RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

DOCUMENTATION OF RISIS DATASETS

Cheetah

Massimiliano Guerini Politecnico di Milano

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 824091

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Outline

1 Basic Characteristics 2

2 Database content 2

2.1 Definition and description of observations 2
2.2 Data acquisition and processing (e.g. data cleaning) 2
2.3 Information on all variables/indicators 5
2.4 Sectorial, temporal and geographical coverage 7
2.5 Quality and accuracy of data 10

3 Technical Specifications 12

3.1 Information on the data base system 12
3.2 Technical variable definition 12
3.3 Description of the Entity Relationship Model (if applicable) 14
3.4 Interfaces for access and to other infrastructures (if applicable) 14

4 Scientific use cases and main references 14

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

1 Basic Characteristics

The Cheetah dataset contains geographical, industry, accounting and ownership information on three cohorts of medium-sized firms that experienced sales or employment fast growth rates in the periods 2008-2011, 2009-2012 and 2010-2013. These fast growing medium-sized firms (FGMFs) are located in 30 European countries (EU-28, Norway and Switzerland) plus Israel. Cheetah covers an overall number of 42,369 firms.

Cheetah has been developed in the context of the WP20 of the project RISIS - Research Infrastructure for Research and Innovation Policy Studies, funded by the European Commission under the Seventh Framework Program. The aim of Cheetah is to cover the long-term economic performance of FGMFs, as one of the main pillars of the European industrial and technological system.

The legal name of the operating organization is POLITECNICO DI MILANO, Department of Management, Economics and Industrial Engineering, located in VIA LAMBRUSCHINI 4/B, MILANO, 20156, Italy represented by Alessandro Perego, Head of Department (or his authorized representative).

2 Database content

2.1 Definition and description of observations

The database includes medium-sized firms that experienced fast-growth rate either in terms of three-years employment growth or three-years turnover growth in at least one of the growth periods of 2008-2011, 2009-2012 and 2010-2013 and are located in 30 European countries in addition to Israel. The main unit of observation is the firm.

The database includes 42,369 FGMFs. Detailed accounting and ownership information has been collected for each firm. For the first cohort of FGMFs (2008-2011) we identified 17,438 firms, for the second cohort (2009-2012) 24,993 firms, and for the third cohort (2010-2013) 16,085 firms. It is worth noting that the sum of firms in the three cohorts is higher than the number of firms in the database. This is because some firms may have experienced high growth in more than one observation period.

2.2 Data acquisition and processing (e.g. data cleaning)

The main source of information is Orbis (www.bvdinfo.com). The criteria for inclusion are as follows:

- firms are located in Europe (28 EU countries, CH, NO) or Israel;
- firms are medium-sized at the beginning of each observation period (2008, 2009, 2010);
- firms experienced fast-growth in at least one of the three observation periods (2008-11; 2009-12; 2010-13).

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

The data collection process consisted in five main steps, which are described in the following:

1. Identification of European and Israeli medium-sized firms (employee, turnover and balance sheet criteria).
2. Imputation of employee information if data is missing (around 40%).
3. Selection of fast growing firms by applying OECD definition.
4. Data collection of accounting and ownership information.
5. Geocoding.

1. Identification of European and Israeli medium-sized firms

Medium-sized firms are defined according to definitions of EUROSTAT and *Entreprise de taille intermŭdiaire* (ETI). EUROSTAT defines medium-sized firms as firms with a number of employees between 50 and 249, and either a turnover of not exceeding €50 million or a balance sheet total of not exceeding €43 million. ETIs are firms with a number of employees between 250 and 4999, and either a turnover of not exceeding €1.5 billion or a balance sheet total of not exceeding €2 billion. A firm that has less than 250 employees but a turnover of more than €50 million and a balance sheet total of more than €43 million is also considered a ETI.

In order to identify European and Israeli medium-sized firms in ORBIS we first applied the turnover and balance sheet thresholds. Accordingly, we selected firms (located in the 30 European countries and Israel) with a turnover lower than €1,5 billion or a balance sheet total lower than €2 billion (measured at 2008, 2009 and 2010). This step led to the identification of 13,034,346 firms.

During the data cleaning process all the observations for which the turnover data in 2008, 2009 and 2010 were missing have been removed. In addition, we excluded firms without BvD ID code (the firms unique identification numbers in the ORBIS database). Finally, the result of the cleaning process was a preliminary list of 8,609,209 European potential medium-sized firms.

2. Imputation of employee information

In order to apply the employee threshold (i.e. number of employees between 50 and 4999), we had first to deal with the problem of missing data. Indeed, the information for the number of employees was missing for about 40% of firms in our preliminary list of potential medium-sized firms. To overcome this issue, we imputed data by predicting the number of employees for the missing values. We used a Tobit regression model to obtain the predicted value of the number of employees on the basis of firms turnover, age, industry dummies, country dummies, and year dummies.

Eventually, after predicting the missing values for number of employees we applied the number of employees threshold to define the final population of European and Israeli medium-sized firms. We applied this threshold for the firms at the beginning of each growth period (2008, 2009 and 2010) in order to assure that only medium-sized firms enter each growth observation period. The final population of medium-sized firms includes 306,278 firms.

3. Selection of fast growing firms by applying OECD definition

After identifying the population of medium-sized firms, we selected those firms that experienced fast growth rates in the periods 2008-2011, 2009-2012 and 2010-2013. To do so, we applied the

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

definition of fast growing firms developed by OECD. According to this definition all firms with average annualized growth of greater than 20% over a three-year period, should be considered as high-growth firms. Growth is measured by either the number of employees or by turnover.

Following the above definition, we identified FGMFs using both employment and turnover growth thresholds for each of the three observation periods (2008-2011, 2009-2012 and 2010-2013). The final population of FGMFs consists of 42,369 firms.

4. Data collection of accounting and ownership information

For the final population of FGMFs, we downloaded from ORBIS accounting information (balance sheet, income statement and cash-flow) for the period 2008-2014.

We also exercised additional data collection efforts in order to identify business groups. We checked whether a firm was NOT controlled (i.e. independent) by another firm/State in the period 2007- 2013. To this aim, we used the ORBIS Ownership sub-database, which contains information on a firms equity ownership structure, specifically:

- names of owners;
- ownership shares (direct or total);
- type of owner (firms, individuals, State…).

Based on the ownership shares and their type we checked whether a firm was independent in each year. A firm is considered independent if:

- the sum of the shares of individual shareholders is greater than 50% OR
- the highest share of shareholders that are firms/State is lower than 50%.

We consider the status based on direct shares, and total shares if information on direct shares was missing.

5. Geocoding

Firms have been geocoded using Google API (i.e. retrieving latitude and longitude coordinates). We applied the following procedure:

1. geocoding based on full addresses or zipcode plus country (the most complete information available);
2. geocoding based on city plus country;
3. calculation of the geographical distance by using the coordinates obtained at the previous points;
4. for cases in which the geographical distance was more than 40km (9% of cases), we manually checked the correct latitude and longitude;
5. in all other cases we kept as final coordinates the ones associated with the full address or zipcode.

We were able to geolocalize the vast majority of the companies (42,284 companies on the total of 42,369 companies). On the basis of the geographical coordinates, we associated information on NUTS regions and Functional Urban Areas (FUAs).

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

2.3 Information on all variables/indicators

Table 1 describes variables names and descriptions.

Table 1. Variables descriptions

Variable Name	Variable Description
ID	Firm identification code.
name	Firm Name - updated on 2017.
country_code	Country code (ISO).
country	Country name.
city	City name.
foundation	Incorporation year.
nace_core	Core industry code (NACE rev. 2 4 digits).
post_code	Postcode.
address	Street, no., building etc, line 1.
latitude	Latitude.
longitude	Longitude.
region	Region of the City (as reported in ORBIS).
region_type	Type of Region (as reported in ORBIS).
nuts1	NUTS1 region.
nuts2	NUTS2 region.
nuts3	NUTS3 region.
FUA_name	Name of the Functional Urban Area.
FUA_code	Code of the Functional Urban Area.
tel	Telephone Number.
fax	Fax Number.
website	Website address.
email	E-mail address.
guo	Global Ultimate Owner (year 2017).
turnoverYEAR	Operating Revenue [Turnover] YEAR = 2007-2014 Thousands Euros.
tot_assetYEAR	Total Assets YEAR = 2007-2014 Thousands Euros.
num_empYEAR	Number of Employees YEAR = 2007-2014.
cost_empYEAR	Cost of Employees YEAR = 2007-2014 Thousands Euros.
salesYEAR	Sales YEAR = 2007-2014 Thousands Euros.
added_valueYEAR	Added Value YEAR = 2007-2014 Thousands Euros.
ebitdaYEAR	EBITDA YEAR = 2007-2014 Thousands Euros.
ebitYEAR	Operating P/L [=EBIT] YEAR = 2007-2014 Thousands Euros.
depreciationYEAR	Total Depreciation & Amortization YEAR = 2007-2014 Thousands Euros.
net_incomeYEAR	Net Income YEAR = 2007-2014 Thousands Euros.
cashflowYEAR	Cash Flow YEAR = 2007-2014 Thousands Euros.
cash_equivalentYEAR	Cash & Equivalent YEAR = 2007-2014 Thousands Euros.
intang_assetYEAR	Intangible Fixed Assets YEAR = 2007-2014 Thousands Euros.
tang_assetYEAR	Tangible Fixed Assets YEAR = 2007-2014 Thousands Euros.
fix_assetYEAR	Total Fix Assets YEAR = 2007-2014 Thousands Euros.
other_fix_assetsYEAR	Other Fixed Assets YEAR = 2007-2014 Thousands Euros.
current_assetYEAR	Current Assets YEAR = 2007-2014 Thousands Euros.

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

other_current_assetYEAR	Other Current Assets YEAR = 2007-2014 Thousands Euros.
working_capYEAR	Working Capital YEAR = 2007-2014 Thousands Euros.
longterm_debtYEAR	Long Term Debt YEAR = 2007-2014 Thousands Euros.
current_liabYEAR	Current Liabilities YEAR = 2007-2014 Thousands Euros.
tot_liability_fundsYEAR	Total Shareholders Funds and Liabilities YEAR = 2007-2014 Thousands Euros.
tot_equityYEAR	Total Shareholders Funds YEAR = 2007-2014 Thousands Euros.
interest_expensesYEAR	Total interest expenses YEAR = 2007-2014 Thousands Euros.
income_before_taxYEAR	Profit before tax YEAR = 2007-2014 Thousands Euros.
indep_aroundYEAR	Independency dummy based on Direct and total Ownership YEAR = 2008-2013.
FG_turnover_08_11	Average annualized growth rate based on turnover Cohort 2008-2011.
FG_turnover_09_12	Average annualized growth rate based on turnover Cohort 2009-2012.
FG_turnover_10_13	Average annualized growth rate based on turnover Cohort 2010-2013.
FG_employee_08_11	Average annualized growth rate based on employees Cohort 2008-2011.
FG_employee_09_12	Average annualized growth rate based on employees Cohort 2009-2012.
FG_employee_10_13	Average annualized growth rate based on employees Cohort 2010-2013.
FG_turnover_sample_08_11	Fast Growing Firm dummy based on turnover Cohort 2008-2011.
FG_turnover_sample_09_12	Fast Growing Firm dummy based on turnover Cohort 2009-2012.
FG_turnover_sample_10_13	Fast Growing Firm dummy based on turnover Cohort 2010-2013.
FG_employee_sample_08_11	Fast Growing Firm dummy based on employees Cohort 2008-2011.
FG_employee_sample_09_12	Fast Growing Firm dummy based on employees Cohort 2009-2012.
FG_employee_sample_10_13	Fast Growing Firm dummy based on employees Cohort 2010-2013.
sample0_2008	Mid-sized dummy based on actual employees 2008.
sample0_2009	Mid-sized dummy based on actual employees 2009.
sample0_2010	Mid-sized dummy based on actual employees 2010.
sample1_2008	Mid-sized dummy based on imputed employees 2008.
sample1_2009	Mid-sized dummy based on imputed employees 2009.
sample1_2010	Mid-sized dummy based on imputed employees 2010.
cohort2008	Dummy = 1 if sample0_2008 = 1 & (FG_turnover_sample_08_11 =1 OR FG_employee_sample_08_11 = 1).
cohort2009	Dummy = 1 if sample0_2009 = 1 & (FG_turnover_sample_09_12 =1 OR FG_employee_sample_09_12 = 1).
cohort2010	Dummy = 1 if sample0_2010 = 1 & (FG_turnover_sample_10_13 =1 OR FG_employee_sample_10_13 = 1).
sum_cohort	#periods firms are FGMFs based on actual employees.
cohort2008_ext	Dummy = 1 if sample1_2008 = 1 & (FG_turnover_sample_08_11 =1 OR FG_employee_sample_08_11 = 1).
cohort2009_ext	Dummy = 1 if sample1_2009 = 1 & (FG_turnover_sample_09_12 =1 OR FG_employee_sample_09_12 = 1).
cohort2010_ext	Dummy = 1 if sample1_2010 = 1 & (FG_turnover_sample_10_13 =1 OR FG_employee_sample_10_13 = 1).
sum_cohort_ext	#periods firms are FGMFs based on imputed employees.

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

2.4 Sectorial, temporal and geographical coverage

Sectorial classification follows the NACE Rev. 2 classification. Table 2 show firms according to the NACE Division (2-digit NACE code).

Table 2. Sectoral coverage

NACE code	NACE Description	N.	%
01	Crop and animal production, hunting and related service activities	549	1.3
02	Forestry and logging	97	0.23
03	Fishing and aquaculture	40	0.09
05	Mining of coal and lignite	12	0.03
06	Extraction of crude petroleum and natural gas	81	0.19
07	Mining of metal ores	41	0.1
08	Other mining and quarrying	97	0.23
09	Mining support service activities	168	0.4
10	Manufacture of food products	1,529	3.61
11	Manufacture of beverages	144	0.34
12	Manufacture of tobacco products	26	0.06
13	Manufacture of textiles	312	0.74
14	Manufacture of wearing apparel	523	1.23
15	Manufacture of leather and related products	330	0.78
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	339	0.8
17	Manufacture of paper and paper products	254	0.6
18	Printing and reproduction of recorded media	177	0.42
19	Manufacture of coke and refined petroleum products	60	0.14
20	Manufacture of chemicals and chemical products	610	1.44
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	289	0.68
22	Manufacture of rubber and plastic products	818	1.93
23	Manufacture of other non-metallic mineral products	394	0.93
24	Manufacture of basic metals	591	1.4
25	Manufacture of fabricated metal products, except machinery and equipment	1,854	4.38
26	Manufacture of computer, electronic and optical products	809	1.91
27	Manufacture of electrical equipment	682	1.61
28	Manufacture of machinery and equipment n.e.c.	1,829	4.32
29	Manufacture of motor vehicles, trailers and semi-trailers	774	1.83
30	Manufacture of other transport equipment	275	0.65
31	Manufacture of furniture	349	0.82
32	Other manufacturing	417	0.98
33	Repair and installation of machinery and equipment	358	0.85
35	Electricity, gas, steam and air conditioning supply	334	0.79
36	Water collection, treatment and supply	87	0.21
37	Sewerage	31	0.07
38	Waste collection, treatment and disposal activities; materials recovery	406	0.96
39	Remediation activities and other waste management services	29	0.07
41	Construction of buildings	1,308	3.09

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

42	Civil engineering	784	1.85
43	Specialised construction activities	1,100	2.6
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	700	1.65
46	Wholesale trade, except of motor vehicles and motorcycles	2,920	6.89
47	Retail trade, except of motor vehicles and motorcycles	1,714	4.05
49	Land transport and transport via pipelines	1,154	2.72
50	Water transport	84	0.2
51	Air transport	57	0.13
52	Warehousing and support activities for transportation	772	1.82
53	Postal and courier activities	71	0.17
55	Accommodation	542	1.28
56	Food and beverage service activities	529	1.25
58	Publishing activities	238	0.56
59	Motion picture, video and television programme production, sound recording and music publishing activities	120	0.28
60	Programming and broadcasting activities	63	0.15
61	Telecommunications	271	0.64
62	Computer programming, consultancy and related activities	1,232	2.91
63	Information service activities	182	0.43
64	Financial service activities, except insurance and pension funding	1,590	3.75
65	Insurance, reinsurance and pension funding, except compulsory social security	274	0.65
66	Activities auxiliary to financial services and insurance activities	337	0.8
68	Real estate activities	585	1.38
69	Legal and accounting activities	200	0.47
70	Activities of head offices; management consultancy activities	1,446	3.41
71	Architectural and engineering activities; technical testing and analysis	645	1.52
72	Scientific research and development	222	0.52
73	Advertising and market research	303	0.72
74	Other professional, scientific and technical activities	239	0.56
75	Veterinary activities	9	0.02
77	Rental and leasing activities	190	0.45
78	Employment activities	1,062	2.51
79	Travel agency, tour operator and other reservation service and related activities	163	0.38
80	Security and investigation activities	487	1.15
81	Services to buildings and landscape activities	733	1.73
82	Office administrative, office support and other business support activities	1,043	2.46
84	Public administration and defence; compulsory social security	133	0.31
85	Education	475	1.12
86	Human health activities	799	1.89
87	Residential care activities	446	1.05
88	Social work activities without accommodation	381	0.9
90	Creative, arts and entertainment activities	104	0.25
91	Libraries, archives, museums and other cultural activities	68	0.16
92	Gambling and betting activities	135	0.32
93	Sports activities and amusement and recreation activities	268	0.63
94	Activities of membership organisations	103	0.24

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

95	Repair of computers and personal and household goods	47	0.11
96	Other personal service activities	310	0.73
97	Activities of households as employers of domestic personnel	1	0
99	Activities of extraterritorial organisations and bodies	4	0.01

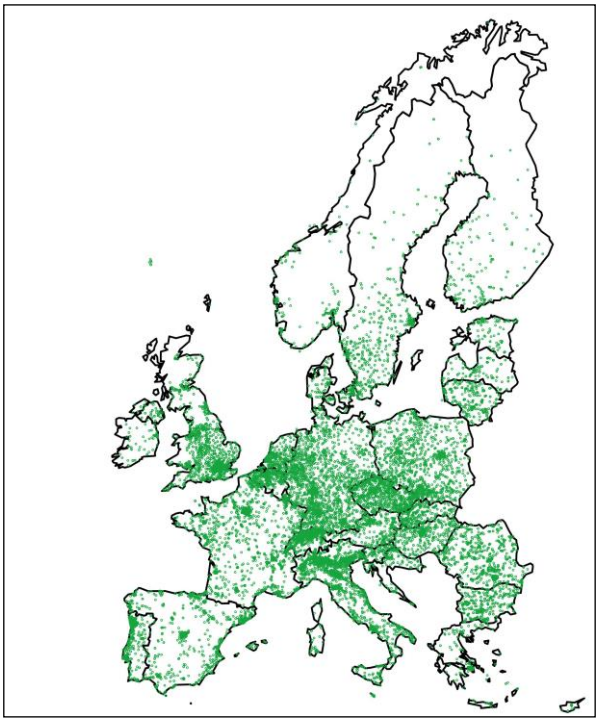
As to the temporal coverage, accounting information is available from 2007 to 2014 and ownership information from 2008 to 2013. Overall, FGMFs are relatively young. The median foundation year of is 1996. Table 3 show the distribution of FGMFs by foundation period. More than 70% of FGMFs have been incorporated from 1990 onwards.

Table 3. Temporal coverage

Foundation period	N.	%
Before 1980	6,632	16.08
1980-1989	4,695	11.39
1990-1999	14,021	34.00
2000-2009	15,889	38.53

As to the geographical coverage, the database includes information on the country, NUTS region (https://ec.europa.eu/eurostat/web/regions-and-cities/overview), FUA (https://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm), and exact geographical coordinates (i.e. latitude and longitude). Overall the database covers 30 European countries plus Israel. Figure 1 shows the distribution of European FGMFs based on exact geographical coordinates. Table 4 shows the distribution of FGMFs by country.

Figure 1. Geographical coverage



This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Table 4. Geographical coverage

Country	N.	%
Austria	693	1.64
Belgium	675	1.59
Bulgaria	1,613	3.81
Croatia	294	0.69
Cyprus	20	0.05
Czech Republic	2,235	5.28
Denmark	333	0.79
Estonia	296	0.7
Finland	584	1.38
France	2,746	6.44
Germany	4,135	9.76
Greece	429	1.01
Hungary	944	2.23
Ireland	292	0.69
Israel	120	0.28
Italy	3,551	8.38
Latvia	526	1.24
Lithuania	941	2.22
Luxembourg	92	0.22
Malta	46	0.11
Netherlands	876	2.07
Norway	447	1.06
Poland	3,344	7.89
Portugal	801	1.89
Romania	2,347	5.54
Slovakia	660	1.56
Slovenia	155	0.37
Spain	2,272	5.15
Sweden	2,176	5.14
Switzerland	1,285	3.03
United Kingdom	7,440	17.56

2.5 Quality and accuracy of data

Table 5 reports the number of missing values and the percentage of missing values with respect to the number of observations (i.e. 42,369 firms) for the all variables included in Cheetah apart from those concerning accounting and ownership information.

Table 5. Missing values

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Variable Name	N. of missing values	% of missing values
ID	0	0
name	0	0
country_code	0	0
country	0	0
city	74	0.17
foundation	1,132	2.67
nace_core	11	0.03
post_code	1,614	3.81
address	506	1.19
latitude	85	0.20
longitude	85	0.20
region	37,246	87.91
region_type	37,246	87.91
nuts1	204	0.48
nuts2	204	0.48
nuts3	204	0.48
FUA_name	18,402	43.43
FUA_code	18,402	43.43
tel	40,933	96.61
fax	41,578	98.13
website	6,376	15.05
email	14,240	33.61
guo	8,963	21.15
FG_turnover_08_11	6,963	16.43
FG_turnover_09_12	4,211	9.94
FG_turnover_10_13	8,698	20.53
FG_employee_08_11	11,364	26.82
FG_employee_09_12	8,600	20.3
FG_employee_10_13	12,472	29.44
FG_turnover_sample_08_11	6,963	16.43
FG_turnover_sample_09_12	4,211	9.94
FG_turnover_sample_10_13	8,698	20.53
FG_employee_sample_08_11	11,364	26.82
FG_employee_sample_09_12	8,600	20.30
FG_employee_sample_10_13	12,472	29.44
sample0_2008	2,740	6.47
sample0_2009	2,204	5.20
sample0_2010	2,972	7.01
sample1_2008	0	0.00
sample1_2009	0	0.00
sample1_2010	0	0.00
cohort2008	0	0.00
cohort2009	0	0.00
cohort2010	0	0.00
sum_cohort	0	0.00

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

cohort2008_ext	0	0.00
cohort2009_ext	0	0.00
cohort2010_ext	0	0.00
sum_cohort_ext	0	0.00

3 Technical Specifications

3.1 Information on the data base system

The database is currently available in STATA format (.dta).

3.2 Technical variable definition

The database is organized in one table. Table 6 reports names and types of variables in Cheetah. The unique identifier (ID) is reported in bold.

Table 6. Variables types

Variable Name	Variable Type
ID	**String**
name	String
country_code	String
country	String
city	String
foundation	Integer
nace_core	String
post_code	String
address	String
latitude	Float
longitude	Float
region	String
region_type	String
nuts1	String
nuts2	String
nuts3	String
FUA_name	String
FUA_code	String
tel	String
fax	String
website	String
email	String
guo	String
turnoverYEAR	Float
tot_assetYEAR	Float

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

num_empYEAR	Float
cost_empYEAR	Float
salesYEAR	Float
added_valueYEAR	Float
ebitdaYEAR	Float
ebitYEAR	Float
depreciationYEAR	Float
net_incomeYEAR	Float
cashflowYEAR	Float
cash_equivalentYEAR	Float
intang_assetYEAR	Float
tang_assetYEAR	Float
fix_assetYEAR	Float
other_fix_assetsYEAR	Float
current_assetYEAR	Float
other_current_assetYEAR	Float
working_capYEAR	Float
longterm_debtYEAR	Float
current_liabYEAR	Float
tot_liability_fundsYEAR	Float
tot_equityYEAR	Float
interest_expensesYEAR	Float
income_before_taxYEAR	Float
indep_aroundYEAR	Boolean
FG_turnover_08_11	Float
FG_turnover_09_12	Float
FG_turnover_10_13	Float
FG_employee_08_11	Float
FG_employee_09_12	Float
FG_employee_10_13	Float
FG_turnover_sample_08_11	Boolean
FG_turnover_sample_09_12	Boolean
FG_turnover_sample_10_13	Boolean
FG_employee_sample_08_11	Boolean
FG_employee_sample_09_12	Boolean
FG_employee_sample_10_13	Boolean
sample0_2008	Boolean
sample0_2009	Boolean
sample0_2010	Boolean
sample1_2008	Boolean
sample1_2009	Boolean
sample1_2010	Boolean
cohort2008	Boolean
cohort2009	Boolean
cohort2010	Boolean
sum_cohort	Integer

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

cohort2008_ext	Boolean
cohort2009_ext	Boolean
cohort2010_ext	Boolean
sum_cohort_ext	Integer

3.3 Description of the Entity Relationship Model (if applicable)

Not applicable.

3.4 Interfaces for access and to other infrastructures (if applicable)

Cheeath is integrated with the other firm-level RISIS databases through the FirmReg facility. FirmReg is a central facility within RISIS, which aims at uniquely identifying and tracking over time the businesses included in firms datasets that are part of RISIS, such as CIB on the largest innovative industrial firms worldwide, Cheetah on European mid-size fast growing firms and VICO. For each entity (company), Cheeath is connected to FirmReg with an unambiguous and stable (over time) identifier for the entity (ID).
At the geographical level, Cheeath is harmonized with the other research infrastructures in RISIS by adopting the NUTS classification of administrative units (as provided by Eurostat) and the FUA classification for urban areas (as provided by the OECD).
Integration of Cheetah in the RCF will be allowed with anonymised company names, which is essential for legal issues due to data retrieval from commercial databases. Only in specific cases when a high level of data protection is guaranteed, such as in case of physical visits and direct research collaborations with the datasets owners, companies names may be disclosed.

4 Scientific use cases and main references

In recent years, high growth firms received considerable attention by academic scholars and policymakers (Coad et al., 2014). One of the reasons for this increased interest is the fact that high growth firms play a crucial role in creating new jobs and employment. However, at the same time, the literature has pointed out that firm performance appears to be highly skewed and most firms do not experience high growth (Stanley et al., 1996; Bottazzi and Secchi, 2006), that high growth is difficult to foresee, and the determinants of high growth are difficult to predict (Marsili, 2001; Coad, 2009), and that high growth tends not to persist over time (Coad, 2007; Daunfeldt and Halvarsson, 2012). In addition, the literature on high growth firms (or fast growing firms) has focused on several characteristics of these firms, including: their size (Delmar and Davidsson, 1998; Delmar et al., 2003), age (Delmar et al., 2003; Haltiwanger et al., 2013), sector (Delmar et al., 2003; Davidsson and Delmar, 2003, 2006), or country (Schreyer, 2000; Bravo-Biosca, 2010).
Within the extant literature on fast growing firms, one category of firms received somewhat less attention: medium-sized firms. Most studies, in fact, tend to focus on either small entrepreneurial fast growing firms or large ones.
Cheetah dataset offers the unique opportunity to address this research gap, providing original data about a large sample of FGMFs. For instance, the role of FGMFs within the local ecosystems

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

could be analysed thanks to the large geographical coverage of the database (EU28, Norway, Switzerland, and Israel), while the sectorial classification of FGMFs could be explored more in depth in order to find different growth patterns (understanding if countries tend to specialize within particular sectors). Furthermore, the panel data structure of the dataset (regarding the financial data) could be exploited to investigate the determinants of growth persistence over time. Please refer to the final policy brief of the previous RISIS project for preliminary evidence based on Cheetah.

References

Bottazzi, G., and Secchi, A. (2006), Explaining the distribution of firm growth rates, Rand Journal of Economics, 37(2), 235-256.

Bravo-Biosca, A. (2010), Growth dynamics: exploring business growth and contraction in Europe and the US, NESTA: London, UK.

Coad, A. (2007), A closer look at serial growth rate correlation, Review of Industrial Organization, 31, 69 82.

Coad, A. (2009), The Growth of Firms: A Survey of Theories and Empirical Evidence. Edward Elgar: Cheltenham, UK.

Coad, A., Daunfeldt, S.O., Hυlzl, W., Johansson, D., and Nightingale, P. (2014). High-growth firms: introduction to the special section. Industrial and Corporate Change, 23(1), 91-112.

Daunfeldt, S.O., and Halvarsson, D. (2012), Are high-growth firms one hit wonders? Evidence from Sweden, HUI Working Paper No 70, HUI Research: Stockholm, Sweden.

Davidsson, P., and Delmar, F. (2003), Hunting for new employment: the role of high growth firms, in D. A. Kirby and A. Watson (eds), Small Firms and Economic Development in Developed and Transition Economies: A Reader. Ashgate Publishing: Hampshire, UK, pp. 7-19.

Davidsson, P., and Delmar, F. (2006), High-growth firms and their contribution to employment: The Case of Sweden, in P. Davidsson, F. Delmar and J. Wiklund (eds), Entrepreneurship and the Growth of Firms. Edward Elgar: Cheltenham, UK and Northampton, MA, pp. 156-178.

Delmar, F., and Davidsson, P. (1998), A taxonomy of high-growth firms, in P. D. Reynolds and W. D. Bygrave (eds), Frontiers of Entrepreneurship Research. Babson College: Wellesley, MA, pp. 399-413.

Delmar, F., Davidsson, P., and Gartner W.B. (2003), Arriving at the high-growth firm, Journal of Business Venturing, 18(2), 189-216.

Marsili, O. (2001), The Anatomy and Evolution of Industries. Edward Elgar: Cheltenham, UK.

Schreyer, P. (2000), High-growth firms and employment, OECD Science, Technology and Industry Working Papers. No. 2000/03.

Stanley, M.H., Amaral, L.A., Buldyrev, S.V., and Havlin, S. (1996). Scaling behaviour in the growth of companies. Nature, 379, 804-806.