so far. Threshold values for the score have been set country by country, according to the score distribution within the country. Figure 2 reports the distributions for France, Italy, and Sweden, the three countries covered by the KEINS database.

Figure 2 – Distribution of total similarity scores, by country



** Max score for all countries exceeds 100. Cumulated frequency at 100 score is ≥99% for all countries*

Distributions of similarity scores for France and Italy have very similar profiles: the mode value is zero, a score exhibited by about 25% of inventors' pairs; the median value is around 15. This suggested us to set the threshold value of the total similarity score at 15: inventors in pairs with score equal or higher than 15 are then presumed to be the same person, and assigned the same CODINV code. Manual checking suggests that no Type 2 error is introduced with this choice (no pair of inventors are assigned erroneously the same CODINV code), although some Type 1 error remains (some pairs of inventors who are indeed the same person have scores <15 and are not given the same CODINV code).

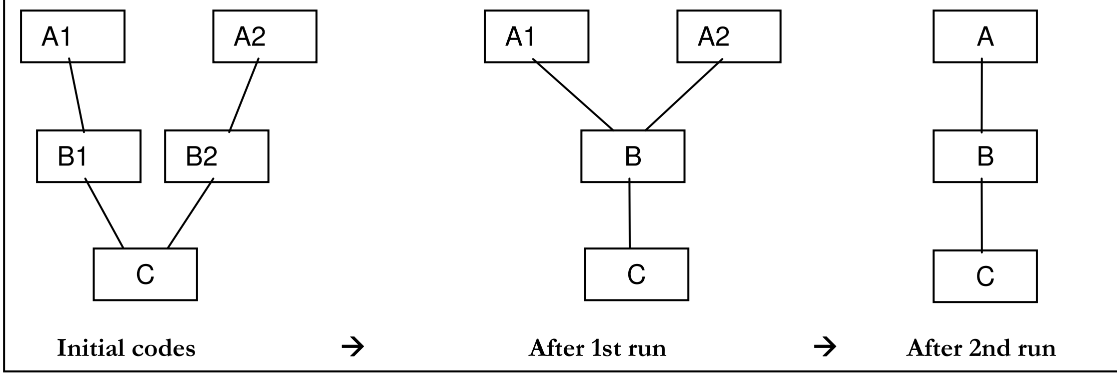
As for Sweden, the frequency of a few common surnames (such as Larsson and Andersson) is so high that most pairs of inventors are given a "-5" for sharing the surname. As a result the average total score is lower than in France and Italy, the median value being zero. Considering that choosing such value as the threshold one for assigning common CODINV codes would have been too risky (it would have been possibly introduced Type 2 errors), we decided to adopt for Sweden the same threshold value of the other two countries. This complicated somehow the following inventor-professor matching exercise (see section 4 below).

This exercise of re-assignment of CODINV codes reduced considerably our estimate of the number of inventors active in each country between 1978 and 2004. French inventors pre- and post-CODINV re-assignment were respectively 119'625 and 98'227; the same figures for Italy were 39'934 and 37'784, and for Sweden 28'163 and 25'882.

The post-CODINV re-assignment figures, however, are not definitive. The re-assignment exercise is necessarily a recursive one, due to the presence of relation criteria in the calculation of the similarity index. Figure 3 illustrates this point: after a first run of CODINV re-assignment a few inventors the same name and address (such as those coded A1 and A2 in figure) find themselves at less than 3 degrees of separation, while they were previously at 4 (or more); this occurs because the first run of CODINV re-assignment identified the related inventors B1 and B2 as the same one, now coded B. Being at less than 3 degrees of separation increases the similarity index for A1 and A2, up to a point that they may now be identified as the same person (A). And so on.

Figure 3 – Recursive re-assignment of CODINV codes

Name Surname, City (CY)	**Initial CODINV codes**	**CODINV codes after 1st run**	**CODINV codes after 2nd run**
Francesco Lissoni, Manchester (UK)	A1	A1	A
Francesco Lissoni, Milan (IT)	A2	A2	A
Bulat Sanditov, Maastricht (NL)	B1	B	B
Bulat Sanditov, Moscow (RU)	B2	B	B
Gianluca Tarasconi, Milan (IT)	C	C	C
nr. of inventors	*5*	*4*	*3*



3. National PROFLISTs

Parallel to the creation of the EP-INV database we proceeded to the collection of biographical information on academic scientists in the three countries of interest. The collection effort was directed at the so-called "hard sciences", which exclude all humanities and social sciences. The main reason for this exclusion was the need to avoid including in the PROFLISTs as many homonyms as possible, who could complicate a lot the matching exercise described in section 4. The exclusion comes at the cost of disregarding a few academic inventors, especially in the social sciences, who are engineers by training and may have patents in information technology[10].

Each PROFLIST comes with a highly idiosyncratic disciplinary classification systems (in the case of Sweden we have indeed two classification systems, which overlap only partially, and are not exhaustive of the professors' list). For the purposes of the KEINS project we have produced an 18 classes disciplinary classification, loosely based on the French classification system, to which each national classification can be converted (see table A2.7 in Appendix 2).

Three partners teams in the KEINS were involved at this stage: CESPRI, BETA and CHALMERS

3.1 Italian PROFLIFST

CESPRI produced the Italian PROFLIST, starting from data already published in Balconi et al. (2004). Those data were based on the complete list of all Italian university professors (assistant, associate, full) active in 2000, provided by the Italian Ministry of Education. A new list, updated to 2004, was obtained from the Ministry (thanks to Margherita Balconi's kind help). Professors in the two lists did not come with a common code, so CESPRI matched them in the 2000 and 2004 lists by surname, first name, and the date of birth.

Whatever their rank, Italian professors in public universities are tenured civil servants, and even those working in private universities are tenured and recorded for administrative purpose in the Ministry's list. However, the Ministry does not keep central records of PhD students nor of the numerous contract-based researchers and instructors who populate Italian universities.

Table 2 summarizes the main contents of the Italian PROFLIST. Notice that information of the professors' rank and affiliation both in 2000 and 2004 allows for panel data or pooled cross section analysis of academic careers. All data are stored in SAS tables, whose contents and relations are described in Appendix 2, alongside with the variable names and their explanations.

[10] For instance, Peter Magnusson (http://www.ctf.kau.se/People/PeterMagnusson.shtml) works for economic department at KAU, but he is an author of several patents in IT as he is an engineer by training, and used to work in R&D department of large telecom companies.

Table 2 – Contents of Italian PROFLIST

Surname
Name
Gender
Date of birth(1)
University affiliation in 2000
University affiliation in 2004
Public institution (2)
Rank in 2000 (3)
Rank in 2004 (3)
Honorary status(4)
Date of nomination (5)
Disciplinary field in 2000 (6)
Disciplinary field in 2004 (6)
Faculty in 2000
Faculty in 2004

(1) Available only for professors still active in 2004

(2) Dummy variable (=1 for public unv.)

(3) Assistant/Associate/Full

(4) Dummy variable (=1 for professors temporarily out of job); available only for 2004

(5) It indicates when the professor acquires the rank he had in 2004

(6) Classification of fields changed from 2000 to 2004 (conversion tables available)

Obs: 27844 (2000) and 32886 (2004) [only scientific and technical disciplines]

3.2. French PROFLIFST

BETA (Bureau d'Economie Théorique et Appliquée), a research centre of the Université "Louis Pasteur" in Strasbourg, compiled a French PROFLIST also based upon Ministerial records and similar to the Italian one (French professors are, like the Italian, tenured civil servants whose wages and careers depend upon the national administration).

The French PROFLIST is the result of separate records for the medical and nonmedical disciplines (only scientific and technical ones). It refers to academic professors of various ranks ("maitre a conference" or "professeur"), active in 2005; it also contains . Table 3 summarizes its main contents; Appendix 2, provides more details.

Table 3 – Contents of French PROFLIST

Surname
Name
Gender
Date of birth
University affiliation
Rank
Date of nomination (5)
Disciplinary field

* All info refer to 2005

Obs: 32006 [only scientific and technical disciplines]

3.3 Swedish PROFLIFST

Swedish academic personnel are not civil servants, so no list of university professors could be obtained from the Swedish Ministry of Education. Ingrid Schild (Dept. of Sociology, Umea Univ.) took upon her the task of collecting list of personnel from as many Swedish academic institutions as possible, and to work with CESPRI in order to standardize and integrate them. Table 4 provides an inventory of all Swedish universities, pointing out those that contribute or not to the Swedish PROFLIST. Most of the non-contributing ones do not host scientific or technical faculties.

Table 4 – Swedish universities contributing/not contributing to Swedish PROFLIST

Contributing	**NOT contributing**
Blekinge tekniska högskola (BTH)	Högskolan i Halmstad (HH)
Chalmers tekniska högskola (CHA)	Handelshögskolan i Stockholm (HHS)
Göteborgs universitet (GU)	Högskolan i Kalmar (HIK)
Högskolan Dalarna (DU)	Högskolan Kristianstad (HKR)
Högskolan i Borås (HB)	Idrottshögskolan i Stockholm
Högskolan i Gävle (HIG)	Konstfack
Högskolan i Jönköping (HJ)	Kungl. Konsthög-skolan
Högskolan i Skövde (HIS)	Kungl. Musikhög-skolan i Stockholm
Högskolan i Trollhättan/Uddevalla (HTU)	Örebro universitet (ORU)
Högskolan på Gotland (HGO) (2)	Teaterhögskolan i Stockholm
Karlstads universitet (KAU)	Dramatiska institutet
Karolinska institutet (KI)	Danshögskolan
Kungl. Tekniska högskolan (KTH)	
Lärarhögskolan i Stockholm (LAR) (1)	
Linköpings universitet (LIU)	
Luleå tekniska universitet (LTU)	
Lunds universitet (LU)	
Mälardalens högskola (MDH)	
Malmö högskola (MAH)	
Mittuniversitetet (MIU) (1)	
Operahögskolan i Stockholm (OH) (3)	
SLU (Sveriges lantbruksuniversitet)	
Södertörns högskola (SH)	
Stockholms universitet (SU)	
Umeå universitet (UMU)	
Uppsala universitet (UU)	
Växjö universitet (VXU)	

(1) No information on year of birth

(2) No information on private addresses

(3) No information on faculty/discipline

Obs: 25'196 [only employees with academic positions, social sciences and humanities excluded]

The contents of the resulting Swedish PROFLIST are summarized in table 4, and described to a greater extent in Appendix 2.

Table 5 – Contents of Swedish PROFLIST

Surname
Name
Gender
Date of birth
University affiliation
Faculty
Department
Rank
Date of nomination (5)
Disciplinary field
Private address

* All info refer to 2005

Obs: 25'196 [only employees with academic positions, social sciences and humanities excluded]

Two major differences in contents between the Swedish PROFLIST and the others is that the former includes not only the teaching body, but also all the university technical and support staff, as well as the PhD student; and that in most cases we also have information on the individuals' private addresses (which turn out to be useful for the professor-inventor matching exercise; see below). The major drawbacks of the Swedish PROFLIST is that information on the professors' rank and discipline is not uniform across universities, each of which adopts its own classification. As a result, we had to create our own classifications, based entirely on our own translation of Swedish terminology into English.

4. From the EP-INV to the KEINS database: inventor-professor matching

The identification of academic inventors was pursued in two steps. We first matched inventors from the EP-INV database with professors in the national PROFLISTs, by name and surname, and then sent e-mails and/or made phone calls to the resulting matched professors to ask for confirmation or their inventor status

We describe the first step in this section, and the second step in section 5.

The matching exercise also consisted of three steps.

1. A "narrow" matching exercise based upon inventors' and professors' full names (surname+ first name+middle names)
2. A "broad" matching exercise, directed at the inventors and professors which escaped the first-step matching, by surname+first name (i.e. with middle names excluded).
3. A "filtering-out" exercise, aimed at eliminating incongruous inventor-professors matches, on the basis of age- and discipline-based criteria. The "age filter" required that, by the time of the patent application filing for the *first patent* attributed to a matched professor, the latter was not younger than 21. The "discipline filter" was based on a list of "incompatible" disciplines (as from the national PROFLISTs) and IPC 3-digit codes (i.e. technologies, where IPC stands for International Patent Classification, a 12-digit classification system adopted by EPO). For instance, disciplines such as "Astronomie, astrophysique" (CNU=3400 in the French PROFLIST) were considered incompatible with technologies such as "Baking; Edible Doughs" (IPC-3digit=B02). Since the "incompatibility" list was based on our own common sense, and not on any expert's opinion, we kept it at a minimum, including in it only the most noticeable clashes. In case where any of patents filed by CODINV2 and information on PROFCODE is "caught" by either of the filters corresponding CODINV-PROFCODE match assumed to be spurious and has been deleted from the list of matches.

In what follows we describe how this procedure applied to French professors from non-medical disciplines, which can be considered an exemplary case.

The list of French non-medical professors has 46’552 names of which 25’825 are names of professors in “hard sciences”.

Inventors with a French address in the EP-INV database were originally to 119’625 names (120’313 CODINV2), reduced to 98’227 by applying the Massacrator routine described in section 3.

The “narrow” matching exercise resulted in 4'503 matches. Of these 261 matches were filtered out based on the age filter, while 647 were caught by the discipline filter, for a total of 857 filtered-out matches (61 matches were caught by both filters).

As for the “broad” matching exercise, this requires first to split the original “name+surname+middle names” filed according to which EP-INV inventors are classified into different fields for surname, first name, and middle names. In order to do so, the following preliminary steps were undertaken, in order to take into account specificities of the French language:

1. ‘M’ and ‘M.’ (which stands for Mr or Ms) before surname were excluded
2. First words in the original “surname+first name+middle names” string were placed in a separate “surname” field. Some substrings were identified as default parts of surnames. They were: 'BEN', 'DA', 'DEL', 'DI', 'DU', 'EL', 'LA', 'LE', 'VON', 'SAINT', 'SAINTE', 'DAL', 'MC', 'DO', 'DOS', 'DES', ‘DE’, ‘DE LA’, ‘DE SAINT’, ‘VAN’, ‘VAN DE/DEN/DER’, ‘AB DER’ in the beginning of the “surname+first name+middle names” string; and ‘DE’, ‘DE LA’, ‘DES’, ‘DA’, ‘VAN DER’, ‘VON’ in the middle of the string.
3. After surnames were extracted from the “surname+first name+middle names” string, we checked manually the remaining “first name+middle names” the string, whenever they contained blanks, to check for double surnames as opposed to names-middles names. For example, BURNOUF RADOSEVICH MIRYANA was split automatically into BURNOUF, RADOSEVICH MIRYANA while the correct split should have been BURNOUF RADOSEVICH, MIRYANA: the manual checking allowed to correct the mistake.
4. First or only words in the checked “first name+middle names” string were placed in a separate “first name field”.

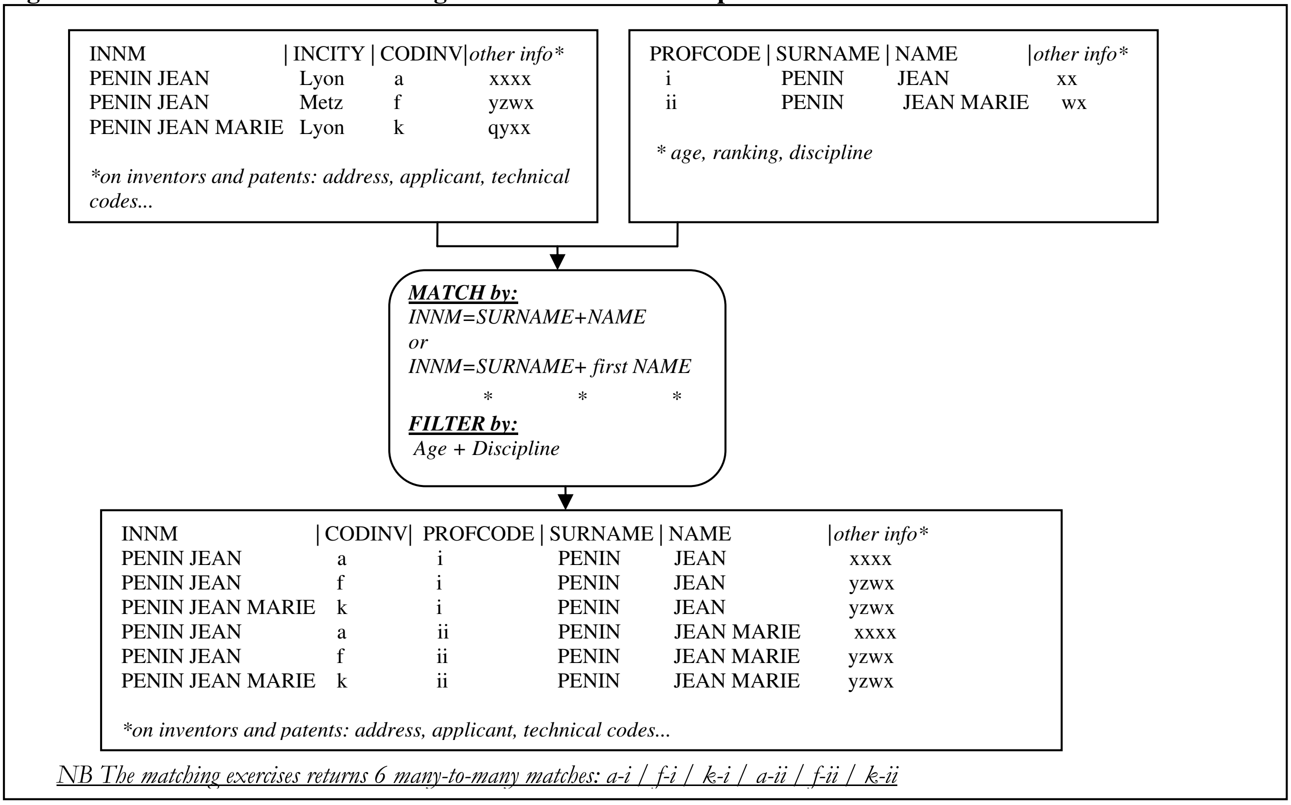
A similar procedure was then applied to names in the French PROFLIST. This allowed matching again professors and inventors, this time only by surname and first name. “Broad” matching resulted in 9’270 CODINV-PROFCODE pairs, reduced to 7’100 after filtering by age and disciplines. As the “broad” match includes 3'646 pairs from the “narrow” match, the net result of “broad” matching and filtering is 3'454 “extra” CODINV-PROFCODE pairs.

A further reduction followed from the exclusion of professors-inventor matches with same surnames and names, but different middle names. This leaves only matches with a professor (inventor) with both first name and middle names, and an inventor (professor) with first name only[11]. This reduced the result of the "broad" matching to 1'019 pairs.

Finally, we combined the results of the narrow and broad matches, for a total of 4'731 pairs, to be checked according to the procedure we describe in the next section.

One drawback of the matching exercise so described is that it does not return only unique (i.e. one professor-to-one inventor) matches, but also a certain number of one-to-many, many-to-one, and many-to-many matches (see figure 4).

Figure 4 – Problems in name-matching between inventors and professors



INNM	INCITY	CODINV	*other info**
PENIN JEAN	Lyon	a	xxxx
PENIN JEAN	Metz	f	yzwx
PENIN JEAN MARIE	Lyon	k	qyxx

**on inventors and patents: address, applicant, technical codes...*

PROFCODE	SURNAME	NAME	*other info**
i	PENIN	JEAN	xx
ii	PENIN	JEAN MARIE	wx

** age, ranking, discipline*

MATCH by:
INNM=SURNAME+NAME
or
INNM=SURNAME+ first NAME
* * *
FILTER by:
Age + Discipline

INNM	CODINV	PROFCODE	SURNAME	NAME	*other info**
PENIN JEAN	a	i	PENIN	JEAN	xxxx
PENIN JEAN	f	i	PENIN	JEAN	yzwx
PENIN JEAN MARIE	k	i	PENIN	JEAN	yzwx
PENIN JEAN	a	ii	PENIN	JEAN MARIE	xxxx
PENIN JEAN	f	ii	PENIN	JEAN MARIE	yzwx
PENIN JEAN MARIE	k	ii	PENIN	JEAN MARIE	yzwx

**on inventors and patents: address, applicant, technical codes...*

NB The matching exercises returns 6 many-to-many matches: a-i / f-i / k-i / a-ii / f-ii / k-ii

These multiple matches may be due either to homonymy (e.g. two inventors with identical name and surname match one professors with that name and surname) or to incomplete identification of inventors by Massacrator© (e.g. the two inventors with identical name and surname matched to one professors are indeed the same person, even though Massacrator© assigned them a score lower than the threshold value). To mitigate the latter problem all pairs of inventors (CODINV) matching the same professor were manually checked and 156 cases with incomplete identification were found and corrected. Thus the result of matching for French non-medical professors could be further reduced to 4'575 pairs. However the problem of homonymy could not be solved on the basis of available information. Only e-mail and phone contacts can clarify these issues.

[11] For example, matches such as 'MARTIN JEAN PIERRE'-'MARTIN JEAN MICHEL' were excluded (first names coincide, middle names do not), while matches such as 'MARTIN JEAN'-'MARTIN JEAN MICHEL' were retained

5. E-mail/phone checking

It is important to remark that at this stage we could not say whether the matched professor-inventor resulting from the previous steps are indeed the same person, or just homonyms. Further information is necessary to clarify this issue.

Some useful information is contained in the EP-CESPRI database, which reports the names of the patent applicants. Whenever the matched inventor is found to be designated on at least one patent application by either a university, a public research organization, or a no-profit institution known for sponsoring academic research, we can conclude that the professor-inventor match is a sound one (i.e. not a case of homonymy) and could be retained as a "true" academic inventor.

Furthermore, co-inventors of academic inventors identified in this way can be also assumed to be academic inventors. This allows to apply an iterative procedure.

In the exemplary case of French non-medical professors (see previous section), this procedure allowed to confirm 1116 academic inventors and 164 academic co-inventors, for a total of 1'148. The remaining 3025 professor-inventor matches had to be checked by contacting the professor-inventor matches by e-mail or phone. This in turn required first to retrieve the e-mail address or phone number of the professors, and then to ask them to confirm to be the authors of the patents our matching exercise attributed to them.

Both steps were performed by producing an Access database to be used by research assistants, which is fully described in Appendix 3. Each record of the database contains information on one professor, on the inventor(s) who have been matched to him/her, and on these inventor(s)' patents (such as IPC codes, applicant, priority date and title); it also contains a few blank fields to be filled in by the mask user (such as e-mail address, phone number, and a number of yes/no fields to indicate whether the professor has been contacted, and whether he/she has confirmed to be the same person as the matched inventor).

The database was endowed with a mask (MatchMask©) with a number of useful action keys and windows[12]:

- *Google key and window*: it connects to Google search engine, feeds it with all the information on the professor, and returns the search results → if these results include e-mail and phone contacts the mask users cut-and-paste them into the dedicated fields;
- *Patent window*: it contains the full list of all patents signed by the matched inventors, including their title, IPC code and applicant; the mask user may open/close it according to his/her needs;
- *Co-inventor window*: it contains the name and surname of all co-inventors listed on the patents which appear in the "patent window" and have also been matched to one or more professors; over

[12] Francesco Lissoni and Christian Catalini are the authors of MatchMask©

phone interviews with one professor, the interviewer can ask the interviewee to confirm whether the listed co-inventors also were academic scientists at the time of the patent;

- *E-mail key*: once the e-mail provided has been filled, clicking the E-mail key allows sending a default letter to the e-mail address owner, asking to confirm he/she is the inventor of the listed patents and information on the identity of the co-inventors.

An important warning is due about France: due to the large numbers of original professor-inventor matches, we could not check ALL matches by e-mail or phone. Therefore we decide to check only the professor-inventor pairs wherein the inventor's latest patent had been filed after 1993, in order to maximize our chances that the inventors would still be active and reachable. Any cross-country comparison based on the KEINS database should therefore be based only on academic inventors and inventors still patenting after 1993.

6. Results and summary statistics

6.1 Italy

As mentioned in section 3.1 the Italian part of the KEINS database has been built upon the earlier data published in Balconi et al. (2004). From the original list of "hard science" professors active in Italian universities in 2000 (PROFLIST 2000 henceforth), 918 academic inventors were identified, through a matching exercise with inventors active from 1978 to 1999. In 2005 a new list of professors was obtained (PROFLIST 2005). We then run 2 complementary matching exercises:

1. "New" professors, i.e. professors who appear PROFLIST 2005, but not PROFLIST 2000[13], were matched to all inventors from the EP-INV database (this allowed us to identify academic patents signed by these new professors both after and before their formal appointment);
2. "Old" professors, from PROFLIST 2000, were matched to "new" inventors from the EP-INV, i.e. inventors who signed their first patents after 1999 (this allowed us to identify new patents signed by those academic inventors already identified by previous research; or patents taken by "old" professors who turned to patenting only after 1999).

Matching procedures described in section 4 produced 690 professor-inventor pairs. Direct contacts (phone calls and e-mailing) allowed to us collect information on 295 pairs, 237 of which were confirmed as good ones (i.e. the professor confirmed to be the inventor[14]). Confirmed academic inventors, or information from collected CVs, allowed us to check 346 more pairs, 240 of which turned out to be good ones. Only

[13] That is to say that the "new" professors are either those who have entered academia between 2000 and 2005.

[14] Notice that, being the results of Massacrator© imperfect, it could still be the case that one professor was matched to more than one inventor with the same names and surnames; by confirming he was the same person as all of these inventors, we realized these inventors were the same person. This implies that the number of matches is higher than the number of professors involved

45 professors (corresponding to 49 professor-inventor pairs) were not traceable with internet nor answered to our e-mails and phone calls. As any of 45 non-responding professors may be an academic inventor we present our results as a range with the number of professors for whom match has been confirmed as a lower bound, and the sum of this number and the number of non-responding professors as an upper bound.

The overall result is 1271 identified academic inventors, and in addition 45 possible academic inventors (corresponding to 49 matches), i.e. the true number of academic inventors for Italy is between 1271 and 1313.

Table 6 – Distribution of Italian academic inventors across academic disciplines

DISCIPLINE	*Total number of professors*	*Acad.inventors (min)*[15]	*Acad.inventors (max)*[16]	***AI min(%)***[17]	***AI max(%)***[18]
Information science	688	11	12	1.60	1.74
Nuclear Physics	173	3	3	1.73	1.73
Astronomy & Astrophysics	187	0	0	0.00	0.00
Physics (other fields)	2186	65	67	2.97	3.06
Chemistry (theoretical)	1402	102	103	7.28	7.35
Organic & Industrial Chemistry; Materials	1182	159	160	13.45	13.54
Pharmaceutical chemistry	676	91	91	13.46	13.46
Earth sciences	1305	4	4	0.31	0.31
Biological disciplines (others)	1430	15	17	1.05	1.19
Pharmacology & pharmacological biology	762	52	55	6.82	7.22
Life sciences (biological disciplines)	3002	152	153	5.06	5.10
Life sciences (medical disciplines)	2278	94	97	4.13	4.26
Medical disciplines (others)	8820	120	133	1.36	1.51
Agricultural & Veterinary sciences	2009	37	38	1.84	1.89
Mechanical & Civil engineering	3132	60	68	1.92	2.17
Information & Electronic engineering	2316	192	195	8.29	8.42
Chemical engineering; Energy	1192	114	117	9.56	9.82
TOTAL:	32740	1271	1313	3.88	4.01

Table 6 presents the distribution of Italian academic inventors by academic fields. Table 7 presents the distribution of “academic patents”, i.e. patents where at least one of the inventors is a university professor.

Table 7 – Distribution of Italian academic patents across academic disciplines

DISCIPLINE_NAME	*Academic Patents (min)*	*Academic Patents (max)*
Information science	14	16
Nuclear Physics	3	3
Physics (other fields)	77	79
Chemistry (theoretical)	178	179
Organic & Industrial Chemistry; Materials	420	424
Pharmaceutical chemistry	168	168
Earth sciences	7	7
Biological disciplines (others)	14	16
Pharmacology & pharmacological biology	72	76
Life sciences (biological disciplines)	225	227
Life sciences (medical disciplines)	128	141
Medical disciplines (others)	187	230
Agricultural & Veterinary sciences	46	49
Mechanical & Civil engineering	97	124
Information & Electronic engineering	378	380
Chemical engineering; Energy	189	192
TOTAL:	2203	2311

6.2 France

As explained in section, 5 due to the large number of professor-inventor pairs we have chosen to focus on the pairs with inventors having their latest patents after 1993, in total 3951 pairs. Information on 2884

[15] Only confirmed matches (lower bound).

[16] Confirmed and possible academic inventors (upper bound). See explanation in the text.

[17] Confirmed academic inventors as percentage of total number of Italian professors in the academic field.

[18] Confirmed and possible academic inventors as percentage of total number of Italian professors in the academic field

pairs was collected through direct contact (2400 pairs) and via examining professors CVs, publications etc. or inquiring their academic co-inventors (484 pairs). Of these 2884 pairs 1324 pairs correspond to 1235 academic inventors. 587 professors (corresponding to 1067 pairs) either were not traceable via internet or never answered any e-mail or phone call, nor had posted any useful on their websites, if they had any.

The distributions of academic inventors and their patents across academic disciplines are shown in Tables 8 and 9 (with same notations as in Tables 6 and 7).

Table 8 – Distribution of French academic inventors across academic disciplines*

DISCIPLINE	*Total number of professors*	*Acad.inventors (min)*	*Acad.inventors (max)*	***AI min(%)***	***AI max(%)***
Mathematics	3335	13	77	0.39	2.31
Information science	2935	22	80	0.75	2.73
Nuclear Physics	448	7	12	1.56	2.68
Astronomy & Astrophysics	155	0	1	0.00	0.65
Physics (other fields)	2212	61	103	2.76	4.66
Chemistry (theoretical)	963	50	61	5.19	6.33
Organic & Industrial Chemistry; Materials	2327	235	289	10.10	12.42
Pharmaceutical chemistry	539	45	59	8.35	10.95
Earth sciences	1090	1	16	0.09	1.47
Biological disciplines (others)	1476	33	52	2.24	3.52
Pharmacology & pharmacological biology	1259	65	90	5.16	7.15
Life sciences (biological disciplines)	2710	132	171	4.87	6.31
Life sciences (medical disciplines)	2674	130	167	4.86	6.25
Medical disciplines (others)	3507	118	186	3.36	5.30
Mechanical & Civil engineering	2052	35	71	1.71	3.46
Information & Electronic engineering	3300	219	289	6.64	8.76
Chemical engineering; Energy	1024	69	98	6.74	9.57
TOTALS:	32006	1235	1822	3.86	5.69

* Only academic inventors whose last patent dated after 1993

Table 9 – Distribution of French academic patents across academic disciplines

DISCIPLINE	*Academic Patents (min)*	*Academic Patents (max)*
Mathematics	52	358
Information science	50	369
Nuclear Physics	8	25
Astronomy & Astrophysics	0	1
Physics (other fields)	168	315
Chemistry (theoretical)	101	148
Organic & Industrial Chemistry; Materials	702	925
Pharmaceutical chemistry	98	146
Earth sciences	3	63
Biological disciplines (others)	44	113
Pharmacology & pharmacological biology	183	254
Life sciences (biological disciplines)	313	439
Life sciences (medical disciplines)	308	402
Medical disciplines (others)	232	411
Mechanical & Civil engineering	46	107
Information & Electronic engineering	422	722
Chemical engineering; Energy	200	296
TOTAL:	2930	5094

6.3 Sweden

The major challenge to matching Swedish PROFLIST with EPO data was to handle the problem with few frequent surnames (see also section 2). The matching procedure described in section 4 applied to Swedish data produced over 11’000 professor-inventor pairs. Furthermore, filtering by academic discipline was not always possible as for many professors the academic discipline was missing.

Therefore we took a different way to address the problem of filtering. As in most cases Swedish PROFLIST includes the professor’s private address, we chose filtering by postcode (with consequent manual check for inventor and professor addresses). In addition to that we have done matching by postcode and filtering by name.[19] In this way we have identify 570 academic inventors corresponding to 570 professor-inventor pairs.

Next we matched co-inventors of 570 academic inventors with the PROFLIST (excluding already identified academic inventors). The result was 827 professor-inventor pairs. Their examination revealed further 22 pairs matching by address.[20] Further 132 pairs were contacted (with email and fax), 621 pairs were checked with information available on internet (professors’ CVs, publications etc.) or via their co-inventors-known academic inventors. The other 52 pairs were either not traceable via internet or never answered any e-mail or fax. The overall result is 726 confirmed academic inventors (751 pairs) and 48 possible academic inventors.

The distributions of academic inventors and their patents across academic disciplines are shown in Tables 10 and 11 (academic inventors having academic discipline assigned).

Table 10 – Distribution of Swedish academic inventors across academic disciplines[21]

DISCIPLINE	*Total number of professors*	*Acad.inventors (min)*	*Acad.inventors (max)*	***AI min(%)***	***AI max(%)***
Information science	937	9	9	0.96	0.96
Nuclear Physics	200	0	0	0.00	0.00
Astronomy & Astrophysics	7	0	0	0.00	0.00
Physics (other fields)	25	0	0	0.00	0.00
Chemistry (theoretical)	1005	28	30	2.79	2.99
Organic & Industrial Chemistry; Materials	976	40	41	4.10	4.20
Pharmaceutical chemistry	431	19	19	4.41	4.41
Earth sciences	401	0	0	0.00	0.00
Biological disciplines (others)	1518	29	31	1.91	2.04
Pharmacology & pharmacological biology	441	30	34	6.80	7.71
Life sciences (biological disciplines)	306	21	21	6.86	6.86
Life sciences (medical disciplines)	1139	73	74	6.41	6.50
Medical disciplines (others)	1635	19	20	1.16	1.22
Agricultural & Veterinary sciences	1112	21	22	1.89	1.98
Mechanical & Civil engineering	2087	27	28	1.29	1.34
Information & Electronic engineering	2051	69	73	3.36	3.56
Chemical engineering; Energy	419	26	27	6.21	6.44
TOTAL:	14690	411	429	2.80	2.92

[19] Using Levenshtein edit distance equal to 3 as a threshold with consequent manual check.

[20] Those pairs had a mistake in the postal code, and were not captured through postal code filter.

[21] For notations see Table 6.

Table 11 – Distribution of Swedish academic patents across academic disciplines

DISCIPLINE_NAME	*Academic Patents (min)*	*Academic Patents (max)*
Mathematics	17	17
Physics (other fields)	56	61
Chemistry (theoretical)	75	103
Organic & Industrial Chemistry; Materials	39	39
Biological disciplines (others)	38	41
Pharmacology & pharmacological biology	73	84
Life sciences (biological disciplines)	50	50
Life sciences (medical disciplines)	155	157
Medical disciplines (others)	33	38
Agricultural & Veterinary sciences	31	34
Mechanical & Civil engineering	44	45
Information & Electronic engineering	271	289
Chemical engineering; Energy	62	63
TOTAL:	1574	1710

7. Conclusions: database access rules and research plans

The KEINS database contains sensitive information, such as the names, surnames, gender and patenting activity of a number of academic professors. It also contains data whose exploitation by the KEINS research partners will require some time. Accordingly, a number of related dataset have been created, which vary in terms of contents and will be delivered at different times and with different rules of access. In Appendix 4, we report the agreement reached by the KEINS research partners and the European Commission on these matters. Researchers interested in data from the KEINS database may wish first to familiarize with that agreement, and then contact CESPRI for further enquiries.

The quality of data for France and Sweden is susceptible of improvement. For France, one ought to check also professor-inventor pairs with patents dating back before 1993. For Sweden, the present PROFLIST is likely to include too large a number of lecturers whose task do not include research, which in turn create a downward bias in estimations of the share of academic inventors over universities' scientific staff; in addition, less restrictive matching exercise could reveal the existence of more academic inventors. In due course, CESPRI will release updated versions of various tables of the KEINS addressing these problems.

KEINS partners, either jointly or individually, will use the data described in this paper to first and foremost to re-assess current evaluations and beliefs on the contribution of European universities to technology transfer, at least for that part of transfer than can be measured through patent data. As displayed in section 6, no less than 3.5% of French and Italian academic scientists are inventors, that is have signed at least one European patent; in Sweden, this figure is just under 3%, but we have reason to believe is slightly underestimated. By taking into account national patents these figures could be revised upward in all countries. It is a far cry from the conventional wisdom, that portray European academic research as detached from the industrial one, one which deserves further investigation: how valuable are these patents? how many of them are owned by business companies (as a result of contract research) or by public funding agencies (as a result of research grants) or by universities (possibly as a result of disclosures to technology liaison offices)? how many of the funding agency- or university-owned patents are licensed, and bring some revenues to their owners? are there differences across disciplines and universities? Answers to these questions will provide the first, broad quantitative assessment of academic patenting in Europe, finally amenable to comparisons with the extensive literature on the US case.

Further research questions, which the KEINS research partners will try to answer to, relate to how contemporary science relate to inventive activity. how much do these patents owe to their inventors' scientific activity, as described in their publications? are they the results of applied, possibly menial research efforts, which distract scientists from fundamental research projects, or do they follow the latter, and possibly bring resources for it? does patenting help a scientists' career? in what fields? how influential are academic patents and inventors for the development of a specific technology?

Some of these research questions, and many more that perspective users of the KEINS database may come up with, call for the collection of complementary data, such as patent citations, information on licensing, and scientific publications. All of these data can be connected to the KEINS database via the patent application number, or the name and surname of the individual scientist (although access to the latter may require special arrangements; see above).

In addition, the methodology followed for the KEINS database can be easily applied to data from other countries.

References

Balconi M., Breschi S., Lissoni F. (2004) “Networks of inventors and the role of academia: an exploration of Italian patent data” , *Research Policy* 33/1, pp. 127-145

Breschi S., Cassi L., Malerba F. (2004), *Data and methodological issues on patents and patent citations*, STI-NET Project Report to the European Commission, http://www.stinet.org/docs/cespri_2.pdf

Breschi S., Lissoni F. (2004) “Knowledge networks from patent data: Methodological issues and research targets”, in: Glänzel W., Moed H., Schmoch U. (eds), *Handbook of Quantitative S&T Research*, Kluwer Academic Publishers

Lissoni F., Llerena P., Penin J., McKelvey M., Sanditov B., Schild I. (2006), “Academic patents in Europe 1990-2004: comparative data for France, Italy, and Sweden”, mimeo

OST (2004), *Indicateurs de Sciences et de Technologies – Rapport 2004*, Observatoire de Sciences et de Technologies, Paris (http://www.obs-ost.fr/services/rapport_ost/)

Singh J. (2005) “Social Networks as Determinants of Knowledge Diffusion Patterns”, *Management Science*

APPENDIX 1 – THE EP-CESPRI AND THE EP-INV DATABASES

The EP-CESPRI database contains, split into different relational tables, three families of data:

- PATENT DATA: application year, IPC class, title, abstract, equivalences, priority dates, designated countries, granted flag…

- APPLICANT DATA: Applicant name, address, company and group to whom the applicant belongs to, SIC codes, history of the applicant (merging, acquisitions…)

- CITATIONS DATA : Patents and non patent literature cited by the patents

The EP-INV database consists of one further family of data, namely:

- INVENTOR DATA: Inventors' name and address

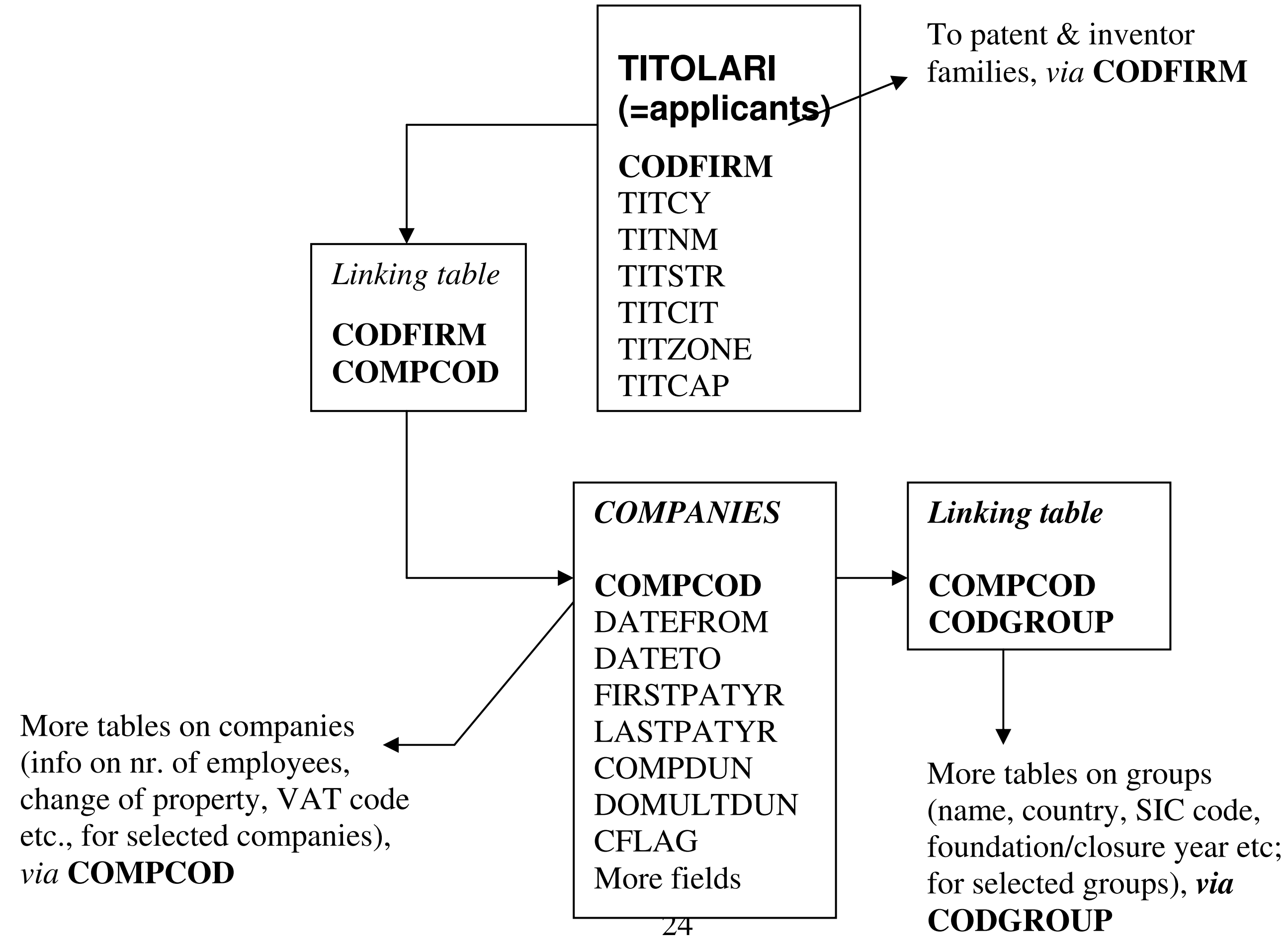
One more table (COMP_FOUNDERS) can be considered part of both applicants 'and inventors' family, since it contains the founders of the companies and contains a unique inventor's code (CODINV), if the founder is an inventor.

The main link among the four families is the Patent Publication Number (PUNR).

Below follow a simplified description of the EP-CESPRI and EP-INV tables, by family, and a description of the main fields

<u>APPLICANT TABLES FAMILY</u> <u>(selected tables and fields)</u>

In ***italics:*** Table name
In **bold**: Key Fields in the table


TITOLARI
(=applicants)
CODFIRM
TITCY
TITNM
TITSTR
TITCIT
TITZONE
TITCAP
To patent & inventor families, via CODFIRM
Linking table
CODFIRM
COMPCOD
COMPANIES
COMPCOD
DATEFROM
DATETO
FIRSTPATYR
LASTPATYR
COMPDUN
DOMULTDUN
CFLAG
More fields
Linking table
COMPCOD
CODGROUP
More tables on companies (info on nr. of employees, change of property, VAT code etc., for selected companies), via COMPCOD
More tables on groups (name, country, SIC code, foundation/closure year etc; for selected groups), via CODGROUP

LEGEND (by selected table)

TITOLARI **-** Anagraphic data of applicants, as derived from EPO documents, after parsing & normalization[1]

Name of field	*Format*	*Contents of field*
CODFIRM	**9(8)**	**Progressive applicant number**
TITCY	$(2)	Applicant's country
TITNM	$(255)	Applicant's name
TITTRDNM	$(255)	Applicant's trade name
TITSTR	$(255)	Applicant's address (street and number)
TITCIT	$(255)	Applicant's city
TITZONE	$(75)	Applicant's "zone" → 3 fields, by level of detail: province/region/state (if applicable)
TITCAP	$(10)	Applicant's ZIP code

[1] Applicants whose names and/or address differ for less than 3 characters are assigned the same CODFIRM, and the name and address of the modal observation

COMPANIES **-** Anagraphic data of companies, as derived from various sources (various research projects[1])

Name of field	*Format*	*Contents of field*
COMPCOD	9(8)	Progressive company number
DATEFROM	9(4)	Starting date
DATETO	9(4)	Closing date
FIRSTPATYR	9(4)	First patent's year
LASTPATYR	9(4)	Last patent's year
COMPDUN	$(10)	company's dun number
DOMULTDUN	$(10)	Dun&Bradstreet number of ultimate domestic parent
CFLAG	$(1)	Origin of data [1]

[1] CFLAG contents

A = Dun&Bradstreet plus manual checking, from STINET research project (Semiconductors; Pharmaceuticals… worldwide)

B= DFpowerstudio plus manual checking from STINET research project

C = Dun&Bradstreet plus manual checking, from STINET research project (but no info on global parent company)

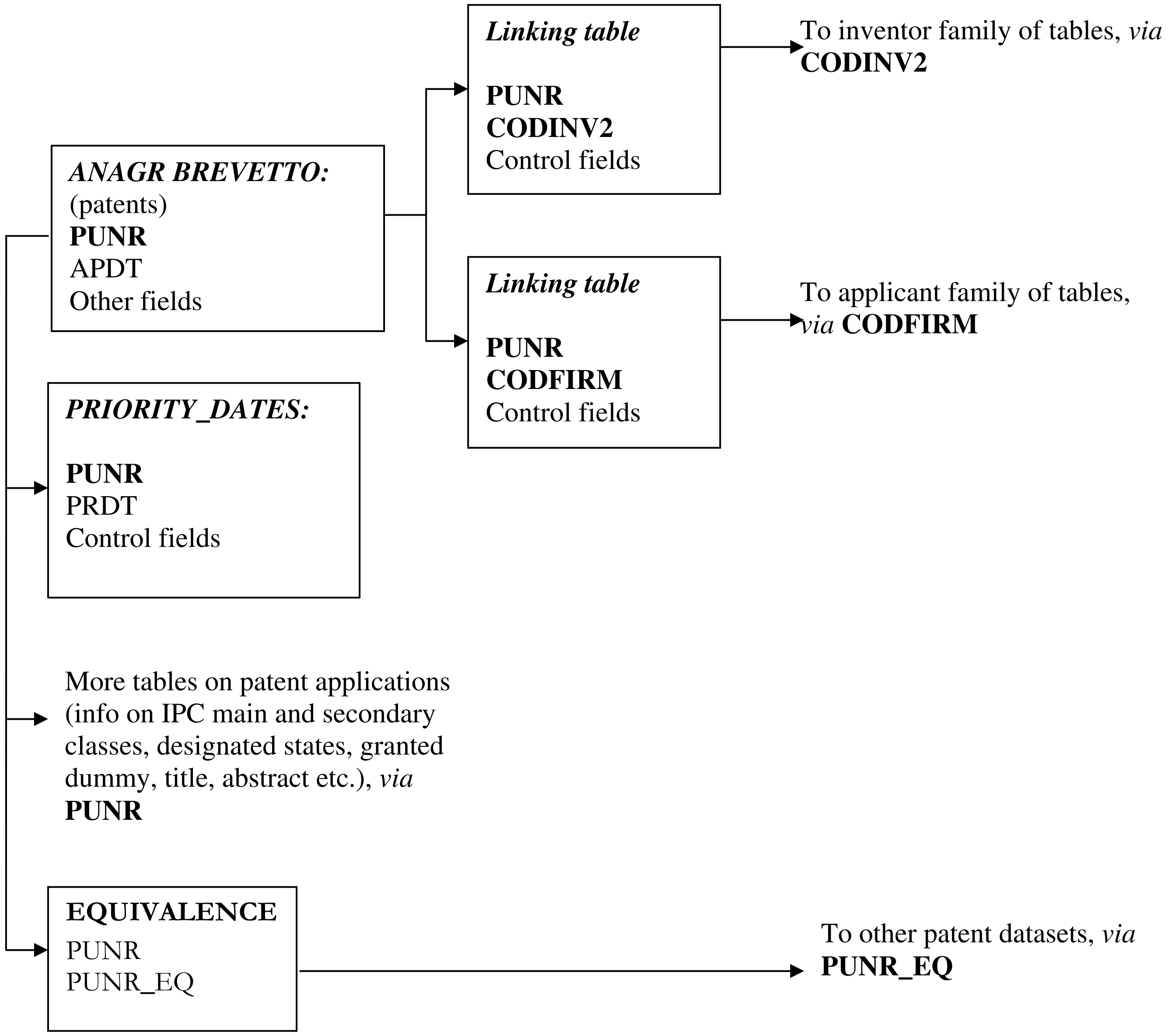
D = Manual checking (from STINET research project)

F = Fabio Montobbio's manual checking

NB: CODFIRM identifies applicants by normalized name and address
COMPCOD identifies applicants by normalized name and country

PATENT TABLES FAMILY (selected tables and fields)

In ***italics:*** Table name
In **bold**: Key Fields in the table

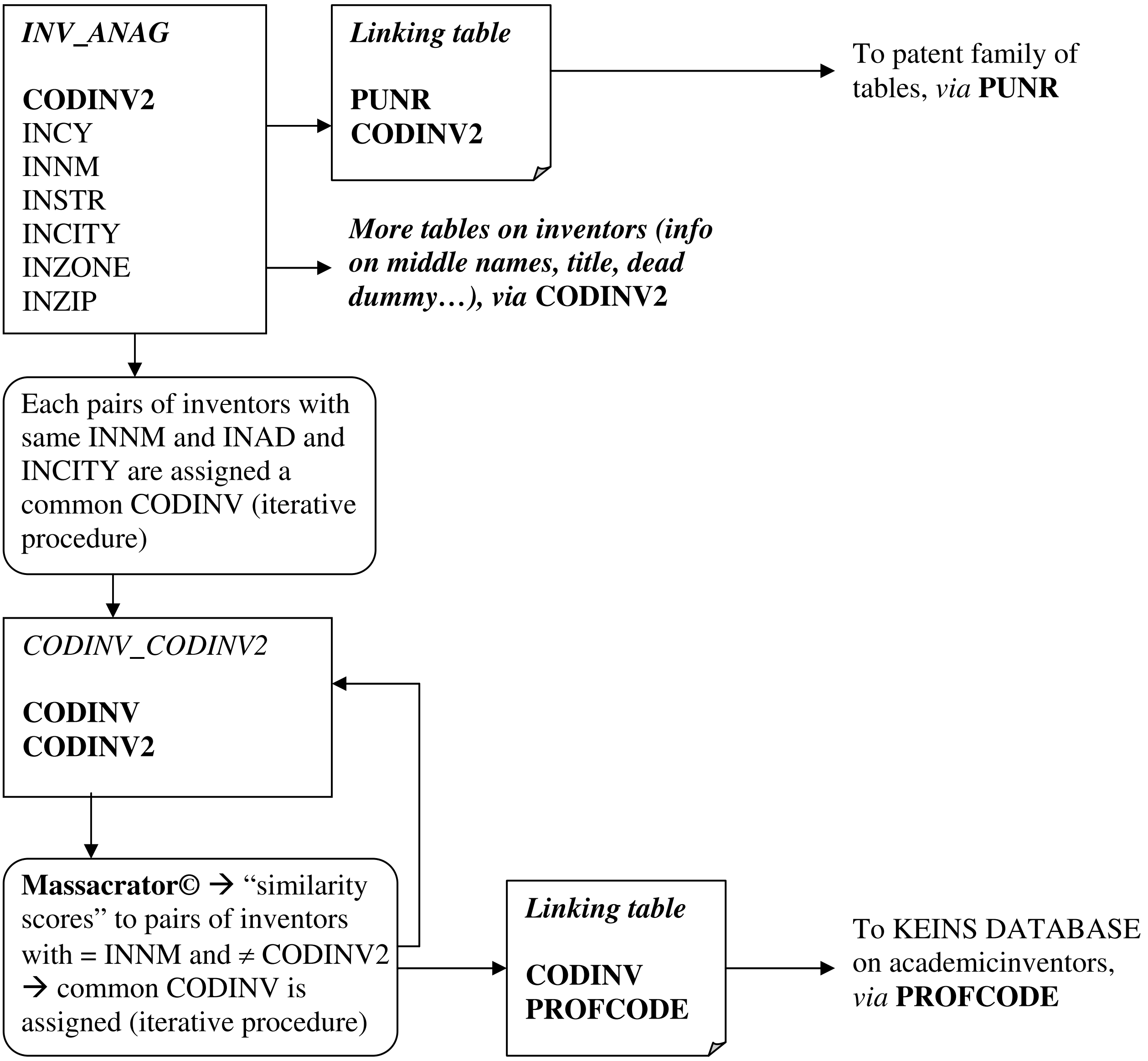

Linking table
PUNR
CODINV2
Control fields
To inventor family of tables, *via* **CODINV2**
ANAGR BREVETTO:
(patents)
PUNR
APDT
Other fields
Linking table
PUNR
CODFIRM
Control fields
To applicant family of tables, *via* **CODFIRM**
PRIORITY_DATES:
PUNR
PRDT
Control fields
More tables on patent applications (info on IPC main and secondary classes, designated states, granted dummy, title, abstract etc.), *via* **PUNR**
EQUIVALENCE
PUNR
PUNR_EQ
To other patent datasets, *via* **PUNR_EQ**

LEGEND (selected fields)

Name of field	*Format*	*Contents of field*
PUNR	9(8)	Progressive EPO Number
APDT	dd/mm/yy	Date of filing at EPO
CODFIRM	9(8)	Progressive applicant number
CODINV2	9(8)	Inventor code; unique for address / name
PUNR_EQ	$(12)	Equivalent patent; format AA999999(B) Where: AA=patent office acronym ; 999999=patent nr
PRDT	9(8)	Priority date

INVENTOR TABLES FAMILY / EP-INV DATABASE (selected tables and fields)

In *italics:* Table name
In **bold**: Key Fields in the table



INV_ANAG

CODINV2
INCY
INNM
INSTR
INCITY
INZONE
INZIP

Linking table

PUNR
CODINV2

To patent family of tables, *via* **PUNR**

***More tables on inventors (info on middle names, title, dead dummy…), via* CODINV2**

Each pairs of inventors with same INNM and INAD and INCITY are assigned a common CODINV (iterative procedure)

CODINV_CODINV2

CODINV
CODINV2

Massacrator© → “similarity scores” to pairs of inventors with = INNM and ≠ CODINV2 → common CODINV is assigned (iterative procedure)

Linking table

CODINV
PROFCODE

To KEINS DATABASE on academicinventors, *via* **PROFCODE**

LEGEND (selected fields)

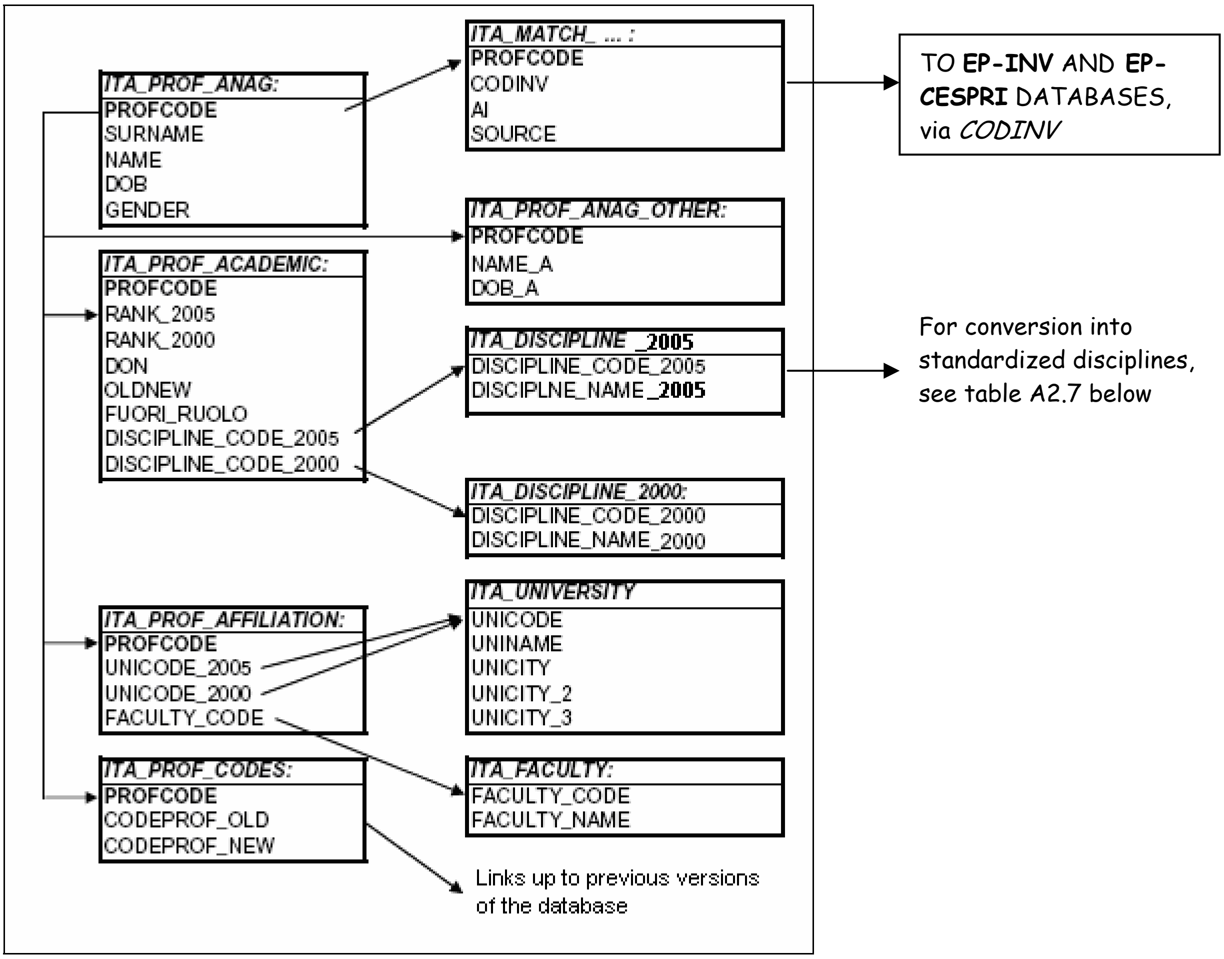
Name of field	*Format*	*Contents of field*
INCY	$(2)	Inventor’s country
INNM	$(255)	Inventor’s name (first name + middle names + surname)
INSTR	$(255)	Inventor’s address (street and number)
INCITY	$(255)	Inventor’s city
INZONE	$(255)	Inventor’s “zone” → 3 fields, by level of detail: province/region/state (if applicable)
INZIP	$(10)	Inventor’s zip code
CODINV2	9(8)	Inventor’s code; unique for INNM+INSTR+INCITY+INCY
CODINV	9(8)	Inventor’s code; unique for INNM [1]
PROFCODE	9(8)	Professor’s code (from KEINS database) [2]

[1] If checked either manually or through Massacrator ©: CODINV and CODINV2 differ for “mobile” inventors (same INNM but different address)

[2] Available for Italy, France and Sweden

APPENDIX 2 –PROFLISTS: FRANCE, ITALY, AND SWEDEN

Table A2.1 – Italian PROFLIST: set of tables



Links up to previous versions of the database

TO **EP-INV** AND **EP-CESPRI** DATABASES, via *CODINV*

For conversion into standardized disciplines, see table A2.7 below

Table A2.2 – Italian PROFLIST: list of variables

ITA_PROF_ANAG:	
Anagraphic information about Italian professors ("hard sciences" , active in 2005)	
PROFCODE	Professor's identification code
SURNAME	Professor's surname
NAME	Professor's first and middle names
DOB	Date of birth (DDMMMYYYY)
GENDER	'M' or 'F'

NB: PROFCODE is a progressive number according to the list of professors in ALL DISCIPLINES (not only "hard sciences") from both professors' lists of 2000 and 2005 (after matching the two lists).

ITA_PROF_ANAG_OTHER:	
Additional anagraphic data due to matching of old and new professors lists. "Alternative" information is one used in the old list.	
PROFCODE	Professor's identification code
NAME_A	Surname as it is in the old list (only if it differs from one in the new list)
DOB_A	DOB as it is in the old list (only if it differs from one in the new list)

ITA_PROF_ACADEMIC:	
Professors' academic profile	
PROFCODE	Professor's identification code
RANK_2005	Academic rank in 2005
RANK_2000	Academic rank in 2000
DON	Date of nomination (for RANK_2005)
OLDNEW	Dummy (=1) on whether professor was in the old dataset (2000)
FR	FR = "FUORI RUOLO" = Soon to retire (in 2005)
DISCIPLINE_CODE_2005	Code of academic discipline according to MIUR 2005
DISCIPLINE_CODE_2000	Code of academic discipline according to MIUR 2000

NB1: Academic rank: "RU"= researcher, "PA"=associated professor, "PO"=ordinary professor

NB2: OLDNEW = 1, if professor is in both old (2000) and new (2005) lists of Italian professors
OLDNEW = 2, if professor is in the new (2005) list, but not in the old list.

ITA_PROF_AFFILIATION:	
Professors' affiliation	
PROFCODE	Professor's identification code
UNICODE_2005	Code of the university where professor worked in 2005
UNICODE_2000	Code of the university where professor worked in 2000
FACULTY_CODE	Code of the faculty at which professor worked in 2005

ITA_PROF_CODES:	
Table of correspondence between old codes (CODEPROF) and new one	
PROFCODE	Professor's identification code
CODEPROF_OLD	Professor's identification code from old list of professors (2000), if any
CODEPROF_NEW	Professor's identification code from new list of professors (2005), if any

NB: The "old list" of professors is one used in Balconi, Breschi, and Lissoni (RP). It consists of professors in (most) "hard sciences" active in 2000. The "new list" of professors is the list from MIUR; it contains all professors across all Italian universities who were active in 2005.

ITA_DISCIPLINE_2005:	
Codes for scientific disciplines in 2005	
DISCIPLINE_CODE_ITALY_2005	Discipline code according to MIUR classification in 2005
DISCIPLINE_NAME_ITALY_2005	Name of scientific discipline according to MIUR classification in 2005

NB: For correspondence b/w 2005 and 2000 codes see http://www.miur.it/UserFiles/117.htm

ITA_DISCIPLINE_2000:	
Codes for scientific disciplines in 2000	
DISCIPLINE_CODE_ITALY_2000	Discipline code according to MIUR classification in 2000
DISCIPLINE_NAME_ITALY_2000	Name of scientific discipline according to MIUR classification in 2000

NB: For correspondence b/w 2005 and 2000 codes see http://www.miur.it/UserFiles/117.htm

Table A2.2 – Italian PROFLIST: list of variables (continues)

ITA_UNIVERSITY:	
Information on Italian universities	
UNICODE	University code
UNINAME	University name
UNICITY	City where university is located
UNICITY_2	Another location if university has more than 1 locations
UNICITY_3	Another location if university has more than 2 locations

ITA_FACULTY:	
Information on Italian faculties	
FACULTY_CODE	Code of the faculty where professor is worked in 2005
FACULTY_NAME	Full name of the faculty where professor is worked in 2005

ITA_MATCH	
Professor-inventor pairs, with indication of results of e-mail/phone/internet checking	
PROFCODE	Professor's code
CODINV	Inventor's code
AI	Dummy for academic inventor
LAST_PAT_PROF	Year in which the professor has last signed a patents
SOURCE	Source of the information for assigning AI

NB1: AI was assigned following values: = 0 , if not an academic inventor; = 1 , if an academic inventor; = . (missing value) , a potential match had not been checked (either the professor was not traceable or did not answer emails/phone calls).

NB2: SOURCE was assigned following values: = 'ADDRESS' , if professor and inventor have the same (private) address; = 'CONTACT' , if professor confirmed/denied that (s)he is the inventor of corresponding patent; = 'INTERNET' , if match was confirmed through studying professor's CV/publications OR the match was confirmed by coinventors; ='OLD' i match comes from early research by in Balconi, Breschi, and Lissoni (RP);= . (missing value) , a potential match had not been checked (either professor was not traceable or did not answer emails/phone calls).

Table A2.3 – French PROFLIST: set of tables

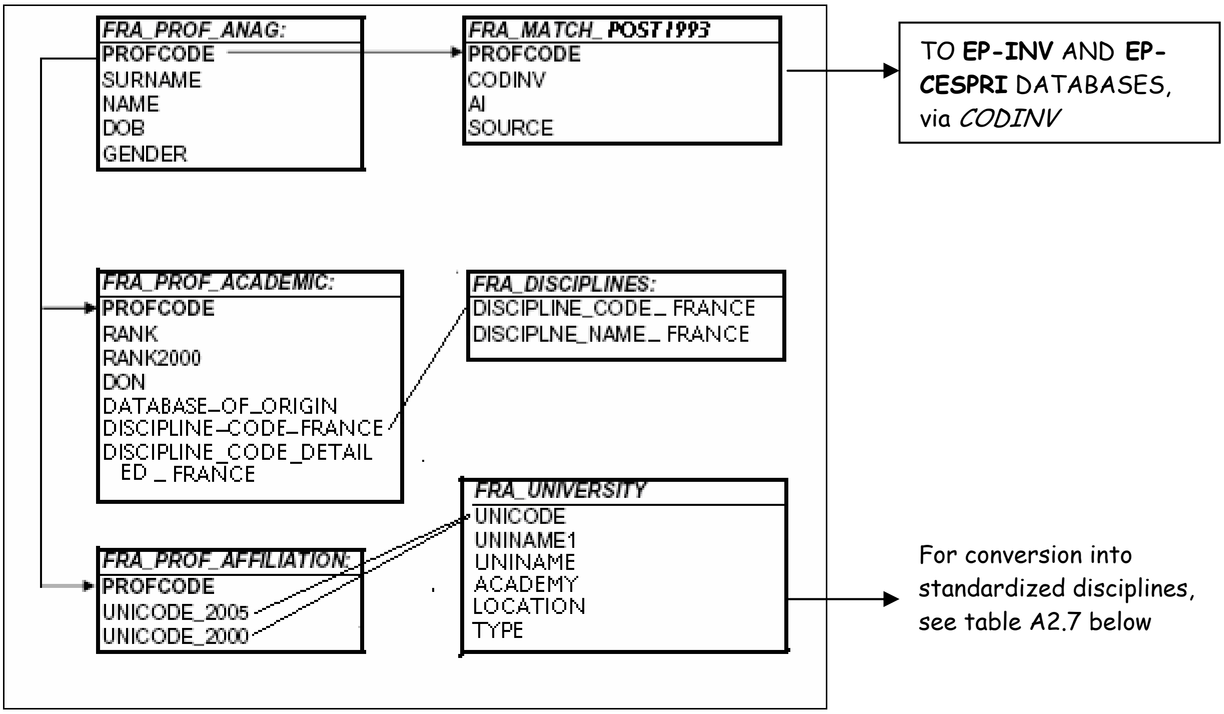


Table A2.4 – French PROFLIST: list of variables

FRA_PROF_ANAG:	
Anagraphic information about French professors in "hard" sciences (CNU 2500-6900), active in 2005.	
PROFCODE	Professor's identification code
SURNAME	Professor's surname
NAME	Professor's first and middle names
DOB	Year of birth
GENDER	'M' or 'F'

NB: PROFCODE is a progressive number according to the list of professors in ALL DISCIPLINES (not only "hard sciences").

FRA_PROF_ACADEMIC:	
Professors' academic profile	
PROFCODE	Professor's identification code
RANK	Academic rank in 2005
RANK_2000	Academic rank in 2000 (only for NONMED database)
DON	Date of nomination (for RANK_2005)
DATABASE_OF_ORIGIN	Indicates whether data come from a MED(ICAL) or a NONMED(ICAL) database (→*NB2*)
DISCIPLINE_CODE_FRANCE	Academic discipline, CNU (2-digit)
DISCIPLINE_CODE_DETAILED_FRANCE	Academic discipline, CNU (4-digit) (only for MED database)

NB1: Academic rank: "MCFU" = Maître de Conférences, "PR" = Professor, "OTH" = Others

NB2: We received different files for professors in MEDICAL and NONMEDICAL disciplines, with slight differences in contents

NB3: See http://www.education.gouv.fr/personnel/enseignant_superieur/enseignant_chercheur/section_cnu.htm contains more info on French disciplines(for "detailed" ones: http://www.education.gouv.fr/personnel/enseignant_superieur/enseignant_chercheur/cnusante.htm)

FRA_PROF_AFFILIATION:	
Professors' affiliation	
PROFCODE	Professor's identification code
UNICODE_2005	Affiliation in 2005
UNICODE_2000	Affiliation in 2000 (only for NONMED database)

FRA_DISCIPLINES:	
Codes for scientific disciplines in 2005	
DISCIPLINE_CODE_FRANCE	Discipline code, 4-digit CNU in 2005
DISCIPLINE_NAME_FRANCE	Name of scientific discipline

NB1: See: http://www.education.gouv.fr/personnel/enseignant_superieur/enseignant_chercheur/cnusante.htm) for names of "detailed" disciplines

FRA_UNIVERSITY:	
Information on French universities	
UNICODE	Unique code for each higher education institution
UNINAME1	Unique name for each higher education institution
UNINAME	Name of the aggregation of higher education institutions (by KEINS)
ACADEMY	Regional aggregation of higher education institutions (administrative unit)
LOCATION	City
TYPE	Type of higher education institution (→ NB)

NB: TYPEs are as follows: UNI = universities (include university hospitals and "instituts universitaires de technologie- IUT"
ENG. SCHOOL = engineering school
GRAND ETAB = Grands etablissements (e.g. Ecole Nationale Superieure)
INP = Institut National Polytechnique de Grenoble
IUFM = Institut Universitaire de Formation des Maîtres (preparatory schools for teachers)

Table A2.4 – French PROFLIST: list of variables (continues)

FRA_MATCH_POST1993:	
Professor-inventor pairs, with indication of results of e-mail/phone/internet checking (→ NB3)	
PROFCODE	Professor's code
CODINV	Inventor's code
AI	Dummy for academic inventor
LAST_PAT_PROF	Year in which the professor has last signed a patents (→ NB3)
SOURCE	Source of the information for assigning AI

NB1: AI was assigned following values: = 0 , if not an academic inventor; = 1 , if an academic inventor; = . (missing value) , a potential match had not been checked (either the professor was not traceable or did not answer emails/phone calls).

NB2: SOURCE was assigned following values: = 'ADDRESS' , if professor and inventor have the same (private) address; = 'CONTACT' , if professor confirmed/denied that (s)he is the inventor of corresponding patent; = 'INTERNET' , if match was confirmed through studying professor's CV/publications OR the match was confirmed by coinventors; = . (missing value) , a potential match had not been checked (either professor was not traceable or did not answer emails/phone calls).

NB3: The dataset contains only professor-inventor matches wherein the inventor's most recent patent was filed after 1993.

Table A2.5 – Swedish PROFLIST: set of tables

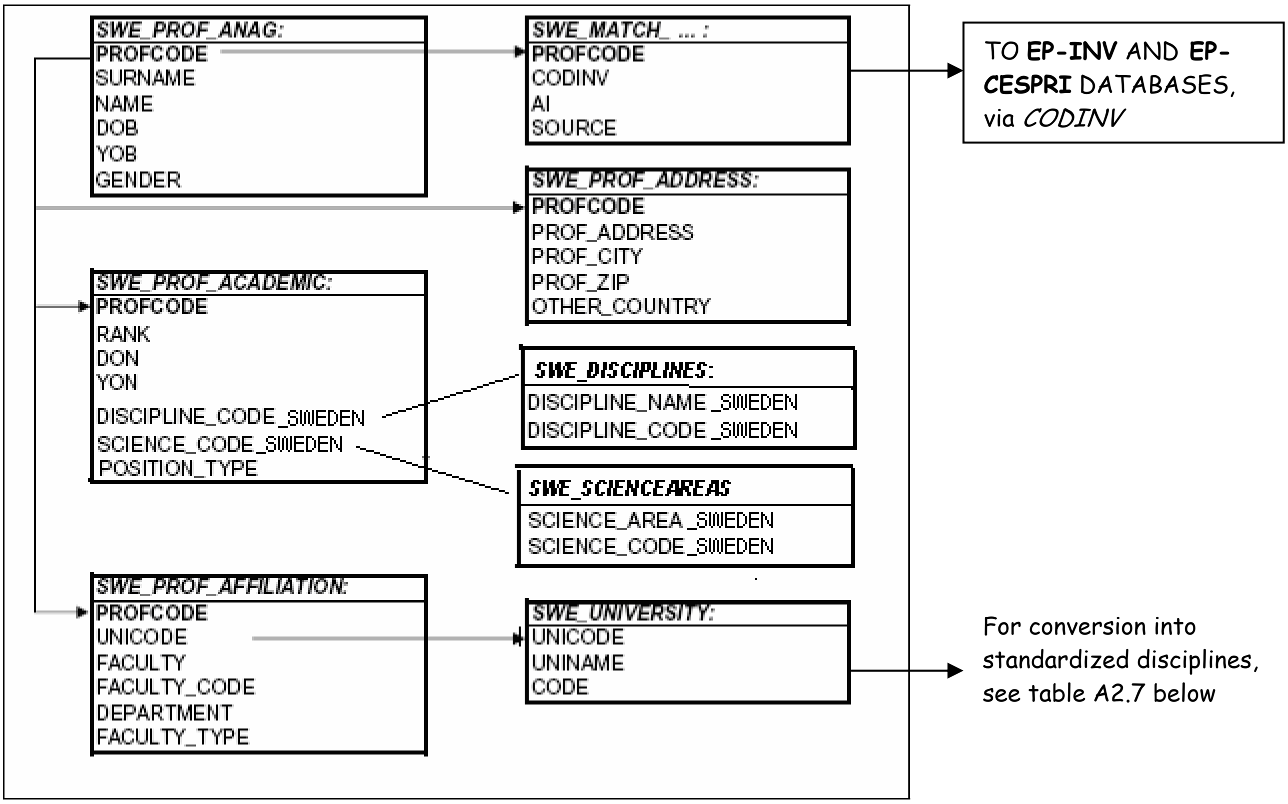


Table A2.6 – Swedish PROFLIST: list of variables

(Tables include all employees of 27 Swedish universities (see Table 4), humanities and social disciplines excluded)[22]

SWE_PROF_ANAG:

Anagraphic information about Swedish professors in “hard” sciences, active in 2005.

PROFCODE	Professor’s identification code
SURNAME	Professor’s surname
NAME	Professor’s first and middle names
DOB	Year of birth
GENDER	‘M’ or ‘F’

NB: PROFCODE is a progressive number according to the list ALL EMPOLYEES (including soft disciplines and non-academic positions).

SWE_PROF_ADDRESS:

Professors’ address (home).

PROFCODE	Professor’s identification code
PROF_ADDRESS	Professor’s street address
PROF_CITY	Professor’s city
PROF_ZIP	Professor’s postal code
OTHER_COUNTRY	Country (if the professor’s address is not in Sweden)

SWE_PROF_ACADEMIC:

Professors’ academic profile

PROFCODE	Professor’s identification code
RANK	Academic rank
DON	Date of employment
YON	Date of employment
DISCIPLINE_CODE_SWEDEN	Original (Swedish) academic discipline, 4-digit code
SCIENCE_CODE_SWEDEN	Science area, 1-digit code
POSITION_TYPE	Position type

SWE_DISCIPLINES:

Swedish disciplines’ names

DISCIPLINE_NAME_SWEDEN	Original (Swedish) academic discipline, description
DISCIPLINE_CODE_SWEDEN	Original (Swedish) academic discipline, 4-digit code

SWE_SCIENCEAREAS:

Swedish science areas’ names

SCIENCE_AREA_SWEDEN	Original (Swedish) science area, description
SCIENCE_CODE_SWEDEN	Original (Swedish) science area, 1-digit code

1: Rank (Academic positions): Professor (*Professor*), Associate Professor (*Adjungerad Professor*), Assistant Professor (*Bidrädande Professor*), Lecturer (*Lektor*), Junior Lecturer (*Adjunkt*), Research Fellow (*Forskarassistent*), Researcher (*Forskare*), Research Assistant (*Forskningsassistent*), and Other.

2: Swedish academic discipline codes have been produced according to a publication-oriented classification: http://www.ub.uu.se/epub/categories/

3: Science code/area have been obtained according to the following classification:
1 - Humanistisk-samhällsvetenskapligt (HUMANITIES-SOCIAL SCIENCES),
2 - Medicinskt (MEDICINE),
3 - Naturvetenskapligt (NATURAL SCIENCES),
4 - Tekniskt (TECHNOLOGY),
5 - SLU (THE SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES),
6 - Gemensamt/övrigt (JOINT/OTHER)
(see also: http://www.scb.se/templates/Standard____24458.asp)

4: Position type = Academic, Ph/Postdoc, Administrative, Technical, Library.

NB: Classifications by “Discipline” and by “Science areas” are alternative: the two classifications do not match completely, and some individuals who are classified by “Science area” are not classified by “Discipline” (and vice versa) → KEINS disciplines are based upon Swedish “Disciplines”, not “Science areas”

[22] Only observations with DISCPLINE_CODE/DISCPLINE_NAME not missing.

Table A2.6 – Swedish PROFLIST: list of variables (continues)

SWE_PROF_AFFILIATION:	
Professors' affiliation	
PROFCODE	Professor's identification code
UNICODE	Code of the university
FACULTY	Name of the faculty
FACULTY_CODE	Code of the faculty
DEPARTMENT	Name of the department
FACULTY_TYPE	Type of the unit (FACULTY) to which the employee is affiliated

NB: FACULTY_TYPE ('ADMINISTRATIVE', 'TECHNICAL', , 'LIBRARY' 'ADMINISTRATIVE/TECHNICAL', 'TECHNICAL-ADMINISTRATIVE') is defined only for non-academic units (FACULTY), and only for observations with FACULTY not missing.

SWE_UNIVERSITY:	
Information on Italian universities	
UNICODE	University code
UNINAME	University name
CODE	Numeric code

NB: CODE was used for building PROFCODE (first two digits).

SWE_MATCH:	
Professor-inventor pairs, with indication of results of e-mail/phone/internet checking	
PROFCODE	Professor's code
CODINV	Inventor's code
AI	Dummy for academic inventor
LAST_PAT_PROF	Year in which the professor has last signed a patents (→ NB3)
SOURCE	Source of the information for assigning AI

NB1: AI was assigned following values: = 0 , if not an academic inventor; = 1 , if an academic inventor; = . (missing value) , a potential match had not been checked (either the professor was not traceable or did not answer emails/phone calls).

NB2: SOURCE was assigned following values: = 'ADDRESS' , if professor and inventor have the same (private) address; = 'CONTACT' , if professor confirmed/denied that (s)he is the inventor of corresponding patent; = 'INTERNET' , if match was confirmed through studying professor's CV/publications OR the match was confirmed by coinventors; = . (missing value) , a potential match had not been checked (either professor was not traceable or did not answer emails/phone calls).

Table A2.7 – KEINS DISCIPLINES *

DISCIPLINE_CODE_KEINS	DISCIPLINE_NAME_KEINS
0	n.c. (not relevant for academic inventorship)
11	Mathematics
12	Information science
21	Nuclear Physics
22	Astronomy & Astrophysics
23	Physics (other fields)
31	Chemistry (theoretical)
32	Organic & Industrial Chemistry; Materials
33	Pharmaceutical chemistry
41	Earth sciences
51	Biological disciplines (others)
52	Pharmacology & pharmacological biology
53	Life sciences (biological disciplines)
61	Life sciences (medical disciplines)
62	Medical disciplines (others)
71	Agricultural & Veterinary sciences
81	Mechanical & Civil engineering
91	Information & Electronic engineering
92	Chemical engineering; Energy

* Conversion tables between national discipline codes and the KEINS discipline codes listed here can be found in the file: **DISCIPLINE_CONVERSION_TABLES.xls**

APPENDIX 3 - MATCH-MASK© FOR E-MAIL AND PHONE INTERVIEWS

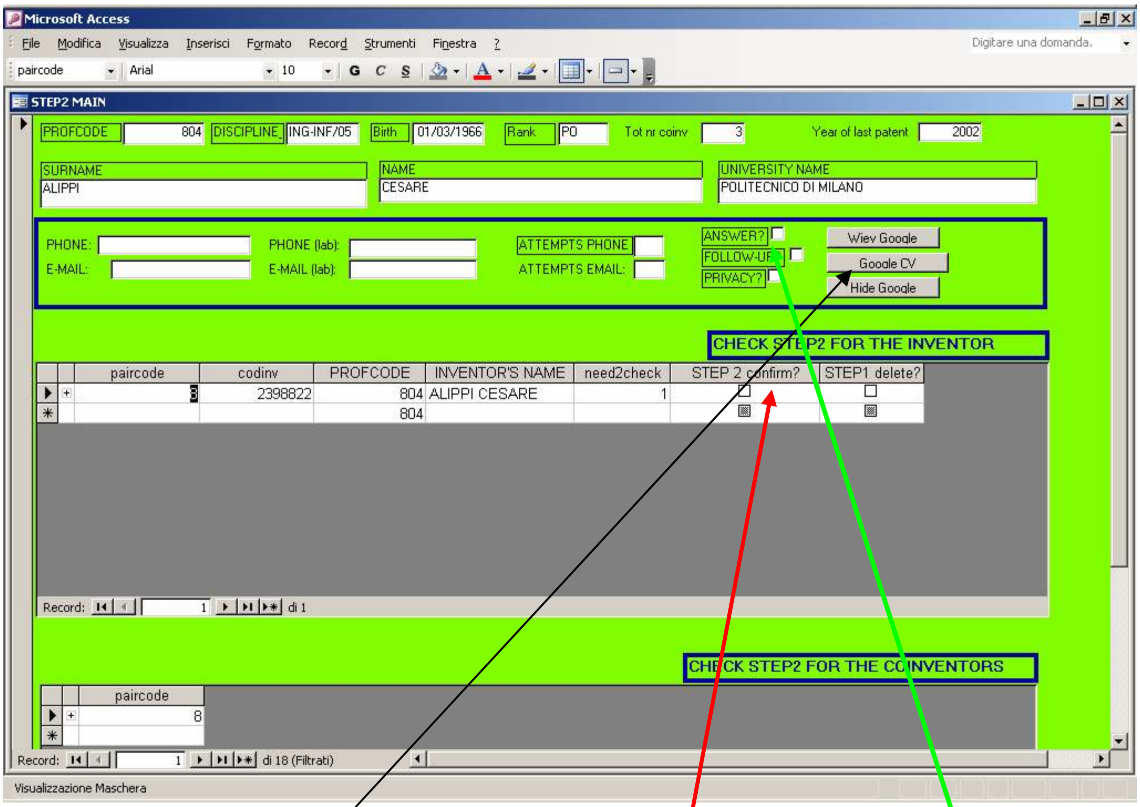
In this section we reproduce the instruction set provided to research assistants in charge with the task of contacting professor-inventor matches to ask for confirmation of their status as academic inventors.

* * *

MATCH-MASK©: HOW TO USE IT

YOUR TASK: you work by professor; to each professor many inventors may be assigned; you must find out (by checking the prof's CV, calling him or e-mailing him) whether the professor correspond to one or all of these professor

- Open the file and go to the "Mask" section
- Open the mask named "2 – STEP2 MAIN". The areas with a blue border are those upon which you must act



PROFCODE 804 DISCIPLINE ING-INF/05 Birth 01/03/1966 Rank PO Tot nr coinv 3 Year of last patent 2002

SURNAME ALIPPI

NAME CESARE

UNIVERSITY NAME POLITECNICO DI MILANO

PHONE:

E-MAIL:

PHONE (lab):

E-MAIL (lab):

ATTEMPTS PHONE

ATTEMPTS EMAIL:

ANSWER? □

FOLLOW-UP? □

PRIVACY? □

CHECK STEP2 FOR THE INVENTOR

	paircode	codinv	PROFCODE	INVENTOR'S NAME	need2check	STEP 2 confirm?	STEP1 delete?
+	8	2398822	804	ALIPPI CESARE	1	□	□
*			804			▨	▨

Record: 1 di 1

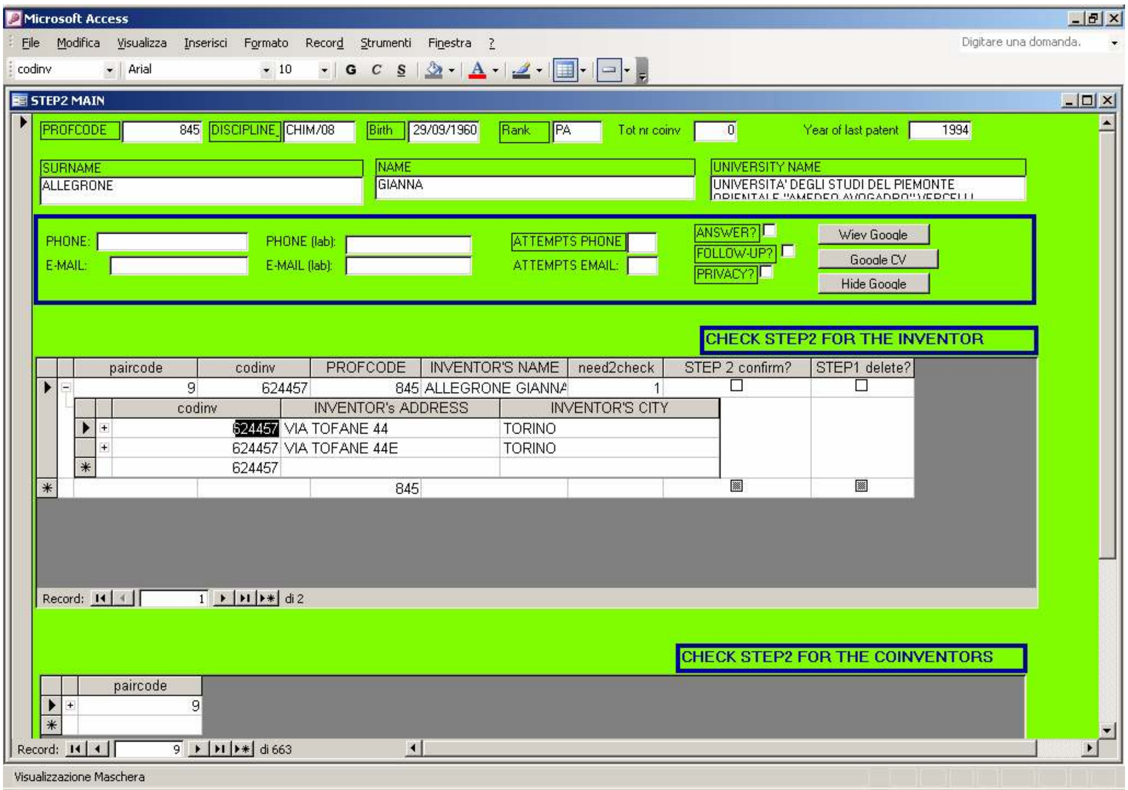
CHECK STEP2 FOR THE COINVENTORS

	paircode
+	8
*	

Record: 1 di 18 (Filtrati)

Click the "Google CV" key to look over the internet for the professor's CV: a web browser will appear, listing all websites that mention the professor. If one of these websites report the professor's CV check whether it mentions his/her patenting activity → if yes, click on the ⊞ key in the "CHECK STEP2 FOR THE INVENTOR MASK" (near "paircode") → a submask listing the inventor's address (more than, if he has changed home) and all of his patents will open → check whether in his CV the professor mentions any of these patents. If yes click on the "STEP2 confirm?" flag and click also the "Answer?" flag.

Here it is how the mask looks like when you click the + key near "paircode": you can now see the inventor's address.



PROFCODE 845 DISCIPLINE CHIM/08 Birth 29/09/1960 Rank PA Tot nr coinv 0 Year of last patent 1994

SURNAME ALLEGRONE

NAME GIANNA

UNIVERSITY NAME UNIVERSITA' DEGLI STUDI DEL PIEMONTE ORIENTALE "AMEDEO AVOGADRO" VERCELLI

PHONE: PHONE (lab): ATTEMPTS PHONE ANSWER? □

E-MAIL: E-MAIL (lab): ATTEMPTS EMAIL: FOLLOW-UP? □ PRIVACY? □

CHECK STEP2 FOR THE INVENTOR

paircode	codinv	PROFCODE	INVENTOR'S NAME	need2check	STEP 2 confirm?	STEP1 delete?
9	624457	845	ALLEGRONE GIANNA	1	□	□
		845				

codinv	INVENTOR's ADDRESS	INVENTOR'S CITY
624457	VIA TOFANE 44	TORINO
624457	VIA TOFANE 44E	TORINO
624457		

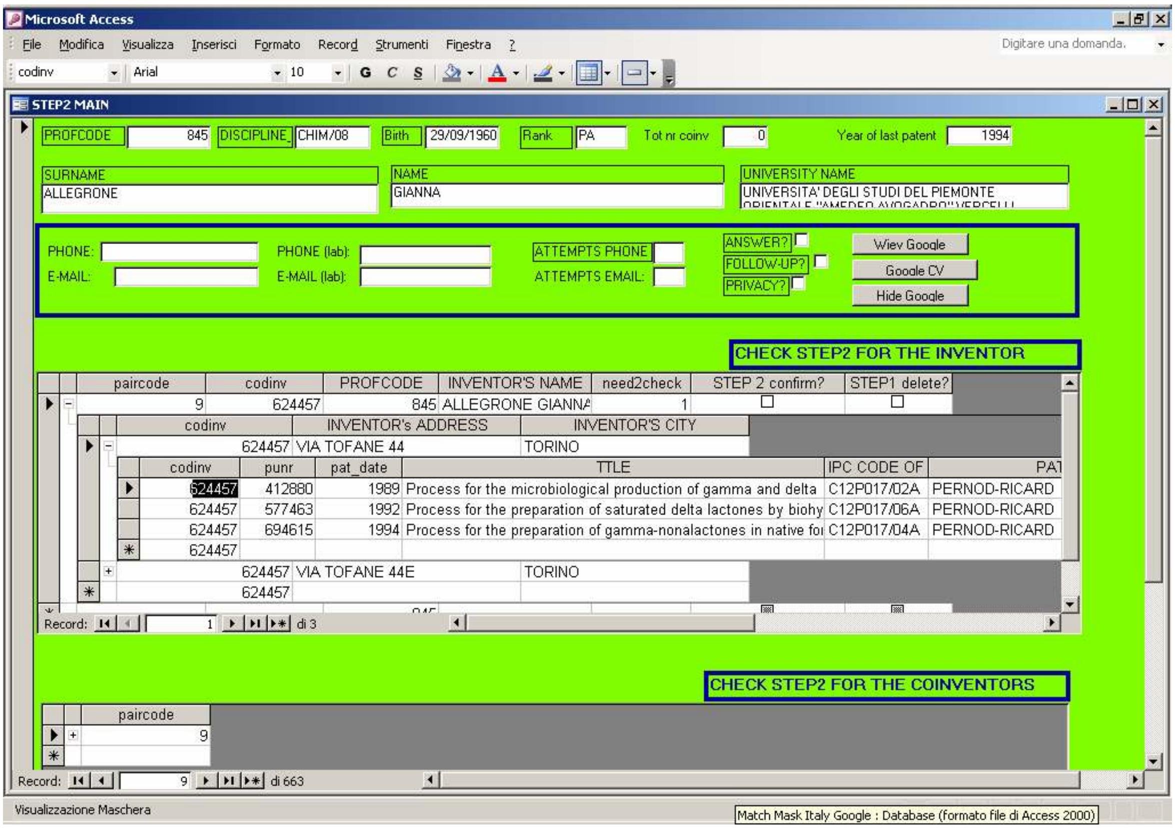
Record: 1 di 2

CHECK STEP2 FOR THE COINVENTORS

paircode
9

Record: 9 di 663

One more click the + key that appears on the left of each inventor's name and you get info on that inventor's patents (to go back to the previous step click the - key which now appears).


Microsoft Access
File Modifica Visualizza Inserisci Formato Record Strumenti Finestra ?
Digitare una domanda.
codinv Arial 10 G C S
STEP2 MAIN
Wiev Google
Google CV
Hide Google
Visualizzazione Maschera
Match Mask Italy Google : Database (formato file di Access 2000)

PROFCODE 845 DISCIPLINE CHIM/08 Birth 29/09/1960 Rank PA Tot nr coinv 0 Year of last patent 1994

SURNAME ALLEGRONE

NAME GIANNA

UNIVERSITY NAME UNIVERSITA' DEGLI STUDI DEL PIEMONTE ORIENTALE "AMEDEO AVOGADRO" VERCELLI

PHONE: PHONE (lab): ATTEMPTS PHONE ANSWER? □

E-MAIL: E-MAIL (lab): ATTEMPTS EMAIL: FOLLOW-UP? □ PRIVACY? □

CHECK STEP2 FOR THE INVENTOR

paircode	codinv	PROFCODE	INVENTOR'S NAME	need2check	STEP 2 confirm?	STEP1 delete?
9	624457	845	ALLEGRONE GIANNA	1	□	□

codinv	INVENTOR's ADDRESS	INVENTOR'S CITY
624457	VIA TOFANE 44	TORINO
624457	VIA TOFANE 44E	TORINO
624457		

codinv	punr	pat_date	TTLE	IPC CODE OF	PAT
624457	412880	1989	Process for the microbiological production of gamma and delta	C12P017/02A	PERNOD-RICARD
624457	577463	1992	Process for the preparation of saturated delta lactones by biohy	C12P017/06A	PERNOD-RICARD
624457	694615	1994	Process for the preparation of gamma-nonalactones in native fo	C12P017/04A	PERNOD-RICARD
624457					

Record: 1 di 3

CHECK STEP2 FOR THE COINVENTORS

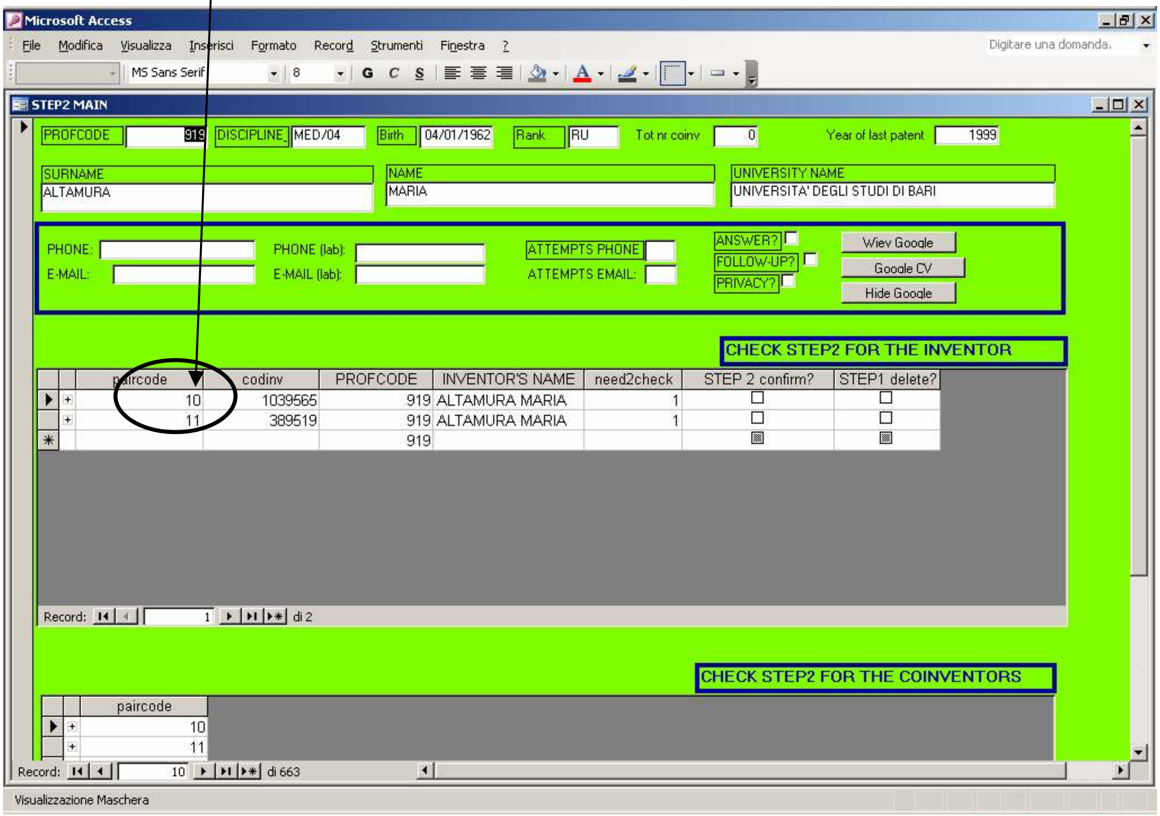
paircode
9

Record: 9 di 663

NB1: When you move to a new professor, you may find out that the "STEP2 confirm?" flag of one or more of his corresponding inventors has already been clicked. This may occur when this professor was a co-inventor of another professor, and the latter has confirmed in his place (see below "check step2 for coinventors"). In this case, don't click it again, otherwise you will cancel the info

NB2: When you move to a new professor, you may find out that the "STEP1 delete?" flag of one or more of his corresponding inventors has already been clicked. It is the result of STEP1 one-to-many procedure. If this is the case skip the inventor, we already know he cannot match the professor

More than one inventors (and as many + keys) may appear, if the professor matches more than one inventor (more than one CODINV). In the example below you find how the masks looks like when you find two inventors corresponding to the same professor,.



PROFCODE 919 DISCIPLINE MED/04 Birth 04/01/1962 Rank RU Tot nr coinv 0 Year of last patent 1999

SURNAME ALTAMURA

NAME MARIA

UNIVERSITY NAME UNIVERSITA' DEGLI STUDI DI BARI

PHONE:

E-MAIL:

PHONE (lab):

E-MAIL (lab):

ATTEMPTS PHONE

ATTEMPTS EMAIL:

ANSWER? □

FOLLOW-UP? □

PRIVACY? □

paircode	codinv	PROFCODE	INVENTOR'S NAME	need2check	STEP 2 confirm?	STEP1 delete?
10	1039565	919	ALTAMURA MARIA	1	□	□
11	389519	919	ALTAMURA MARIA	1	□	□
		919			▨	▨

Record: 1 di 2

paircode
10
11

Record: 10 di 663

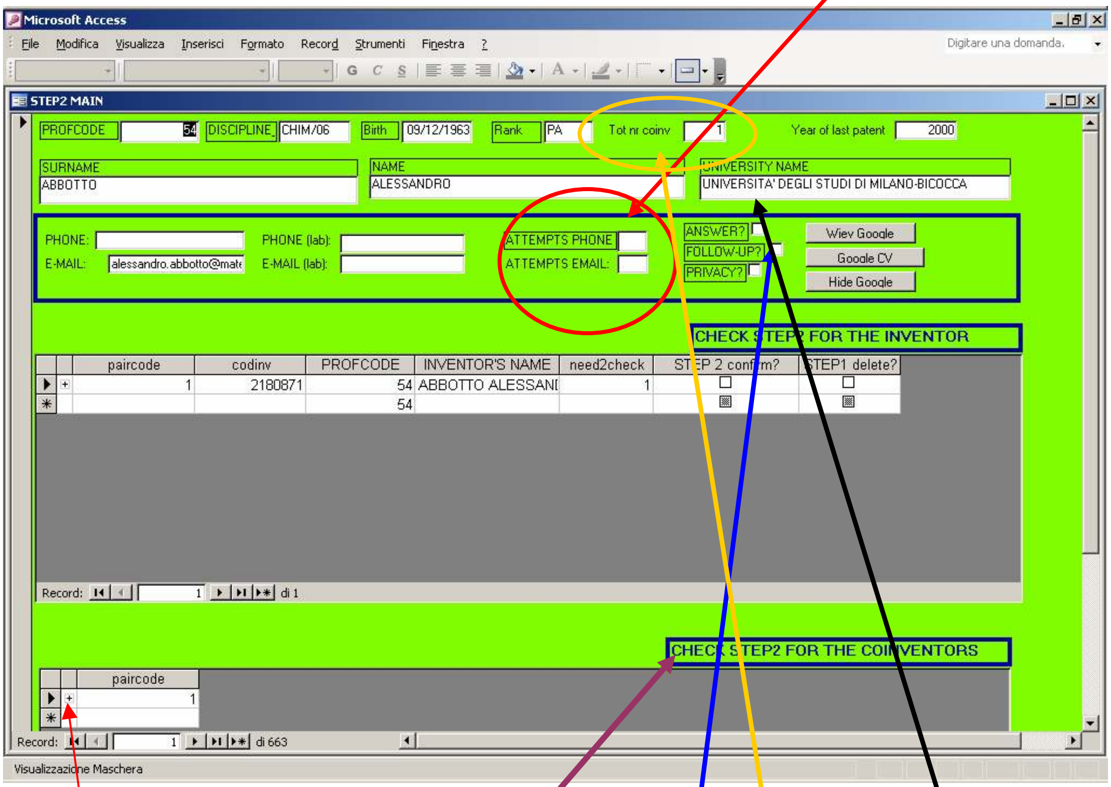
While checking professors' website on Google, check also whether you can get their e-mail and phone number from their universities' websites (or their CV)

→ it yes, copy-and-paste them in the dedicated fields. If you do not find these information for the professor, but more generically for his/her laboratories, proceed similarly, using the dedicated field.

→ if not, click the "View Google" key, which will try again to look for these information, using different criteria. If the search is successful proceed as described above.

*   *   *

After collecting phone numbers and e-mail addresses for all the professors, contact them over the phone or e-mail. Every time you call or e-mail the professor update the “attempt by phone” and “attempts by e-mail” boxes: record the number of phone calls you made or e-mails you set; decide a number over which you stop calling, or postpone calls to another week.


Microsoft Access
File Modifica Visualizza Inserisci Formato Record Strumenti Finestra ?
Digitare una domanda.
STEP2 MAIN
Wiev Google
Google CV
Hide Google
CHECK STEP2 FOR THE INVENTOR
CHECK STEP2 FOR THE COINVENTORS
Visualizzazione Maschera

PROFCODE 54 DISCIPLINE CHIM/06 Birth 09/12/1963 Rank PA Tot nr coinv 1 Year of last patent 2000

SURNAME ABBOTTO

NAME ALESSANDRO

UNIVERSITY NAME UNIVERSITA' DEGLI STUDI DI MILANO-BICOCCA

PHONE:

E-MAIL: alessandro.abbotto@mate

PHONE (lab):

E-MAIL (lab):

ATTEMPTS PHONE

ATTEMPTS EMAIL:

ANSWER? □

FOLLOW-UP? □

PRIVACY? □

		paircode	codinv	PROFCODE	INVENTOR'S NAME	need2check	STEP 2 confirm?	STEP1 delete?
▶	+	1	2180871	54	ABBOTTO ALESSANI	1	□	□
✳				54			▨	▨

Record: 1 di 1

		paircode
▶	+	1
✳		

Record: 1 di 663

PHONE CALLS

Ask the professor to confirm whether he is the author of the patents listed in the “CHECK STEP2 FOR THE INVENTOR MASK” and mark the STEP 2 CONFIRM box accordingly, as explained above.

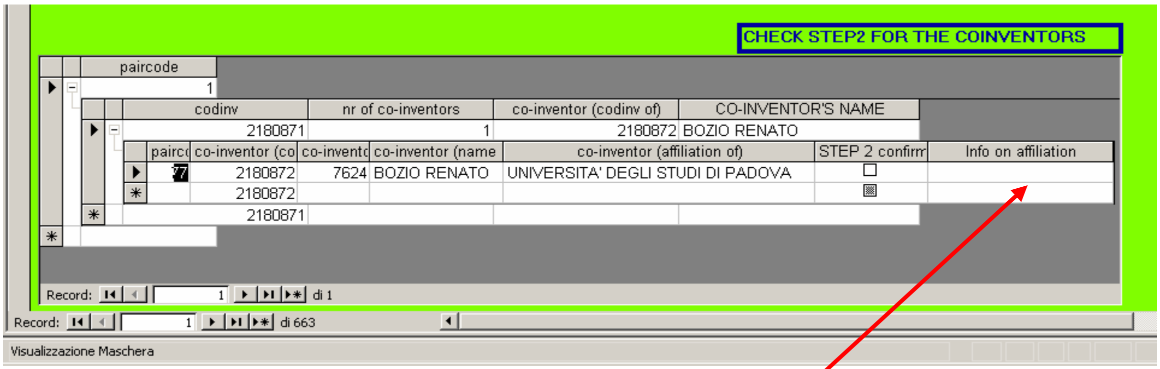
If the professor confirms to be the inventor, ask him whether in the future he would not mind answering a couple more questions on the origin of these patents, in terms of research and financing (do not add much more than this). If they say yes, click the “follow-up” flag.

If the professor gets angry and asks to be deleted from the dataset click the “privacy” flag

If the professor has one or more coinventors (see box “total nr. of inventors”) go to the “check step2 for the coinventors” submask.

At first the “check step2 for the coinventors” submask looks identical to the previous one, but if you click on the + keys you do not get info on the inventor and his patents, but on the inventor’s co-inventors, in particular on those co-inventors we suspect to be professors

Here it is how the submask for co-.inventors looks like.



CHECK STEP2 FOR THE COINVENTORS

paircode
1

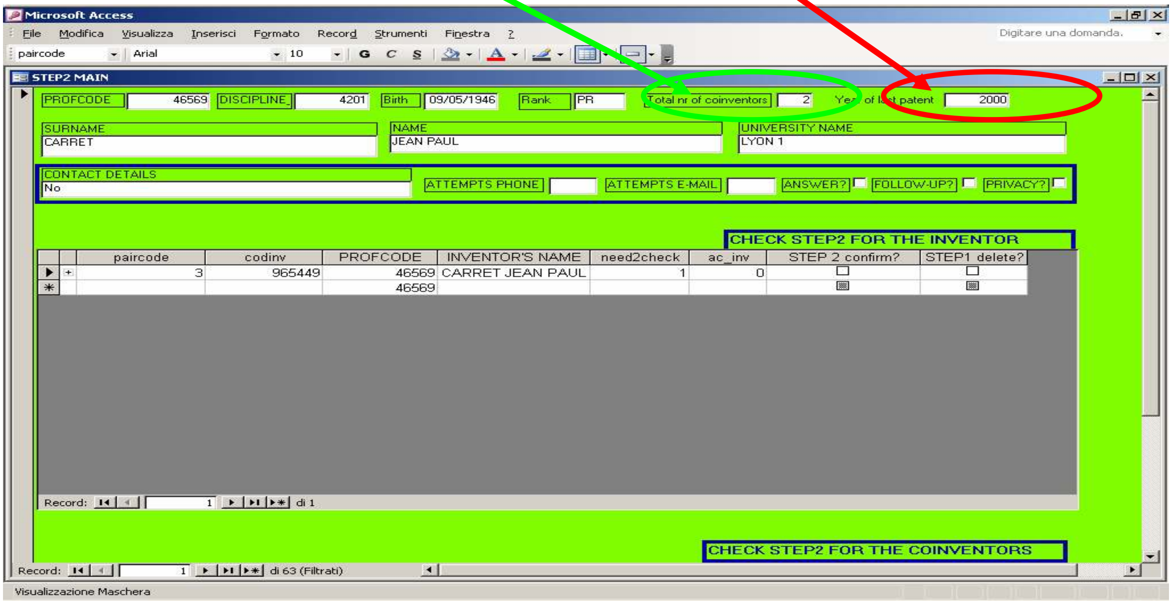
codinv	nr of co-inventors	co-inventor (codinv of)	CO-INVENTOR'S NAME
2180871	1	2180872	BOZIO RENATO
2180871			

pairc	co-inventor (co	co-invent	co-inventor (name	co-inventor (affiliation of)	STEP 2 confirm	Info on affiliation
77	2180872	7624	BOZIO RENATO	UNIVERSITA' DEGLI STUDI DI PADOVA	□	
	2180872				▨	

- Ask the professor whether he knows these co-inventors and can confirm they also are professors. If yes, click on the STEP2 CONFIRM flag of the co-inventor
- When you'll meet again the co-inventors in the next records the STEP2 CONFIRM flag will appear already clicked, and you will not need to contact these professors again.
- If you are told the co-inventor's affiliation is not the same as the affiliation at the time of the patent fill the "Info on affiliation" field with the name of the latter or other relevant info

Two useful features of the mask are:

1. The "Year of last patent" field, which tell us the year in which this professor is thought to have patented for the last time. You can use it to filter out professor-inventors matches whose patents are so old that you may fear to find it hard to get a confirmation. We used it to filter out, in the French case, people with patents applied for before 1994: since French matches were very may, we chose to concentrate on "young" ones, in order to get as many answers as possible.
2. The "Total nr. of co-inventor" fields: it indicates the nr of co-inventors from all patents assigned to the professor-inventor match. You may wish to start contacting people with many co-inventors: if they answer you, you then won't have to contact the co-inventors, thus saving time.



Microsoft Access
File Modifica Visualizza Inserisci Formato Record Strumenti Finestra ?
Digitare una domanda.
paircode
Arial
10
G C S

STEP2 MAIN

PROFCODE	46569	DISCIPLINE	4201	Birth	09/05/1946	Rank	PR	Total nr of coinventors	2
Year of last patent	2000								

SURNAME: CARRET
NAME: JEAN PAUL
UNIVERSITY NAME: LYON 1

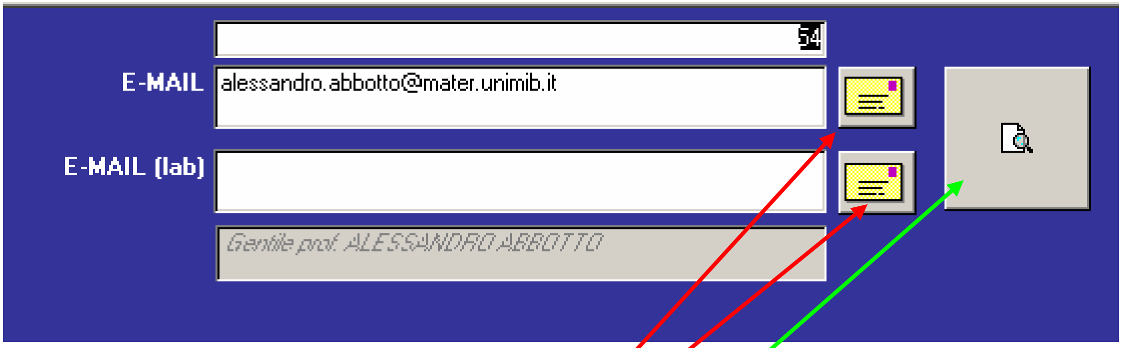
CONTACT DETAILS: No
ATTEMPTS PHONE
ATTEMPTS E-MAIL
ANSWER? □ FOLLOW-UP? □ PRIVACY? □

CHECK STEP2 FOR THE INVENTOR

paircode	codinv	PROFCODE	INVENTOR'S NAME	need2check	ac_inv	STEP 2 confirm?	STEP1 delete?
3	965449	46569	CARRET JEAN PAUL	1	0	□	□
		46569				▨	▨

E-MAILS

Close the “2 - STEP2 MAIN” mask and open the “E-MAIL” mask., which looks like:



54

E-MAIL: alessandro.abbotto@mater.unimib.it

E-MAIL (lab):

Gentile prof. ALESSANDRO ABBOTTO

NB: THE MASK RETRIEVES ONLY THE RECORDS FOR WHICH YOU FILLED IN EITHER THE “E-MAIL” OR THE “E-MAIL (lab)” FIELDS

When clicking one of the two “MAIL” keys your default e-mail programme will open up, ready to send a prepared message and an *.rtf attachment, to the professor identified by the code on the first row, either at his/her personal e-mail address, or at his/her lab’s. The message explains to its recipient why we contact him/her and asks him/her to open the attachment. The latter contains the list of patents and co-inventor attributed to the recipient, who is asked to tick those he confirmed to be his/hers. One more click on the “Send” key of your e-mail programme and the e-mail will be sent.

The PREVIEW key allows to view the attachment before sending the e-mail, and check the patents and co-inventors listed are OK: We suggest to use a few times to check the programme works fine; if the check is positive just send the e-mails.

APPENDIX 4. ACCESS RULES TO THE KEINS DATABASE ON ACADEMIC INVENTORS

(addendum to the scientific report for KEINS WP 5, November 2005)

The joint efforts of CESPRI, BETA and IMIT-CHALMERS will produce a set of 7 (SEVEN) related datasets.

One of them will contain the information due to the European Commission (Deliverable 11) and will be referred to as the "KEINS database" in this document and all the scientific and administrative reports due to the Commission after the Lisbon workshop held on October 14-15 2005.

The other datasets will be made accessible to the contributors according to the rules outlines below, only for research purposes; when citing them, the ensuing publications will also refer broadly to the "KEINS database" and list Deliverable 27 (or equivalent publication on a refereed journal, as indicated by CESPRI in due time) in the references.

The following list provide details on contents (names of included variables) and dissemination rules for all 7 (seven) datasets. If not explained her, the meaning of each variable can be found in the scientific report for KEINS WP 5.

CONTENTS OF DATASETS

DATASET 1. KEINS DATABASE - DELIVERABLE 11

Variables: personal code, age, gender, discipline, academic status, affiliation, nationality of academic inventors, publication number (or equivalent[23]) of each patent, priority date of each patent, IPC code of each patent, applicant of each patent[24].

Countries covered: France, Italy, Sweden.

NB: This is the only dataset that will be eventually delivered to the Commission. Personal codes will be created in such a way to protect the real identity (name, surname, and affiliation) of the academic inventors, for privacy reasons

DATASET 2. ITALIAN INVENTORS

Variables: name, surname, address(es), personal code, publication number of each patent, priority date of each patent, IPC code of each patent, type of applicant of each patent, name of applicant(s) of each patent, for all inventors with an Italian address listed on EPO patent applications, 1978-2003, from the EP-CESPRI database

NB: This dataset is not a deliverable to the Commission

DATASET 3. ITALIAN PROFESSORS

Variables: personal code, age, gender, discipline, academic status, affiliation (university name) of all Italian professors, as from the list compiled by CESPRI for the KEINS project.
NB: This dataset is not a deliverable to the Commission

DATASETS 4. AND 5.

Same as 2. and 3. for France
NB: These datasets are not a deliverable to the Commission

DATASETS 6. AND 7.

[23] The original EPO publication numbers (PUNR) allow for the personal identification of academic inventors, since they can be used to retrieve the original patent document on a number of search engines. Due to concerns regarding privacy laws emerged during the research, CESPRI will retain the option to substitute the original PUNRs with equivalent codes (one PUNR = one code)

[24] Patent applicants will be classified into 3 categories: business companies, open science institutions (i.e. universities and public laboratories), and individuals (as when the inventor's and the applicant's names coincide)

Same as 2. and 3. for Sweden
NB: These datasets are not a deliverable to the Commission

ACCESS TO DATASETS

DATASET 1.

Before completion of Keins project
Access will be granted to Keins WP 5 partners as soon as Deliverables 27 and 28 will be drafted, that is immediately after the Keins workshop due after month 23; or before then, if Keins WP 5 partners will contribute to Deliverables 27 and 28, or one of Deliverables 19-21 based upon the dataset. In addition to the contents of the dataset listed above, and only for the purposes of Keins research, Keins partners will receive names/surnames of inventors and publication number of patents. No access granted to researchers outside Keins WP 5, nor to the Commission.

After the completion of Keins project
Keins WP 5 partners: as before

Researchers outside Keins WP 5: free access upon request to CESPRI or the Commission (see below)

The Commission will dispose freely of the dataset, but it will require users to refer to it in their publications as the “KEINS database”, and to cite Deliverable 27 (or this paper, or equivalent publication on a refereed journal, as indicated by CESPRI in due time) in the references.

DATASET 2.

Access reserved to Cespri, both before and after the completion of Keins project; but dissemination to Keins WP 5 partners only for joint use in cross-country co-authored papers will be allowed.

Researchers who are not KEINS partners, but are interested in the data are encouraged to contact CESPRI

DATASET 3.

Access reserved to Cespri, both before and after the completion of Keins project; but dissemination to Keins WP 5 partners only for joint use in cross-country co-authored papers will be allowed.

Researchers who are not KEINS partners, but are interested in the data are encouraged to contact CESPRI

DATASET 4.

Access reserved to Cespri and Beta, both before and after the completion of Keins project; but dissemination to Keins WP 5 partners only for joint use in cross-country co-authored papers will be allowed.

Researchers who are not KEINS partners, but are interested in the data are encouraged to contact CESPRI

DATASET 5.

Access reserved to Beta, both before and after the completion of Keins project; but dissemination to Keins WP 5 partners only for joint use in cross-country co-authored papers will be allowed.

DATASET 6.

Access reserved to Cespri and Imit-Chalmers/Dept. Sociology, Umeå University, both before and after the completion of Keins project; but dissemination to Keins WP 5 partners for joint use in cross-country co-authored papers will be encouraged.

Researchers who are not KEINS partners, but are interested in the data are encouraged to contact CESPRI

DATASET 7.

Access reserved to Imit-Chalmers/Dept. Sociology, Umeå University, both before and after the completion of Keins project; but dissemination to Keins WP 5 partners for joint use in cross-country co-authored papers will be encouraged.