



# **Internationally comparative dataset on start-up processes and their institutional foundations in Germany, Italy, the UK and the US**

Andrea M. Herrmann  
Björn Fischer  
Lukas Held

## **Document Identifier**

Annex 4 to D5.1 Internationally comparative dataset on start-up processes and their institutional foundations in Germany, Italy, the UK and the US

## **Version**

1.0

## **Date Due**

M36

## **Submission date**

20 May 2018

## **Work Package**

5

## **Lead Beneficiary**

UU



Grant Agreement Number 649378



## Change log

Version	Date	Amended by	Changes
1.0	20 May 2018		-
1.1			

## 1. Executive summary

We here describe the empirical approach taken to collect the “perfect timing (PT) database”, which traces the timing of labour-, finance-, and knowhow-related activities throughout the venture creation process on a monthly basis. The sample for this study was drawn from the Orbis database, which provides internationally comparable company profiles. A rigorous catalogue of selection criteria was developed and applied to arrive at a meaningful sample. To collect data in Germany, the US, the UK, and Italy, national call centers were contracted after a thorough tender process. While these call centers recruited venture founders for an interview, the actual interviews were conducted by an interviewer team of research assistants, based at Utrecht University (the Netherlands), Columbia University (NY, USA), the Universität zu Köln (Germany), and the Heinrich-Heine Universität Düsseldorf (Germany). These interviewers were selected and rigorously trained by the survey coordinators. These data collection efforts enabled the recruitment of 820 interviews and completion of 539 interviews within the FIRES project. Together with the already existing data of the preceding Marie Curie project, the PT dataset includes 1044 cases.

## 2. Internationally comparative dataset on start-up processes and their institutional foundations in Germany, Italy, the UK and the US

### 1. Introduction

The seminal paper by Gartner (1988) led to a paradigm shift in entrepreneurship research. In line with his paper's title, Gartner argued that asking “(w)ho is an entrepreneur (...) is the wrong question” as entrepreneurship research ought to be process- rather than trait-oriented. While the trait-oriented studies of the 1970s and 1980s succeeded in identifying the core characteristics of entrepreneurs, they did not provide insights into the process to be undertaken when setting up a new venture. Such insights are, however, essential in order to learn about the steps needed for venture creation.

Gartner's claim was taken-up by many, thereby initiating the process-oriented school of entrepreneurship research. The need for data on venture creation processes led to the collection of several datasets, of which the Panel Study of Entrepreneurial Dynamics (PSED) is to date the most comprehensive one (Reynolds and Curtin 2007). Around the turn of the millennium, the PSED contributors interviewed - in two waves - ca. 1000 founders of nascent ventures about the steps undertaken during the venture creation process (Reynolds and Curtin 2008). While the PSED study was replicated, often in modified versions, in Argentina, Australia, Canada, Greece, The Netherlands, Norway, Sweden, and the UK, (idem: 167-168), the PSED focuses on identifying reference dates of many activities, such as the moment of corporate inclusion in the yellow pages. Start and end dates of particularly important venture creation activities, e.g. of R&D collaborations, are often missing. Precise time-stamped data is however needed in order to understand how the duration of activities, their timing within the overall venture creation process, as well as their completion in relation to other start-up activities shapes the venture creation process and its outcome.

To uncover patterns of venture creation processes, as well as their (institutional) drivers, the 'Perfect Timing' (PT) dataset was collected between 2011 and 2018. Led by Andrea M. Herrmann, research teams at Utrecht University (the Netherlands), Columbia University (NY, USA), the Universität zu Köln (Germany), and the Heinrich-Heine-Universität Düsseldorf (Germany) collected information on overall 1044 venture creation processes in collaboration with Saul Estrin (London School of Economics) and Luca Grilli (Politecnico di Milano).<sup>1</sup>

---

<sup>1</sup> Between 2011 and 2013, these data collection efforts received funding from a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme, from the QMSS program at Columbia University (New York, USA), and from the Innovation Studies Group at Utrecht University (The Netherlands). Between 2015 and 2018, data collection was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 649378.

These data collection efforts translated into an internationally comparative dataset on start-up processes of alternative energy, as well as information & communication technology ventures, in Germany, Italy, the UK, and the USA.<sup>2</sup> For these countries, the PT database provides monthly information on the activities undertaken to build-up human resources, acquire finance, and develop the core business idea of the new venture. While the PT dataset contains more time-stamped information than PSED, it explicitly includes variables that enable merging PT and PSED data. The PT data thus constitutes an important resource for advancing scholarly understanding of venture creation processes.

The following summarizes the empirical approach taken to collect the PT database. To this end, it illustrates the sampling approach taken, project administration, and the rationale underlying the questionnaire. It furthermore discusses the interview schedule, data cleaning, and data reliability.

## 2. Conceptualization and Data Collection Activities

### 2.1. Sampling Approach and Sample Characteristics

The question of how to arrive at a meaningful sample is intimately linked to the question of how to conceptualise entrepreneurship. As (Reynolds and Curtin 2007: 3) note, the literature uses a variety of concepts and indicators, including “self-employment as a proxy for entrepreneurship, (...) new market entrants, (...) new listings in registries of business organizations, (...) the emergence of new industries or types of organizations (...), retrospective histories [of particularly successful ventures], and a wide range of samples of convenience.”

Combining different aspects of these conceptual and empirical approaches, we define entrepreneurship in this study as *economic activity by one or more individuals that translates into the registration of a new, independent for-profit firm*. Importantly, this definition excludes self-employed individuals that are not incorporated, as well as firms that are registered as a subsidiary or otherwise financially dependent entity of a larger company.

Based on this concept, we used a venture’s *registration year, legal form, country, and industry* as sampling criteria. To obtain an internationally comparative sample, we used the ORBIS database of Bureau van Dijk, which offers ample corporate information including the *registration date of firms* at a chamber of industry and commerce or a comparable authority (see <https://www.bvdinfo.com/en-gb/our-products/data/international/orbis>). Given that data collection started in 2011 and ended in 2018, we decided to include all years of corporate

---

<sup>2</sup> In addition, the dataset contains information on about 50 venture creation processes in the Netherlands. For the Netherlands, data coverage is thus more limited and less systematic.

registration between 2004 and 2014.<sup>3</sup> These registration dates leave sufficient time to trace venture development after registration. Most importantly, this sampling approach has the advantage of including ventures that became successful after registration by making sustainable profits, as well as ventures that failed and were thus dissolved. Furthermore, these registration dates offer the advantage of being sufficiently close to the interview date, so that founders still remembered the venture creation activities undertaken. This is particularly important as numerous venture creation activities already took place before the venture's registration at a chamber of industry and commerce or a comparable authority.<sup>4</sup>

We used the *legal form* under which a venture was registered to exclude public corporations, not-for-profit organisations, foundations, associations, cooperatives, as well as liberal professions and solo self-employed.

Akin to the Global Entrepreneurship Monitor (GEM), also the PT study considers the *country* as an essential unit of analysis (see Reynolds, Bosma et al. 2005: 208), because differences in institutional influences on venture creation processes are likely to be particularly pronounced between national institutions (Hall and Soskice 2001). Accordingly, the FIRES project in general, and the PT survey in particular, decided to study venture creation processes in three European countries with particularly distinct and representative institutions of Continental European, Anglo-Saxon and Mediterranean economics: namely Germany, the UK, and Italy (see Hall and Soskice 2001; Amable 2003; Hancké, Rhodes et al. 2007; Schneider and Paunescu 2012). In addition to these countries, the PT study also covered the US as the latter is said to offer the most favourable institutional environment for entrepreneurship.

In addition to these four countries, the PT sample also focuses on specific sectors and *industries*, namely the alternative energy sector (including solar, wind, and biomass industries), as well as the information and communication (ICT) sector (including information and communication industries alike). While both are forward-looking, the alternative energy sector has been massively subsidized over the past decades, while the ICT sector has not received any sector-specific subsidies. The impact of subsidies on venture creation processes thus becomes particularly well visible in alternative energy ventures.

Both NACE (Rev.2) and ISIC (Rev.4) classifications include a specific industry code for ventures pursuing telecommunications and computer-related activities, so that ICT ventures were directly identifiable. Importantly, though, there are no separate industry codes for alternative energy ventures. Accordingly, alternative energy ventures are included in broader sectoral classifications, such as 'electric power generation, transmission and distribution' – to name just one of the seven, eight, and respectively nine industry codes we used for locating wind, biomass, and solar firms within the broader sample. In a next step, we then used the venture's trade description to manually single out the wind, biomass, and solar

---

<sup>3</sup> Ventures that were registered in 2006-2012 were approached first with the request for an interview, ventures registered in 2004, 2005, 2013, and 2014 were approached once the 2006-2012 samples were exhausted.

<sup>4</sup> For illustrations about data reliability, see section 2.6 below.

ventures from this broader sample.<sup>5</sup> Whenever in doubt, the venture's core activity was cross-checked via a www-search.

Given that 'Computer Assisted Telephone Interviews' (CATI) interviews could only be conducted with founders of ventures with a telephone number, the availability of a phone number *de facto* became an additional sampling criterion.

## 2.2. Project Administration

To secure data quality, reliability, and completeness, the data collection process was split in three parts: First, for each country surveyed, a call center was commissioned to identify and recruit interview partners (mostly founders) who were sufficiently knowledgeable about the start-up process of the ventures sampled. After a thorough tender process, that call center was selected which only worked with native speakers and was sufficiently experienced in founder surveys in the respective country. After a thorough training process, the callers contacted ventures and identified suitable interview partners to participate in the PT survey. In doing so, the callers cross-checked the aforementioned sampling criteria (most importantly, the venture's industry and independence from a mother company).

In a second step, the actual interviews with founders were completed. To this end, research assistants were recruited after a thorough selection process. All research assistants were students (or former students) of different business-, management- or innovation-related study programs at Utrecht University (the Netherlands), Columbia University (NY, USA), the Universität zu Köln (Germany), and the Heinrich-Heine-Universität Düsseldorf (Germany). Next to their overall performance, motivation and time commitment to the project, the interviewers' language skills were a major selection criterion as only native speakers or interviewers with equivalent language skills were hired. In preparation of their first interview, these research assistants were thoroughly trained by the project leader (A.M.Herrmann) and her assistants (most importantly Lukas Held), by listening to audio recordings of previous interviews, and by completing trial interviews amongst each other. Once trained, the research assistants recorded their interviews with founders (whenever permission was granted).

In a third step, the data collected was cross-checked by the survey coordinator: Upon completion of the interview, the research assistants forwarded the interview recordings to the survey coordinator (most importantly, B.Fischer and A.M.Herrmann). The survey coordinator either listened into these recordings or clicked through the online CATI questionnaire in order to clean the data in the questionnaire, and to provide feedback to the interviewers (for details, see below section 2.5 "Data Cleaning"). This constant feedback process did not only assure high data quality, completeness, and reliability; it also constituted an ongoing training process for the interviewers.

---

<sup>5</sup> For US ventures, trade descriptions were not available. We thus took the venture name (for example, for solar ventures, "sol\*", "lux", "green", "energy", "photo", "vol", "helio" – and many more) to identify the venture's industrial activity.

While the aforementioned process of project administration remained the same throughout the entire period of data collection (from 2011 – 2018), it should be noted that data collection took place in two waves: From 2011 – 2014, data collection of the PT database was initiated by A.M.Herrmann within the framework of a Marie Curie fellowship at Columbia University (New York). The questionnaire was developed on the basis of thorough literature studies, in-depth interviews with entrepreneurs and in close collaboration with experts and practitioners in the field. It was tested through interviews with founders of applicable ventures and repeatedly modified until a concise questionnaire had been developed. During the first wave, data collection focused on Germany, the USA, and the Netherlands. Consequently, native-speaking interviewers were not only recruited and based at Utrecht University (the Netherlands), but also at Columbia University (USA), as well as the Universität zu Köln (Germany) and the Heinrich-Heine-Universität Düsseldorf (Germany). These data collection efforts were financed by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme, by the QMSS program at Columbia University (New York, USA), and by the Innovation Studies Group at Utrecht University (The Netherlands).

In view of the scholarly interest that the PT database had triggered already in 2014, the H2020 FIRES consortium decided to finance additional data collection efforts. A second wave of data collection thus took place between 2015 and 2018, which aimed at broadening the existing US and German datasets, and at collecting new data for the UK and Italy. During this second wave, the questionnaire of the first interview wave was broadened by including questions that enable merging the PT and PSED datasets. Furthermore, questions about the socio-economic characteristics of founders, about investment amounts, writing a business plan and patenting activities were added. These questions were added after thorough literature studies and in consultation with various FIRES colleagues, most notably Saul Estrin (London School of Economics), Luca Grilli (Politecnico di Milano), as well as Niels Bosma and Mark Sanders (Utrecht University). In addition to conducting new interviews in Germany, Italy, the UK and the USA, the existing German and US interviews of the first wave were completed by re-calling the interview partners of that time. These follow-up interviews also offered a valuable opportunity to cross-check the data reliability of the PT study (see section 2.6).

### **2.3. Questionnaire Design**

To ensure that dozens of interviewers in different countries could assist in data collection, a structured interview guide was developed for a survey based on ‘Computer Assisted Telephone Interviews’ (CATI). This interview guide made it possible to systematically capture venture creation circumstances and to trace how venture creation processes evolved on a monthly basis. To this end, the questionnaire contains six parts.

Part I of the questionnaire records the venture details and circumstances of venture creation, such as the venture’s *industry, location, year of registration, legal form, business idea (product or service), novelty, degree of innovativeness, and location of core customers.*

Part II captures the length of the venture creation process by identifying its start and end date. In line with the process-oriented entrepreneurship literature (Reynolds and Curtin 2008), the PT uses different ways to determine when venture creation started and ended respectively. These indicators can be used alternatively. Possible start dates include:

- the moment when the *interview partner* first *thought* about setting-up the venture in question,
- the moment when *one of the founders* first *talked* about setting-up the venture in question,
- the moment of *corporate registration* at a chamber of commerce or a comparable register,
- if applicable, the moment when one of the founders started *writing a business plan*.

Possible dates to determine the end of venture creation include the respective moments

- when the new venture first generated *revenues, profits*, or respectively *sustainable profits* for more than 3 months, as well as the moments
- when the venture *merged, was acquired, or dissolved*.

Parts III, IV, and V inquire into the timing of activities that took place during venture creation. Contrary to PSED, the PT survey does not seek to cover a broad variety of activities (such as the moment of inclusion in the yellow pages or of opening a bank account) but rather captures detailed time-stamped information on those activities that can be considered essential for (the success of) any venture creation process. Economic theory, the resource dependence view (Pfeffer and Salancik 1978), and the varieties of capitalism literature (Hall and Soskice 2001) – to name just some particularly influential literature strands– teach us that a company cannot operate without labour (or human capital), finance (or financial capital), and know-how (needed for product development). Accordingly, part III asks about how the venture's labour composition evolved over time; part IV enquires into finance acquisition; and part V captures how the necessary know-how for product/service development was acquired.

More precisely, part III traces – on a monthly basis – how many *founders, employees, and service providers* worked for the venture on a part-time or full-time basis respectively. Furthermore, the socio-economic background of the founders is retrieved, namely their marital status, financially dependent children, highest degree obtained, prior occupation and venture creation experience, as well as their motives for setting up the new venture.

Part IV inquires into the different financial sources that the venture acquired, including:

- *shareholder capital* by founders, their family and friends, as well as corporate investors including venture capital firms and business angels.
- *loans* provided by banks and other corporate investors, as well as by different types of private investors.
- *subsidies, grants and other funding that the venture did not need to pay back*.

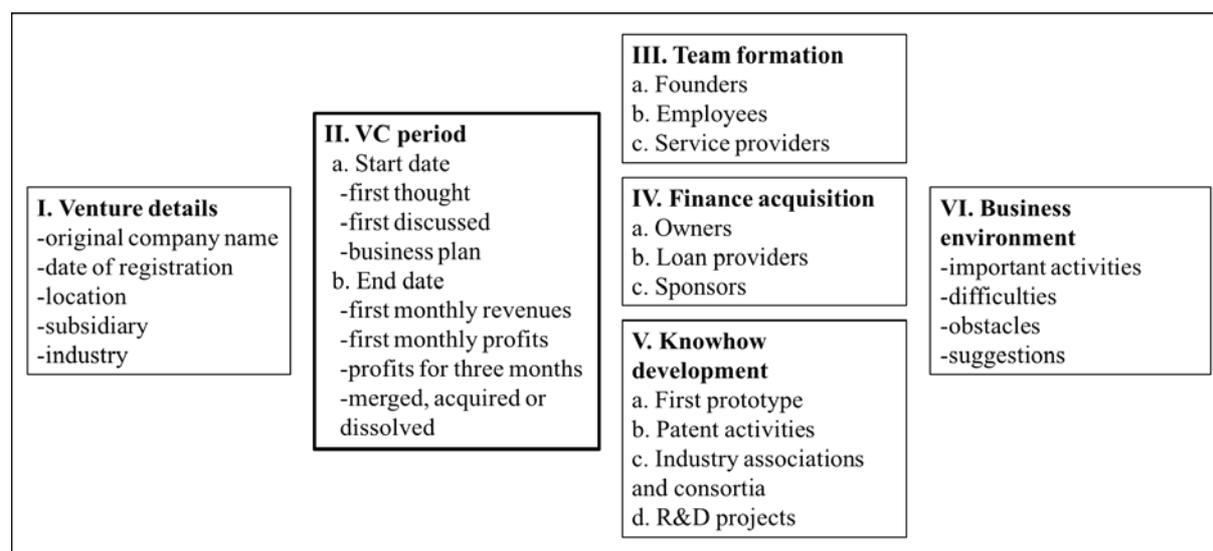
For any of these financial means, part IV asks about the beginning and end dates of funding acquisition, as well as the amount of finance that was invested.

Part V traces the process of product development, that is, the development of the first prototype of the venture’s core business idea. In addition to the timing of prototype development, part V captures whether the venture developed its core business idea on its own, in collaboration with academic or corporate R&D partners, or within the framework of a larger consortium or industry association. Furthermore, if applicable, the timing of patenting activities is recorded.

Importantly, the questionnaire also seeks to identify possible institutional influences on venture creation. Given that founders within one institutional context (i.e. country or region) typically take the latter for granted as they, simply, do not have a cross-institutional comparison, differences in national or regional institutions cannot be established by asking directly about them. Instead, they rather become visible indirectly through systematic differences in economic behaviour or judgements about the venture creation context. To capture such differences, the questionnaire also includes several open questions asking about whether and, if so why, founders were reluctant to give up dependent employment or to hire employees. Furthermore, the questionnaire asks about possible difficulties experienced in obtaining institutional investment and about the types of support obtained from institutional investors. Finally, part VI asks about those activities that founders experienced as particularly important and, respectively, difficult when setting-up their ventures. Furthermore, it enquires into whether founders experienced any regulatory obstacles during venture creation and asks what policy-makers could do to facilitate venture creation.

Figure 1 provides an overview of the six parts covered during in the PT survey as well as the major questions covered within each part.

**Figure 1: Structure of PT Timing Interview Schedule**



## 2.4. Interview Schedule

The master version of the CATI questionnaire was developed in English and then translated into all applicable languages, i.e. German and Italian. These translations were thoroughly cross-checked and reviewed by native speakers. For each language, at least three native speakers participated in the review. Their inputs were consolidated and implemented by the survey coordinator before data collection began.

The duration of each interview varied depending on the complexity of the venture creation process. On average, an interview lasted about 40 minutes. Interviews were conducted typically with the venture founder or, in about 5% of the cases, with another person who was particularly knowledgeable about the venture creation process, namely a family member (wife, sibling of the founder), employee, or core investor. Following the CATI procedure, the interviewer read the questions to the respondent as they appeared on the computer monitor and then entered the answers directly into the online survey. To ensure data **consistency**, the computer program automatically led the interviewer to the relevant follow-up questions.

Each interview started with the interviewer informing the respondent about his/her rights, stipulating the voluntary nature of the interview, the right not to answer, to stay anonymous, raise questions, interrupt - or withdraw from - the interview, as well as the required minimum age of 18 years to participate. The respondents were asked to provide their formal consent to these conditions and were given the opportunity to indicate whether, or not, they would allow for the interview to be recorded. In more than 95% of the cases, this permission was granted.

Once the interview partner was informed about his/her rights, the interviewer proceeded to the actual questionnaire. As mentioned in section 2.3, the main questionnaire contains six central parts recording: (I) venture details, (II) the venture creation period, (III) the team formation process, (IV) finance acquisition, (V) knowhow development and (VI) the founder's opinion about the business environment (see figure 1 for an overview of the interview schedule and its central elements).

During interviews, the most critical point of part I was to determine whether the recruited start-up conformed to the sampling criteria: Interviews were conducted only with founder of those ventures that were (a) founded as the original company and were no successor organization, (b) registered in either Germany, the UK, the USA or Italy, (c) were no subsidiary and (d) active either in the ICT or alternative energy sector. If these criteria were not fulfilled, the interviewer abandoned the interview. Furthermore, essential information about the venture's core business idea (product/service) was collected, such as its degree of novelty and envisioned customer spectrum.

Part I was followed by one central element of the questionnaire: the identification of the time span of venture creation (part II). For the CATI procedure, the start date of a venture creation process was defined as the moment (indicated by the interviewee) when one of the founders first discussed the idea with another person to set up the business in question. The

end date was determined as the moment when the venture made profits for three consecutive months, or when it was merged, acquired or dissolved. For those cases where a venture was neither profitable for three consecutive months nor merged, acquired or dissolved, venture creation was considered to be ongoing until the day of the interview.

The time span of venture creation identified in part II served as a reference period for all following sections of the interview schedule. In part III, the respondents were asked to indicate the involvement periods of the venture's founders, employees and service providers specifically during the identified venture creation time frame. Similarly, in part IV, the interviewer enquired into the investment periods of owners, loan providers and sponsors. Likewise, part V asked about the development process of the venture's core business idea (product/service), its patenting activities, R&D projects and participation in industry associations and consortia throughout the venture creation process. The restriction of data collection to a specific timeframe enabled the interviewers to focus on those events that were most relevant during the early start-up phase.

To investigate the influence of a venture's institutional environment on venture creation, the interview schedule was designed specifically to combine systemic quantitative data with relevant qualitative insights. While most of the questions thus produce fine-grained quantitative data about the timing and sequences of venture creation activities, the interview schedule also incorporates qualitative, open questions. Most notably, section VI poses four open questions to gain a more profound understanding of the business environment in which the new venture was set up.

All items included in the questionnaire have been carefully crafted and tested to provide a comprehensive picture of the events that occur during venture creation. To facilitate the interview procedure, the underlying rationale and focus of each question were thoroughly explained to the interviewers. This approach did not only create the necessary interviewer confidence to guide the respondents, it also was at the basis of collecting meaningful data. Overall, a total of 886 interviewees report complete data on all items of the questionnaire. An additional 158 interviews provide useful data on at least parts of the questionnaire, so that the entire dataset contains overall 1044 cases. Table 1 provides an overview, including separate information (by country and industry) of the interviews recruited and conducted during the first (Marie Curie) and the second (H2020) data collection wave.

*Table 2: Interviews Recruited and Completed (by Country and Industry)*

Overview by Country and Industry	Data Collection Marie Curie		FIRES Interviews Envisaged		FIRES Recruited	Data Collection FIRES		PT Dataset Overall (Marie Curie + FIRES)	
	All interviews conducted	Complete interviews conducted	Additional interviews envisaged	Overall number of interviews envisaged	Number of interviews recruited	All interviews conducted	Complete interviews conducted	Overall number of interviews available	Overall number of complete interviews available
<b>Germany</b>	<b>265</b>	<b>213</b>	<b>100</b>	<b>300</b>	<b>168</b>	<b>113</b>	<b>100</b>	<b>378</b>	<b>313</b>
- of which IT	169	137			n.a.	49	45	218	182
- of which AE	91	71			n.a.	62	55	153	126
<b>USA</b>	<b>187</b>	<b>163</b>	<b>100</b>	<b>300</b>	<b>172</b>	<b>101</b>	<b>72</b>	<b>288</b>	<b>235</b>
- of which IT	135	117			n.a.	54	41	189	158
- of which AE	48	40			61	45	31	93	71
<b>UK</b>	<b>0</b>	<b>0</b>	<b>300</b>	<b>300</b>	<b>301</b>	<b>181</b>	<b>158</b>	<b>181</b>	<b>158</b>
- of which IT					250	142	126	142	126
- of which AE					51	39	32	39	32
<b>Italy</b>	<b>0</b>	<b>0</b>	<b>300</b>	<b>300</b>	<b>179</b>	<b>144</b>	<b>133</b>	<b>144</b>	<b>133</b>
- of which IT					112	98	90	98	90
- of which AE					67	44	43	44	43
<b>NL</b>	<b>53</b>	<b>47</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>53</b>	<b>47</b>
- of which IT	26	21				0	0	26	21
- of which AE	27	26				0	0	27	26
<b>Σ</b>	<b>505</b>		<b>800</b>	<b>1200</b>	<b>820</b>	<b>539</b>	<b>463</b>	<b>1044</b>	<b>886</b>

## **2.5. Data Cleaning**

A comprehensive cleaning procedure was established to ensure that all data gathered is consistent. Each interview conducted was individually reviewed for quality and consistency by at least one survey coordinator, who carefully listened to the interview and, if necessary, cleaned the data in the CATI questionnaire in line with the responses of the interviewee. In particular, the survey coordinator cross-checked the chronology of events, the industry and product of the new venture, its degree of innovativeness, whether breaks between time periods were meaningful, whether all notes taken were understandable to outsiders, and whether all mentioned dates for team formation, investments and knowhow development were within the time frame of the venture creation process. To do this, the survey coordinators relied not only on the information provided by the respondent, but also on venture information that was available online.

Minor inconsistencies could be corrected directly by the survey coordinator. Such minor inconsistencies generally referred to situations in which a response was not indicated correctly, such as misconceived months or years, an incorrect chronological order of collaborators or investors, or a mistaken degree of innovativeness of the venture's products. In some cases, larger inconsistencies or missing data occurred that could not be corrected directly by the survey coordinator. Here, the survey coordinator asked the interviewer either for an explanation or, if no explanation could be given, to call the respondent again in order to clarify inconsistencies or collect missing data. This happened, for example, when the interviewer omitted information about a service provider or an institutional investor, who had been mentioned by the respondent. Most times, such additional information yielded useful data that made it possible to correct for such inconsistencies. In some rare cases, however, major inconsistencies could not be corrected (for instance when it was no longer possible to reach a respondent). In these cases, the corresponding parts of the interview were excluded from the dataset. If the entire interview had major mistakes that could not be corrected (for example a venture active in the wrong industry, or too vague responses by the interviewee), the interview was excluded from the dataset altogether.

For each interview of the FIRES wave, the survey coordinator created an individual feedback file in which s/he duly noted all inconsistencies and adjustments made. As such, the feedback file did not only help data users to keep track of the adjustments made, but also allowed for a continuous training of the interviewers, thereby ensuring that all interviews were completed homogeneously. At times, the interviews were conducted by particularly experienced interviewers who had received plenty of training. In these cases, the survey coordinators did not listen into the audio recordings but only clicked through the online questionnaire to look for inconsistencies. In doing so, the survey coordinator followed basically the same procedure as for the interviews that required listening. The only difference was that, for each interview, the survey coordinator kept detailed logs of the email conversations with the corresponding interviewer on top (or instead) of the feedback files. In this vein, clarifications by the interviewers for all types of inconsistencies were stored.

## 2.6. Data Reliability

Data reliability generally refers to the degree to which a measure of a concept is replicable and stable over time. Anybody following the same measurement procedure should be able to arrive at the same findings, irrespective of individual judgments made by the observers or researchers. While the measures and data collection procedures for the PT dataset have clearly been specified for repeated use, one essential point that could nevertheless have affected data reliability is that the data collected for each venture relies on the responses of one single interviewee. Even though the founders interviewed appeared not to have difficulties to recollect all major events that occurred throughout the venture creation process, the question remains: (How) can we be sure that the responses by this single respondent are accurate and reliable?

To create confidence in the reliability of the founders' responses, we administered two different reliability tests. First, we designed a procedure to assess the *reliability of responses over time* using *follow-up interviews*. For this procedure, 155 founders in Germany and the USA were interviewed twice: first in 2011-2013 and a second time in 2017-2018. During this second interview with the same founders, the same questions were asked. As the questions remained the same, the respective founders were asked to remember events that, in 2017-2018, dated back about five more years than in 2011-2013. Importantly, the interviewees did not receive any assistance or knowledge in relation to the responses they had given during their first interview. Finally, we calculated the overlap between the responses given during the first (2011-2013) and second (2017-2018) interview. Despite the 5-year time gap, responses overlapped for more than 70%. Importantly, the interviewees also reported that differences in the overlap were not to be attributed to a lack of the respondents' memory, but rather to the more limited experience of interviewees during the first wave of interviews.

A second procedure was developed to assess the *reliability across respondents* by contacting a *co-founder* of the same venture. Those interviewees who reported that more than one founder was involved in the set-up of the new venture were asked for the contact details of their co-founders. A few months after the interview with the main founder, 20 co-founders in Italy and the UK were called and asked exactly the same questions about venture creation as the main founder. More concretely, the co-founders were asked to verify the information provided by the main founder, and to recall if any significant information was missing. Subsequently, we again calculated the overlap between the answers by the main founder and the co-founder. We found that in these cases – where the interview with the co-founder took place only a few months after the interview with the main founder – the response overlap was almost 100%.

The results of the two abovementioned reliability tests indicate that the responses provided by the founders interviewed can be considered reliable. Duplicate measures as well as similar procedures designed to tap into the same concepts have produced remarkably uniform results. Those interested in studying early-stage entrepreneurial processes can therefore be assured that the PT dataset contains only reliable and consistent data.

### 3. Conclusions

To date, the PT database constitutes one of the most complete and reliable databases on venture creation processes. Focusing on activities related to team formation, finance acquisition and product/service development, the database contains precise time-stamped data on a monthly basis. While it is mergeable with other databases, in particular the PSED study, the PT database is unique with regard to the detailed time-stamped information it provides. Furthermore, it is directly internationally comparative as it traces venture creation processes in Germany, Italy, the USA, the UK, and the Netherlands.

It might be interesting to learn that the more limited number of cases for Italy and the UK were caused by the more limited samples available, in particular for alternative energy ventures. In the UK, entrepreneurs often shied away from setting-up alternative energy ventures because of frequent and, thus, unpredictable regulatory changes. In Italy, on the contrary, entrepreneurs repeatedly indicated to set-up alternative energy ventures, (thus obtained the related subsidies,) but ultimately abstained from executing their plans. Irrespective of any subsidy provisions, venture creation in the information and communication industry was generally more limited in Italy.

Despite the slightly more limited database for Italy and the UK, the PT dataset has been extremely well received by national and international scholars as it is of high quality, that is complete, consistent, and reliable alike. Accordingly, several research collaborations with leading scholars in the field are already pursued.

## Literature

- Amable, B. (2003). The diversity of modern capitalism. Oxford, Oxford University Press.
- Gartner, W. B. (1988). "Who is an entrepreneur? Is the wrong question." American journal of small business **12**(4): 11-32.
- Hall, P. and D. Soskice (2001). Varieties of capitalism, Oxford University Press Oxford.
- Hancké, B., M. Rhodes, et al., Eds. (2007). Beyond Varieties of Capitalism. Oxford, Oxford University Press.
- Pfeffer, J. and G. R. Salancik (1978). The External Control of Organizations: A Resource Dependence Perspective. New York, NY, Harper and Row.
- Reynolds, P., N. Bosma, et al. (2005). "Global Entrepreneurship Monitor: Data Collection Design and Implementation 1998–2003." Small Business Economics **24**(3): 205-231.
- Reynolds, P. D. and R. T. Curtin (2007). Panel Study of Entrepreneurial Dynamics Program - Rationale and Description. <http://www.psed.isr.umich.edu/psed/background>, University of Michigan.
- Reynolds, P. D. and R. T. Curtin (2008). "Business Creation in the United States: Panel Study of Entrepreneurial Dynamics II Initial Assessment." Foundations and Trends in Entrepreneurship **4**(3): 155-307.
- Schneider, M. R. and M. Paunescu (2012). "Changing varieties of capitalism and revealed comparative advantages from 1990 to 2005: A test of the Hall and Soskice claims " Socio-Economic Review **10**(4): 731–753.