Austrian Journal of Statistics
Volume 33 (2004), Number 1&2, 211-235

The Development of an Integrated and Systematized Information System for Economic and Policy Impact Analysis

Filippo Oropallo and Francesca Inglese
ISTAT (Italian Statistical Institute), Rom[1]

Abstract: The paper addresses the integration problems that have been faced in reconciling administrative and survey sources and combining them into one multi-source database. It shows the architecture of the integration process that has been adopted and the exploitation of the integrated database for economic and policy impact analysis at a micro level. The integration of administrative and survey data is performed by exact matching when the same unit is identified otherwise it is performed by statistical matching techniques. To apply these techniques, matching variables are required: one quite apparent option is to use firm characteristics as provided by the business register. The development of the Enterprise Integrated and Systematized Information System (EISIS) opens new possibility in microsimulation analysis to study the tax burden and the economic performance of enterprises through the construction of micro-founded indicators. IT (Information Technology) features of the whole process are also described that are the formalization of the integration process and the structure of the user friendly interface of the integration software. Confidentiality is satisfied by remote processing on a protected server that is only accessible to granted users of the National Statistical Institute.

Zusammenfassung: Das Manuskript behandelt Probleme des Zusammenführens von administrativen Daten mit Daten aus statistischen Erhebungen zum Erstellen eines integrierten Datenbestandes von Unternehmensdaten für Zwecke von Mikro-Simulationen. Die Struktur des Integrationsprozesses wird vorgestellt. Zum Zusammenführen der Daten wird exaktes Matching bei Vorhandensein identischer Objekte in den zu integrierenden Datenbeständen, andernfalls statistisches Matching verwendet. Als Matching-Variable bieten sich Charakteristika der Unternehmen an, wie sie im Unternehmensregister zur Verfügung stehen. Die Entwicklung von EISIS (Enterprise Integrated and Systematized Information System) ermöglicht Mikro-Simulationen, mit deren Hilfe durch Analyse von Mikro-Indikatoren die Auswirkungen von verschiedenen

[1] The work is developed under the DIECOFIS (Development of a System of Indicators on competitiveness and Fiscal Impact on Enterprise Performance) project, financed by the Information Society Technologies Programme (IST-2000-31125) coordinated by Paolo Roberti ISTAT (Italian Institute of Statistics).
The authors are grateful to Maria Grazia Calza, Alessandro Fazio, Lorenzo Lo Cascio and Danilo Orsini for their assistance and valuable suggestions. This paper draws heavily on the research they carried out under the DIECOFIS project.

Maßnahmen auf die steuerliche Belastung und auf die Wettbewerbssituation untersucht werden. Das Manuskript beschreibt die informationstechnologische Realisierung der Daten-Integration und die Benützerschnittstelle der entwickelten Software. Anforderungen des Datenschutzes werden durch Implementierung des Systems auf einem geschützten Server und durch Zutrittsbeschränkungen sichergestellt.

Keywords: Multisource Database Integration, Exact and Statistical Matching, Imputation, Microsimulation, Micro-founded Indicators

1 Introduction

The Italian National Statistical Institute (ISTAT) is currently involved in a number of EU IST projects with the aim of supporting the Lisbon objectives, EU governance and national policy making processes with “best” EU-wide and national policy impact and evaluation analyses. The knowledge for policy impact analysis which exists in the EU is limited and inadequte. . The gap with the USA is remarkable. The "facts" on the impact of policies continue to be charted only at the aggregate level and with a high degree of approximation. The aggregate indicators which are normally used have well-known pitfalls and drawbacks. Understanding how policies affect economic performance and developing better indicators to gauge their effects is essential to endow the EU with efficient and fair policies. The DIECOFIS EU-FP5 project has taken up the challenge of reducing this gap in the field of taxation. Results have been quite encouraging and have open new vistas for future work. Particularly notable has been the development of a system of micro-founded indicators, based on factuals and counterfactuals, estimated through micro-simulation models. Micro-founded indicators on enterprise performance and fiscal microsimulation models require a great deal of information. This is normally scattered in different statistical surveys and administrative sources. Each different data source is conceived to serve different purposes and, in many instances, may refer to different units or different definitions may be used for the same unit. Any attempt to bring together data from different sources has to overcome complex problems in terms of sheer access, and integration and systematisation.

Generally the integration process of various sources for the development of integrated systems can be performed through three different methodologies: merging, record linkage and statistical matching. The first and the second deal with the identification of the units in two or more different files, the third deals with the problem of integration when units in different files are not the same. When integration deals with sources whose units are identified through a univocal identification code, deterministic exact matching is possible; if the involved units do not have univocal identity code, connection criteria among different information coming from different sources have to be established: in this case other record linkage techniques can be used. These techniques have the scope of identifying pairs of records from two data sources related with the same unit (cf. Jabine, Sheuren 1986). If the sources do not have the same unit, but they have common information, then statistical matching can be used in order to collect information of similar units respecting some criteria (cf. Paass, 1985).

This paper presents the creation of a multi-source integrated and systematised data base of enterprises data. Paragraphs 2 and 3 describe the sources utilised in the construction of DIECOFIS database; paragraph 4 exposes the architecture of the integration process and the solutions that has been adopted. Finally, for the construction of the datasets for microsimulation purposes, the problems determined by combination of data from the different sources are described: (i) the reconciliation of survey data with administrative data, (ii) the treatment of missing data and (iii) the sample weights adjustment.

2 DIECOFIS Database

In order to build the integrated and systematized information system on enterprises needed to support economic analysis and for the development of tax microsimulation models and micro founded indicators, the first step is to select the "spine" information that will be use as a basis for the integration process. At ISTAT, the "spine" is constituted by the statistical register of Italian active enterprises (ASIA)[2]. This is the result of an integration process of different administrative sources and represents the best "hanger" for data integration purposes. On this hanger, information from the following sources can be put. These include: Large Enterprise Accounts (SCI); Small and Medium Enterprise Survey with less than 100 workers (PMI); Manufacturing Product Survey (PRODCOM) and Cost Structure Survey (ISC); Foreign Trade Archive (COE); Community Innovation Survey (CIS, 1999); ICT Survey (2002). All above ISTAT surveys are based on common EUROSTAT standards and classifications (as shown in chart 1). This implies that the DIECOFIS database can serve to microsimulate the impact of public policies not only in Italy and that a path for the creation of an EU statistical information system has been traced. The main effort is concerned with the development of a methodology that allow the data linkage between the information of the above surveys and the whole enterprise universe , represented by the data register on enterprises. In the ASIA[3] archive, ISTAT files all active enterprises (cf. Eurostat 1999) except for those belonging to Agriculture, Forestry and Fishing (A, B sectors according to NACE classification) and the Public Sector (L, O91, P and Q). This can be used as a starting point or common basis for the linkage of all survey data. In the ASIA archive the following information is included: identifier (internal code, name, fiscal code, vat number, telephone, address); localisation (geographical reference); typology (economic activity and legal form); demographic (status and transformations); size (turnover and employees).

The information coming from the administrative sources that have been integrated in the DIECOFIS database include: Commercial Accounts (CA) data from the Chamber of Commerce annual report that complement ISTAT business survey of account system

[2] Basically, the actual enterprise state of activity is estimated by means of a logistic model, where the probability of existence is a function a various signs of life, drawn from different administrative sources.

[3] The ASIA project started in 1995, its goal is to improve and update the register of all Italian enterprises. It is the result of the integration of external sources with ISTAT Archives (old Sirio-nai archive, 7° Industry Census and survey SK). External sources are: VAT Register of the Ministry of Finances; Chambers of Commerce; INAIL (National Institute of Insurance Against Accidents at Work); INPS (National Social Security Institute); Yellow Pages.

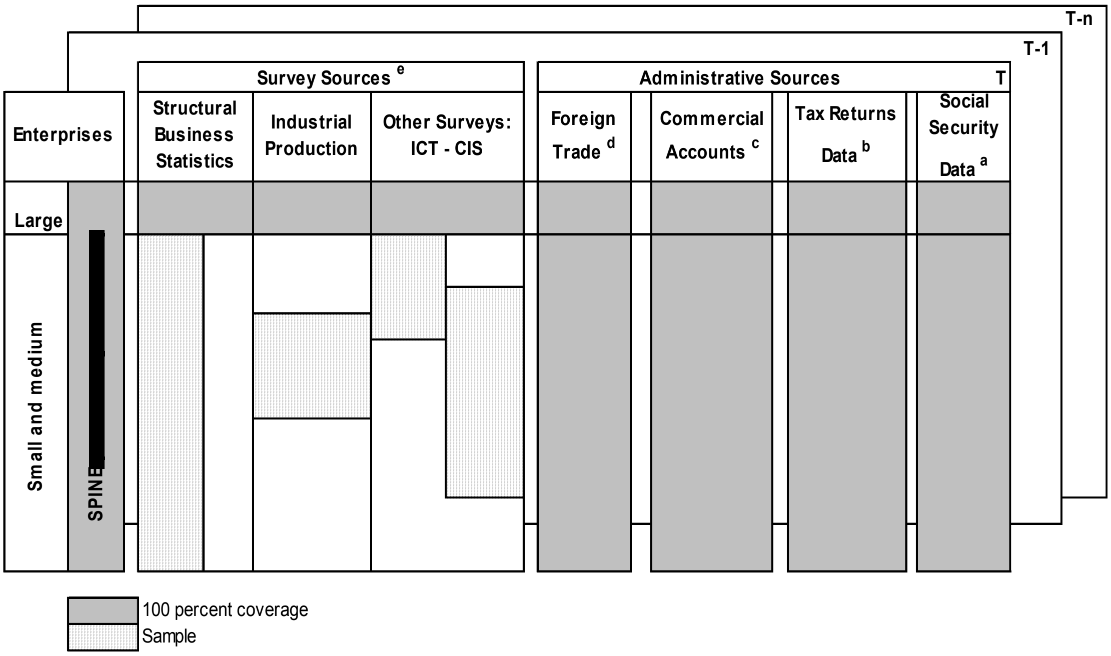
(SCI and PMI) for all corporate, co-operatives and consortium enterprises only; Fiscal data (FISCAL) from Revenue Agency annual tax returns; Social Security data (SSD) from the Italian Social Security Institute (INPS). These two latter sources permit to obtain precise information on tax and social contribution revenues, and thus to calculate the actual tax burden on enterprises, which can be used to test the model's output (e.g. "counterfactuals").

Looking at the quality of the available information, the enterprise size seems to be a "key" variable[4]. In fact, exhaustive information (which covers the whole universe) is available for large enterprises that have at least 100 workers, while small and medium ones are collected data from a sample of enterprises. A second characteristic that appears to be very important is the legal form, as the type of tax that an enterprise is required to pay depends on it.

If the attention is focused on the integration of statistical data with business registers and with administrative data (for a discussion on this topic see, for instance, EUROSTAT 1999, Giovannini, Sorce 2001), the first problem is to identify the business unit. This means basically choosing a variable which can be a unique key and act as a natural bridge between the different sources. In almost all firms' databases the ID code is represented by the VAT code or the fiscal code. Another important question relates to possible changes to the business (cf. Black 2001) during the enterprises' life. In fact, the same enterprise may appear as a different unit because of transformation events. Usually two types of changes are considered: changes involving a single unit (changes in kind of business classification, in size or localisation); changes in the number of units (death, birth, divestitures and splits, mergers and acquisitions). As a consequence of changes or in the presence of new-born firms, it found that the business register doesn't contain all the units of a survey and it is necessary to distinguish between the case of new firms and that of transformed units. In the latter case, a problem of identifying the successor of the initial business can arise. In some cases, the VAT number of the new unit is different but the fiscal code is the same. A correspondence table containing old and new codes or a table containing the fiscal code and the many VAT numbers used by the enterprise has been used in order to solve this kind of problems.

Being able to rely on an integrated and systematised database for a sufficient number of years, in this overall systemic perspective, basically means that it would become possible to go beyond overall indicators (measuring "averages") and measures of inequality and dispersions (measuring overall inequality) and to "*slice*", "*dice*", "*drill up*" and "*drill down*", "*drill through*" and "*drill across*" information *hyper* and *micro cubes*, that is to move horizontally and vertically, across dimensions and over time, and chain link indicators. These *newly-built* indicators would refer to different dimensions and be characterised by systemic features that can be studied to identify factors/areas of weakness or strength; of progress or decline; gains and losses. Accordingly, it would become possible to study aspects relating to structure, composition, distribution and dispersion.

[4] See Denk, Oropallo F. (2003) Overview of the issues in longitudinal and cross-sectional multi-source databases



(a) Enterprise with employees only - This database is shaped by the National Social Security Legislation
(b) All enterprises - This database is shaped by the National Tax Legislation
(c) Incorporated enterprises only (Their account system is regulated by UE directives)
(d) Exporting enterprises only
(e) These Sources are governed by: (i) Council Regulations: no 3924/91 - survey of industrial production; (ii) n. 696/93 - statistical units for the analysis of the production system; (iii) n. 58/97 - structural business statistics; (iv) Commission Regulation: n. 1618/99 - evaluation of quality of structural business statistics;

Figure 1: General Framework

3 First stage of integration

The first step of the DIECOFIS database integration was concerned with the creation of the following datasets.

Survey datasets (Table 2 - Regional datasets), which include information on 55 thousand firms in 1999, coming from several ISTAT surveys. Information is exhaustive for large firms (8.8 thousand firms with 100 workers or more, SCI). Sample data are available for small firms (45.9 thousand firms, PMI).

Administrative datasets (Table 3, Corporate datasets), which include information on 53.5 thousand corporate firms in 1999, coming from the commercial account database. Information is exhaustive for large corporate firms (roughly 7 thousand firms with 100 workers or more). Sample data are available for small corporates (roughly 46.6 thousand firms).

Business Register ASIA (Table 1, from 1996 to 2001) includes basic information on the whole universe of Italian active enterprises, so that the contained characteristics can serve as auxiliary variables in the processes of imputation and estimation. These are: geographical reference, sector of economic activity, legal type dimension (independent workers and employees) and, for a portion of them, the annual turnover. This merging activity makes it possible to proceed with the second stage of integration: the missing data reconstruction, obtained by using matching techniques.

Table 1: Business Register

Year	*1996*	*1997*	*1998*	*1999*	*2000*	*2001*
Large Enterprises	8,091	8,684	8,924	9,240	9,741	10,125
Small-Medium Ent.	3,862,383	3,761,446	4,040,250	4,122,853	4,212,916	4,287,340
Total	3,870,474	3,770,130	4,049,174	4,132,093	4,222,657	4,297,465

The 1999 SCI survey (data are available also for year 1998) contains information about 8,734 enterprises. They refer to the universe of large enterprises with 100 or more workers (with the exclusion of the J division, Financial sector). Among these, there are 7,340 corporate enterprises, about 84 percent of the total. The 1999 PMI survey (data are available also for year 1998) contains information about 45,867 enterprises, of which 15,330 are corporate enterprises.

Table 2: Survey data

	SCI		PMI	
Year	*1998*	*1999*	*1998*	*1999*
Corporate Enterprises	7,124	7,340	15,372	15,330
Non-corporate Ent.	1,330	1,394	32,112	30,617
Total	8,454	8,734	47,484	45,867

With respect to CA data (years ’98-’00), there is, in the 1999, a sample of 53,532 enterprises: 6,911 of these are present in the SCI survey as well and the others are small corporate enterprises to be linked with survey data. The FISCAL dataset contains a targeted sample of tax returns for the year 1999. It contains all large corporates and a very small sample of small enterprises. The SSD contains data on all enterprises which have one employee at least.

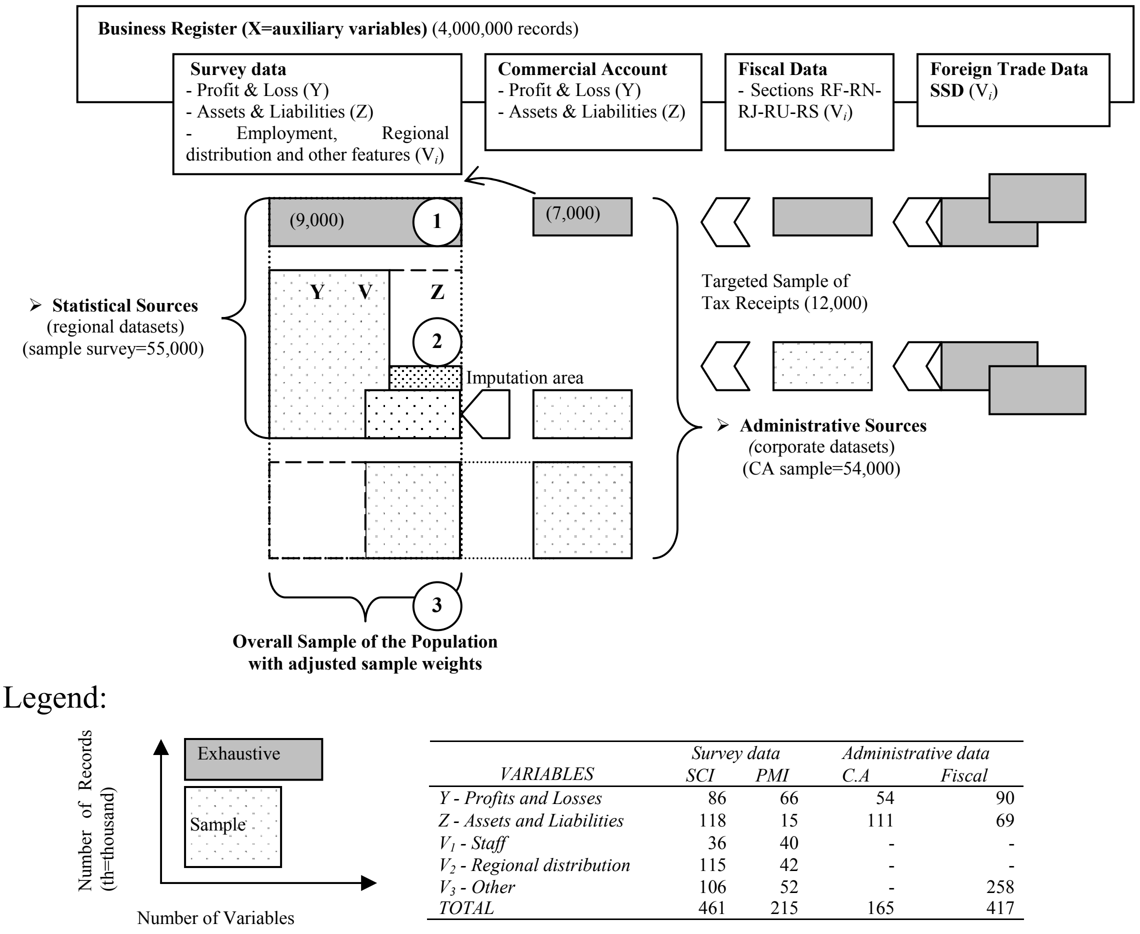
Table 3: Administrative data

	CA			Fiscal	SSD
Year	*1998*	*1999*	*2000*	*1999*	*1999*
Large Enterprises	6,197	6,911	6,082	7,340	9,239
SME Enterprises	48,261	46,621	43,349	4,535	1,061,714
Total	54,458	53,532	49,431	11,875	1,070,953

In this first stage, other surveys have been linked with the business register. 1) *PRODCOM survey* (cf. ISTAT 2001a) is exhaustive for large enterprises and there is a sample of small and medium ones (approximately 35,000 units). It covers the manufacturing sector only. Other sectors, such as trade and services, remain uncovered. Moreover, for small and medium enterprises, there is the problem of the missing link between PMI sample units and units from the PRODCOM sample. 2) *Foreign trade archive* (COE) integrates information about foreign trade for the totality of enterprises (cf. ISTAT 2002). It is derived from custom data and covers all the population of enterprises engaged in foreign trade (approximately 260,000). It contains the value of every product exchanged (according to the Combined Nomenclature with a detail of 8 digits) for each country of destination and origin. 3) *Technological Innovation of enterprises* survey (CIS-Community Innovation Survey) collects information on expenses for innovation projects and on the type of innovation in question. The purpose is to estimate the input and output of the innovation process that takes place in enterprises. This survey is led on a representative sample of 5.3 thousand enterprises in the 1996 and 15.5 thousand enterprises in the 2000 that are part of the population of enterprises with 20 workers or more. This is not an annual survey but is carried out every 4 years. 4) *Information and Communication Technologies (ICT)* survey tries to gather information on enterprises' use of information and communication technologies and electronic commerce, in order to highlight "new economy" activities. Enterprises with 10 or more workers in the manufacturing sector and in part of the services sector are the reference units. The representative sample contains 7,000 units. The first year of issue is 2001.

4 Second stage of integration

The structure of the database and some methodological solutions for the integration procedures were adopted to fit the information requirements of the micro-simulation models. Regarding fiscal micro-simulation (see appendix a), the general aim is to evaluate the effects of the Italian tax system on enterprises' decisions developing different modules of analysis dealing both with indirect taxes (VAT and taxes on production) and with taxes on revenues created by the enterprise (Irpeg, Irap) as well as Social Contributions, that are still an important part of labour costs. The output created for the Irap and Social Contribution modules was obtained uniting the two files containing the surveys after defining and harmonising the variables of the micro-simulation model. The fusion of the two sources with merging technique at the record level, realises a partial matching. So it is possible to acquire some information, lacking in the surveys, from the commercial accounts for those units that have been matched, as it can be seen in the chart below. At this level three main issues can be distinguished: (1) the reconciliation of survey data with administrative data, (2) the treatment of missing data and (3) the problem of sample weights. The chart shows the integration scheme of the sources described: The business register with the auxiliary variables, the statistical sources, the administrative sources, the variables involved and it focuses the three main integration issues.



VARIABLES	*Survey data* SCI	PMI	*Administrative data* C.A	Fiscal
Y - Profits and Losses	86	66	54	90
Z - Assets and Liabilities	118	15	111	69
V_1 - Staff	36	40	-	-
V_2 - Regional distribution	115	42	-	-
V_3 - Other	106	52	-	258
TOTAL	461	215	165	417

Figure 2: Integration Scheme
Reconciling Survey and Administrative Data

The harmonization of variable definitions has been, preliminarily, required in order to produce metadata information. When it has been possible, the same variables from different sources have been compared. Discrepancy analysis, for the couples of records in common between the two data sources, has the objective of picking up and quantifying possible measurement errors. The comparison of the values of the variables included in both sources was performed creating classes on the basis of the size of the percentage gap calculated in the following way: (Y_{CA}/Y_{PMI})-1. An example on two important variables, which can be used as matching variables, is illustrated in the chart below.

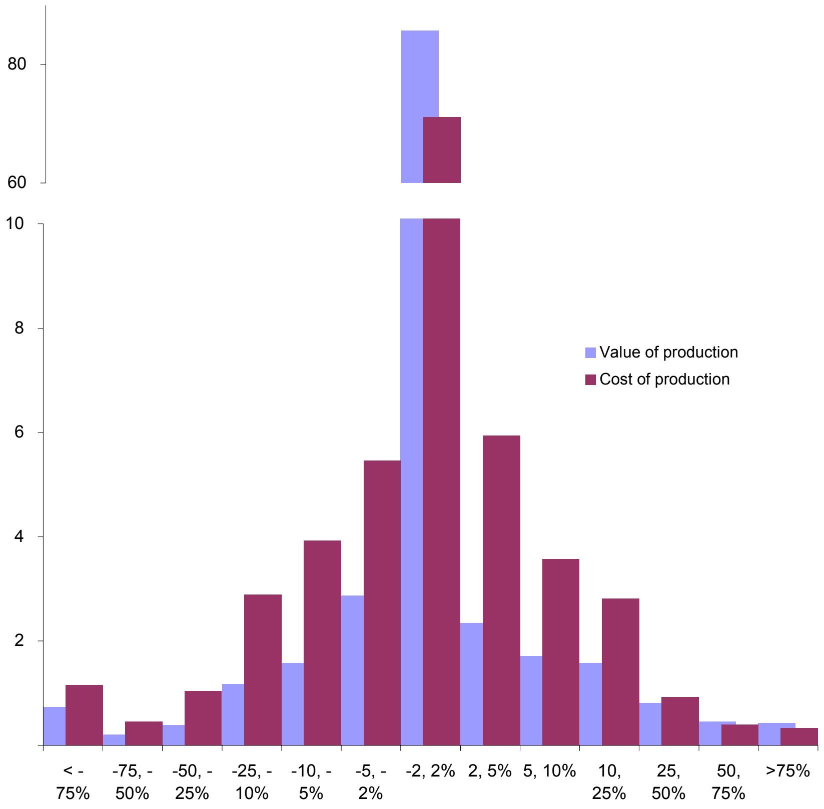


Figure 3: Discrepancy analysis

With regards to the first variable roughly 85,70 percent of PMI under observation have a discrepancy range of ±2 percent, while with regards to the second variable roughly 71 percent of the enterprises fluctuate between ±2 percent.

Some information lacking in the surveys has been reconstructed by using administrative sources. Starting from accounting variables of the surveys, and adding up new information from administrative sources, some variables has been recalculated according to compatibility rules of the balance sheet scheme. At the end the profit or loss of a single unit has been recalculated. The relative difference between the initial value and the value after the reconstruction process is shown in the table 4. Records of classes 1-4 and 18-21 have incoherent values. Considering that all incoherent units have been excluded from the analysis.

Table 4: Reconstruction of information through Administrative data

Class	Range	# records	Percentage value
1	*< -300%*	*10*	*0.1*
2	*-300, -200%*	*5*	*0.0*
3	*-200, -150%*	*7*	*0.1*
4	*-150, -100%*	*22*	*0.2*
5	-100, -75%	16	0.1

6	-75, -50%	39	0.3
7	-50, -25%	157	1.3
8	-25, -10%	392	3.3
9	-10, -5%	467	3.9
10	**-5, -2%**	**664**	**5.6**
11	**-2, 2%**	**8449**	**71.0**
12	**2, 5%**	**771**	**6.5**
13	5, 10%	473	4.0
14	10, 25%	323	2.7
15	25, 50%	76	0.6
16	50, 75%	18	0.2
17	75, 100%	8	0.1
18	*100, 150%*	*4*	*0.0*
19	*150, 200%*	*3*	*0.0*
20	*200, 300%*	*1*	*0.0*
21	*> 300%*	*1*	*0.0*

Data imputation

If the exact matching fails, the problem can be treated as a typical problem of missing data. The table below shows on which basis missing data have to be imputed: being 1999 the year of reference and taking the corporate firms of the Regional dataset (15 thousand on a total of 46 thousand firms), the 78% is linked with the information contained in the administrative data sources. For the units for which no link has been possible, 22%, information has to be reconstructed and missing data has to be imputed (dotted region in chart 2).

Table 5: Cross Sectional integration (year 1999)

Enterprises	**Regional**	**of which corporates**	**Linked with CA**	**Not linked with CA**
SME	45,867	15,330	78%	22%
Large	8,734	7,340	95%	5%
Total	54,601	22,670	83%	17%

Regarding the reconstruction of the missing data, as the mechanism that generates missing data is ignorable, that is the MAR type (missing at random), standard analyses are applied for incomplete data (R.J.A. Little and D.B. Rubin -1987) this is because the two sub-populations, that is the one with exhaustive data and the one with missing data, are not characterised by different distributions of the variables with missing data. Through the use of covariates it is possible to identify the characteristics of the units with missing data so that missing information can be recovered by considering the units with complete data. With regards to the criteria for the reconstruction of missing information two typologies of imputation techniques are considered: donor technique and a parametric model.

In the first case, imputation is performed taking the information from a unit with complete data that is similar to the unit with missing data. The similarity between donor units and host units is found on some selected matching variables on the basis of their correlation with the variables to be imputed. The concept of similarity is translated in mathematical terms as a distance function. The donor imputation based on the minimum distance among observations of the donor dataset (exact units) and the host dataset (incorrect records). The quantification of the distances of the unit with MRP (Partial Missing Response) (cf. Abbate 1997) from all the exact units is a salient aspect of the donor imputation method. The mixed distance function is defined by weighting several elementary distance indicators, each of which is calculated on the basis of the different type of the variables: qualitative variable; hierarchically classified variable; orderable-classified variable; quantitative variable. The weighting of the distances is carried out on the basis of the weights taken from the chi-square tests of independence. Chi-square tests of independence have the merit of being rather simple and determining both the variables for which there is dependence, therefore to be included in the mixed distance function, and the variables for which there is statistical independence, for which the inclusion of any simple distance indicator in the mixed distance function is irrelevant. Moreover, these tests help distinguish the stratum variable from the matching variables, i.e. between the variables that identify the large groups of similar units, within which to limit the calculation of the distances, and the variables to be included in the distance function. The chi-square values, in addition to their use for the independence tests, construct a measurement of the connection between the variable with missing values and all the other variables found on the unit. For each variable with missing values it can be observed that the higher the chi-square value is, the stronger the statistical connection between the two variables. The values contained in the table underneath result from the twin distributions constructed on the conjoined frequencies of the variables' mode with missing values (Y= total asset) and of the modes of the 4 covariates.

Table 6: Chi-square test for verifying the independence among variables

Variables		***Pearson's chi-square test***		
missing	**covariates**	**Chi-square**	**Phi Chi-square**	**Cramer contingency Index**
Y	X1	14714.8477	1.1131	0.7439
	X2	7701.8840	0.8053	0.6272
	X3	1774.5202	0.1093	0.1086
	X4	141.7592	0.3865	0.3606

On the basis of the chi-square values, for imputation of the variable Y, the following variables have been considered: X4 (business sector) as stratum variable; X1 (value of production) and X2 (total employment) as matching variables. While the variable X3 (localisation-region) has been excluded. The effectiveness of the application may be facilitated by the analysis of several control indicators. For each stratum the following are calculated: 1) the number of units with missing values; 2) the number of possible donor units; 3) a table with the actual minimum distances; 4) the number of times a single donor has been used; 5) the usage ratio.

Table 7: Control indicators to evaluate the imputed values

Stratum Variable	**Missing records**	**Donor units**	**Donor used (times)**					**Minimum Distances**			**Usage Ratio**
			1	**2**	**3**	**4-9**	**10-99**	**d=0**	**0<d<1**	**1<=d<10**	
11	1	3	1	0	0	0	0	0	1	0	0.333
14	36	148	23	5	1	0	0	0	36	0	0.243
15	139	529	94	21	1	0	0	0	139	0	0.263
16	1	1	1	0	0	0	0	0	0	1	1.000
17	147	465	85	22	6	0	0	0	147	0	0.316
18	30	130	17	5	1	0	0	0	30	0	0.231
19	15	154	15	0	0	0	0	0	15	0	0.097
20	61	154	38	7	3	0	0	0	61	0	0.396
21	28	95	13	6	1	0	0	0	28	0	0.295
22	50	221	38	6	0	0	0	0	50	0	0.226
23	51	59	22	6	3	2	0	0	51	0	0.864
24	128	381	78	19	4	0	0	0	128	0	0.336
25	31	198	25	3	0	0	0	0	31	0	0.157
26	89	423	66	7	3	0	0	0	89	0	0.210
27	91	256	61	15	0	0	0	0	91	0	0.355
28	66	420	51	6	1	0	0	0	66	0	0.157
29	88	525	74	7	0	0	0	0	88	0	0.168
30	27	54	13	1	1	2	0	0	27	0	0.500
31	59	208	36	7	3	0	0	0	59	0	0.284
32	34	86	23	1	3	0	0	0	34	0	0.395
33	64	191	37	8	1	2	0	0	64	0	0.335
34	42	91	26	5	2	0	0	0	42	0	0.462
35	48	101	27	9	1	0	0	0	48	0	0.475
36	88	263	67	9	1	0	0	0	88	0	0.335
37	24	81	21	0	1	0	0	0	24	0	0.296
40	53	93	29	4	1	3	0	0	53	0	0.570
41	19	30	12	2	1	0	0	0	19	0	0.633
45	49	375	41	4	0	0	0	0	49	0	0.131
50	121	461	72	21	1	1	0	0	121	0	0.262
51	386	1692	267	43	11	0	0	0	386	0	0.228
52	58	396	46	6	0	0	0	0	58	0	0.146
55	43	243	28	6	1	0	0	0	43	0	0.177
60	36	147	25	4	1	0	0	0	36	0	0.245
61	30	59	16	5	0	1	0	0	30	0	0.508
62	9	21	6	0	1	0	0	0	8	1	0.429
63	138	471	98	15	2	1	0	0	138	0	0.293
64	12	35	8	2	0	0	0	0	12	0	0.343
65	3	3	1	1	0	0	0	0	3	0	1.000
70	227	594	124	35	11	0	0	7	220	0	0.382
71	31	138	25	1	0	1	0	0	31	0	0.225
72	210	571	119	29	7	3	0	1	209	0	0.368
73	24	44	7	2	1	2	0	0	24	0	0.545
74	247	577	118	36	13	4	0	0	247	0	0.428
80	28	101	17	3	0	1	0	0	28	0	0.277
85	20	91	14	1	0	1	0	0	20	0	0.220
90	49	113	30	8	1	0	0	0	49	0	0.434
92	109	292	69	15	2	1	0	0	109	0	0.373
93	27	86	13	4	2	0	0	0	27	0	0.314
Tot	**3367**	**11870**	**2137**	**422**	**93**	**25**	**0**	**8**	**3357**	**2**	**0.284**

Furthermore for verifying the quality of the imputation, the dissimilarity index between the distributions of the modes of donor units and those of units with imputed data is calculated:

$$Diss.= (\sum_{j=1}^{J} \left| fo_j - fi_j \right| / 2) * 100 \tag{1}$$

where *fo* indicates the original relative frequencies present in the concealed records and *fi* stands for the relative frequencies reconstructed after the imputation process: *j* indicates the probably values of the variable. This index varies between 0 and 100, so that the closer it is to 0, the more similar the compared distributions are: on the contrary, the closer it is to value 100, the more the distributions differ.

Table 8: Distribution of exact (donor) and reconstructed (correct) records (Y=Total Assets)

	Exact records		**Correct records**	
Classes	**#**	**%**	**#**	**%**
<=4.000	2,380	20.0	671	19.9
4,000 – 1,500,000	3,622	30.5	948	28.2
1,500,000 – 4,000,000	2,956	24.9	822	24.4
4,000,000 – 10,000,000	1,825	15.4	579	17.2
>10,000,000	1,093	9.2	347	10.3
Total	**11,876**	**100.0**	**3367**	**100.0**
	Index of dissimilarity			
	Absolute	0.059		
	Relative	0.029		
	MAX	2.000		

Through the second method the value of the missing datum has been estimated by applying a parametric method based on multiple regression model. Two explanatory variables selected after the analysis of the Pearson Correlation Coefficients and two dummies define the model. This model has been tested considering alternatively a model with all the variables in a linear form and a logarithmic model with two logarithmic variables and two dummies. Statistic tests identify the best models and sensitivity analysis methodologies calculate the data quality indicators in the two different approaches adopted to 'model incomplete data' (single imputation and multiple imputation). The simple imputation techniques consist of the substitution of one single value for every missing value. Multiple imputation techniques consist in the substitution of every missing value by more acceptable values that represent the possibilities distribution (Rubin 1987). The advantage of single imputation is that it utilises standard analytical methods on complete data, even on data sets with imputed data. However, it may lead to an erroneous estimate of the sampling variability as it does not consider the variability of missing values and the uncertainty deriving from the lack of knowledge of the most appropriate non response model. Under a non response model with repeated randomised experiments, multiple imputation consents to obtain valid inferences through the reoccurrence of inferences on the complete data and it permits to consider the additional variability due to missing values.

As regards model imputation technique, the best multiple regression model are Y=f(X1,X2,X4) and Y=f(X1,X2,X3,X4), where Y=total asset, X1=value of production, X2 =total employment, X3=localisation, and X4=business sector (the continuous variables are transformed in logarithmic values).

The multiple imputation method (cf. EC-JRC, ISTAT 2003) imputes several values (M) for each missing value (from the predictive distribution of the missing data), to represent the uncertainty about which values to impute. The M versions of completed datasets are analyzed by standard complete data methods and the results are combined using simple rules to yield single combined estimates (e.g., MSE, regression coefficients), standard errors, p-values, that formally incorporate missing data uncertainty. The pooling of the results of the analyses performed on the multiply imputed datasets, implies that the resulting point estimates are averaged over the M completed sample points, and the resulting standard errors and p-values are adjusted according to the variance of the corresponding M completed sample point estimates. Thus, the "*between imputation variance*", provides a measure of the extra inferential uncertainty due to missing data (which is not reflected in single imputation). The results obtained after 50 imputations for the variable Y (total assets) are shown in table 9. The relative increase in variance (r) due to the multiple imputations is very small, indicating that the statistical uncertainty due to missing data, likewise (variation across the imputed datasets) is small.

Table 9: Estimates of Y (complete data) obtained after m=50 imputations
(Confidence level for interval estimates is 95%)

	Estimate	**BI Variance**	**WI Variance**	**Total Variance**	**Std Error**	**Low End-Point**	**High End-Point**	**r**
Y	6.5725	1.3653(10^{-6})	0.7918	0.7918	0.8898	4.8284	8.3165	1.7589(10^{-6})

Next, the results considering the regression coefficients are exposed in table 10. In this way the robustness of the imputed datasets is explored. Some definitions: T-ratio is defined as the estimate divided by its standard error (appropriate for testing the null hypothesis that the quantity is equal to zero), *df* shows degrees of freedom for Student's *t* approximation, and p-value is for testing the null hypothesis that the quantity is equal to zero, against the two-sided alternative hypothesis that it is not zero.

Table 10: Linear regression coefficients obtained after m=50 imputations
(Confidence level for interval estimates is 95%)

	Estimate	**Standard error**	**t-ratio**	**Df**	**p-value**	**Low End-Pnt**	**High End-Pnt**	**r**
B0	1.83695	0.0319513	57.49	767	0.0000	1.77423	1.89967	0.3380
B1	0.65585	0.0061107	107.33	795	0.0000	0.643853	0.667843	0.3302
B2	0.26916	0.0070641	38.10	1154	0.0000	0.255299	0.283019	0.2595
B3	*-0.01214*	*0.0067513*	*-1.80*	*1949*	*0.0722*	*-0.0253843*	*0.00109661*	*0.1884*
B4	0.06546	0.0054209	12.08	3033	0.0000	0.0548306	0.0760884	0.1456

A comparison of the *t-ratio* and the *t-distribution* for *B3* reveals that the null hypothesis (*B3*=0) is true. So the regressor X3 (localisation) does not contribute to the explanation of "missing-ness", in other words, with respect to X3 observation values are missing at random while for the remaining factors, they do not.

Weight adjustment

For the estimation of the variables of the microsimulation model it is needed to employ weight adjustment techniques. In fact the units of the matrix are a subset of a sample (SME) in which weights have been constructed on the basis of the sampling and estimation strategy of the survey that does not consider in the stratification the legal form of the units and as a consequence does not take it into account in the system of weights to be associated with the considered statistical units either. In general the adjustments of weights can be effected using calibration methods. The calibration methods consist in adapting the estimated sample distribution of some auxiliary variables to the distribution of the same variables known from outside sources.

The auxiliary variables usually used for sample stratification are the business sector (NACE, four digits), 21 geographical areas (Regions) and 5 employment classes. In this case, there is the need to add a new dummy variable, corporate/non corporate firm. The technique that is used is the Generalized Estimation Method (Falorsi et al. 1995), where calibrated estimators are applied. Weights are adjusted through the minimization of a distance function between the initial and final weights. The distance function is subject to boundary conditions in fact, for reason of consistency, correction of weights has to obey the constraints of number of firms and workers as registered in administrative registers (cf. Estevao, Hidiroglou, Sarndal). First results are produced by considering the dummy variable "corp" (0 = non corporate; 1 = corporate) and one only constraint, i.e. the total number of enterprises. At this stage, a first re-weighting procedure has been run, in order to correctly re-calculate sample weights, so that the sum of corporate weights is equal to the number of corporate enterprises. When there are a higher number of constraints, the problem of constrained optimization has to be managed with the GENESEES software (cf. Istat 2002d).

5 Conclusions

The development of economic analysis at micro level (micro-simulation, impact analysis of policies, systemic analysis and indicator construction) can be carried out by joining two or more different sources in order to obtain one single database with more information. The integration process has encounters various problems of different nature. The attainment of economic objectives is strongly conditioned by the availability and quality of information, as well as by the sample and non-sample nature of the involved data sources. Even the choice of integration methodologies to be used depends on the above mentioned factors. From a methodological point of view the statistical analysis, the data treatment, the application of different integration techniques, the estimate of parameters and the investigation of statistical quality indicators for multi-source databases are complex phases of the integration process. For instance, for the completion of the whole integration process of the Irpeg module, partially analysed in this work, it is also needed to reconstruct integrated data-set whit longitudinal information. At a micro level, while for the SCI survey a reconstruction of longitudinal information is possible, for the PMI survey, since there is not a panel sample, it is not possible. For the latter the reconstruction of longitudinal information relative to some variables is possible by constructing a panel of statistical units with commercial accounts data available for the previous year. Furthermore the couples of records in common between the PMI survey and the commercial accounts for the two years can be used to calculate adjustment factors. Finally for the estimation of the variables needed in the microsimulation module and to calibrate microsimulation results, weight adjustment techniques have been employed. In fact the units of the matrix are a subset of a sample (PMI) in which weights have been constructed on the basis of the sampling and estimation strategy of the survey that does not consider in the stratification the legal form of the units and as a consequence does not take it into account in the system of weights to be associated with the considered statistical units either. If, from one or more sources, information sufficiently accurate to be used as a benchmark is available on the values of the parameter to be estimated, than sample data can be used to obtain estimates of the quantities of interest, ensuring that the inferential procedure returns results that are close to the reference value.

Analysis in this paper has tried to describe the main steps and methodological issues of the integration process and to stress its complexity. The process of integration is still under construction which range from census and survey to administrative (including fiscal) data. In this instance, the integration of all available information on enterprises into one multi-sources database realised in the DIECOFIS project gives to the system high potentialities and opportunities in terms of economic analysis. It enables to solve the data requirements of the micro-simulation model for the estimation of the effects of fiscal policies on enterprises performance but also more economic issues can be investigated and analysed in a very detailed perspective through new demand driven methods. The use of a systematised and integrated system makes it possible to create new micro founded indicators that are more appropriate to describe different economic systems and to understand their systemic strength and weakness.

Appendix: Integrated Database for the support Policy Impact Analysis

The final result of the integration process is the integrated overall dataset, which is representative of the universe of enterprises. Data marts are extracted from this database to serve fiscal microsimulation analysis and to produce systemic analyses. The model exposed in the chart has three modules: C.T. (corporate tax), S.C. (social contribution), R.T. (regional tax); and produces tax estimates and indicators for economic analysis.

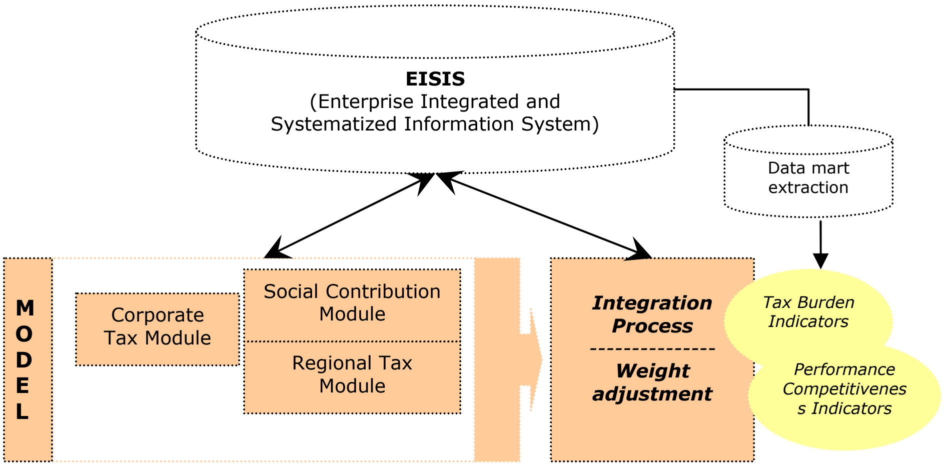


Figure 1a: Overall Scheme

From the integrated data base the model should estimate the taxable yield for every type of tax. Even if ISTAT surveys do not cover all data needs for a precise calculation of taxable incomes, of the voices of deductions of taxable income and of deduction of the tax, hypothesis on the behaviour of the enterprise have been made.

In the DIECOFIS Microsimulation Model there are three modules that calculate three types of taxes. It is structured to simulate hypothetic scenarios of different variation of tax legislation. The three modules are the Social Contribution Module, The Regional Tax Module and the Corporate Tax module. The Social Contributions module calculates the contribution for social security according to professional category (manager, employees, etc.) and type of contract (formation, collaborations, etc.). The calculation of the Regional Tax (IRAP) on Added Value is carried out in the following way:

IRAP = tr (VP - CP - TO). Where: tr represents the share proportion of the regional tax. VP = Value of Production: Income from sales, variations of stocks, other income. CP = Costs of Production: Raw materials and consumables, other external charges, value adjustments, amortisation. *TO* = Other Deductions: INAIL (National Institute of Insurance Against Accidents at Work) Contributions, apprentices' costs, formation contract jobs and costs for disabled persons. The Corporate Tax module calculates the tax burden on corporate enterprises. The equation of Italian corporate tax is the following: $CT = t_g(U + T_{IND.} + Cr + Crd - PP)$. Where: t_g is the legal rate, U is the profit, T_{IND} is the amount of non deductible taxes, Cr tax credits in profit and loss scheme, Crd Tax credit to share dividend and PP the amount of loss brought forward [5].

Appendix: The Software for multi-source data integration

The central task in the Diecofis system is the Multi-source data integration. Here, starting from a generic and extensible way to realise the multi-source data integration, the concept for the SAS environment is exposed. Another central task is the selection of data for the purpose of a micro simulation. The resulting Diecofis Data Mart now serves a micro simulation as data basis and its structure is flexible in order to support the structure of the simulation .

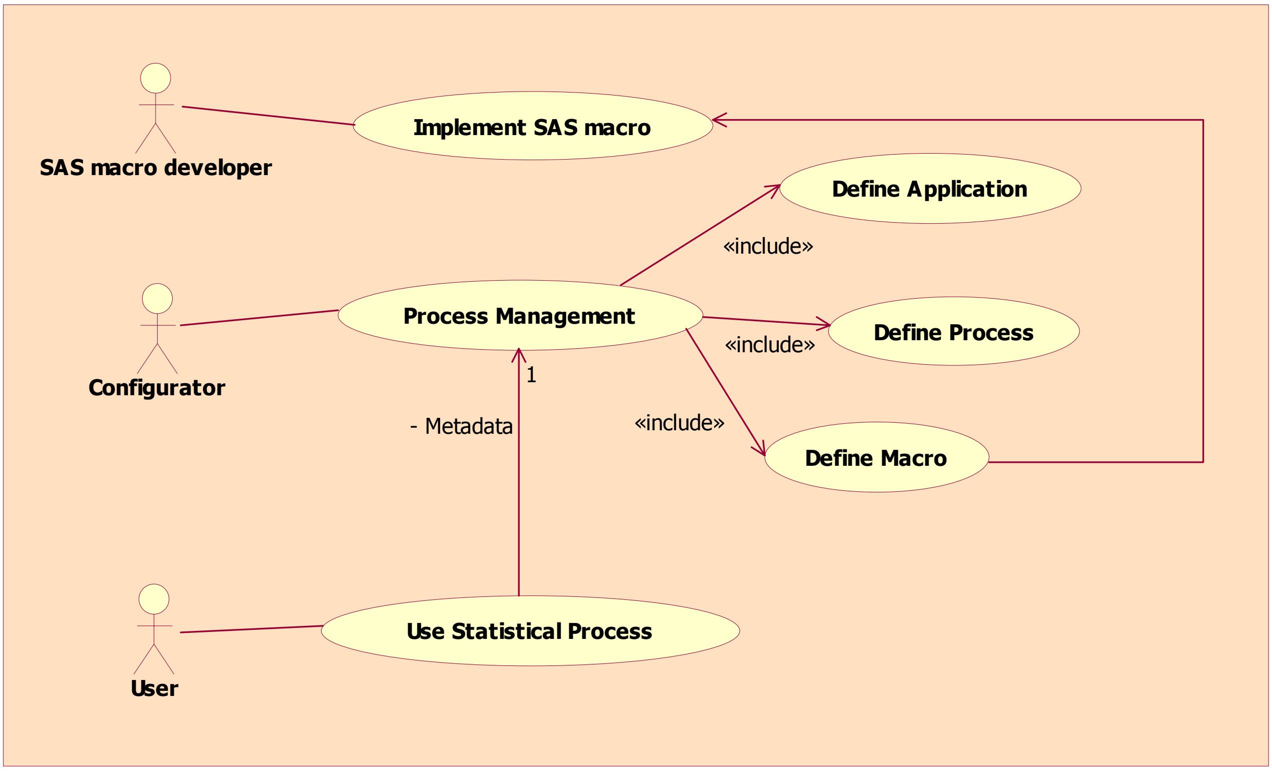


Figure 1b: Diecofis Conceptual Scheme

[5] For further details see Bardazzi, R., F., Pazienza, M.G., Parisi, V. (2003), “The Effects of the Italian Tax Reform on Corporations: a Microsimulation Approach”. http://www.istat.it/diecofis.

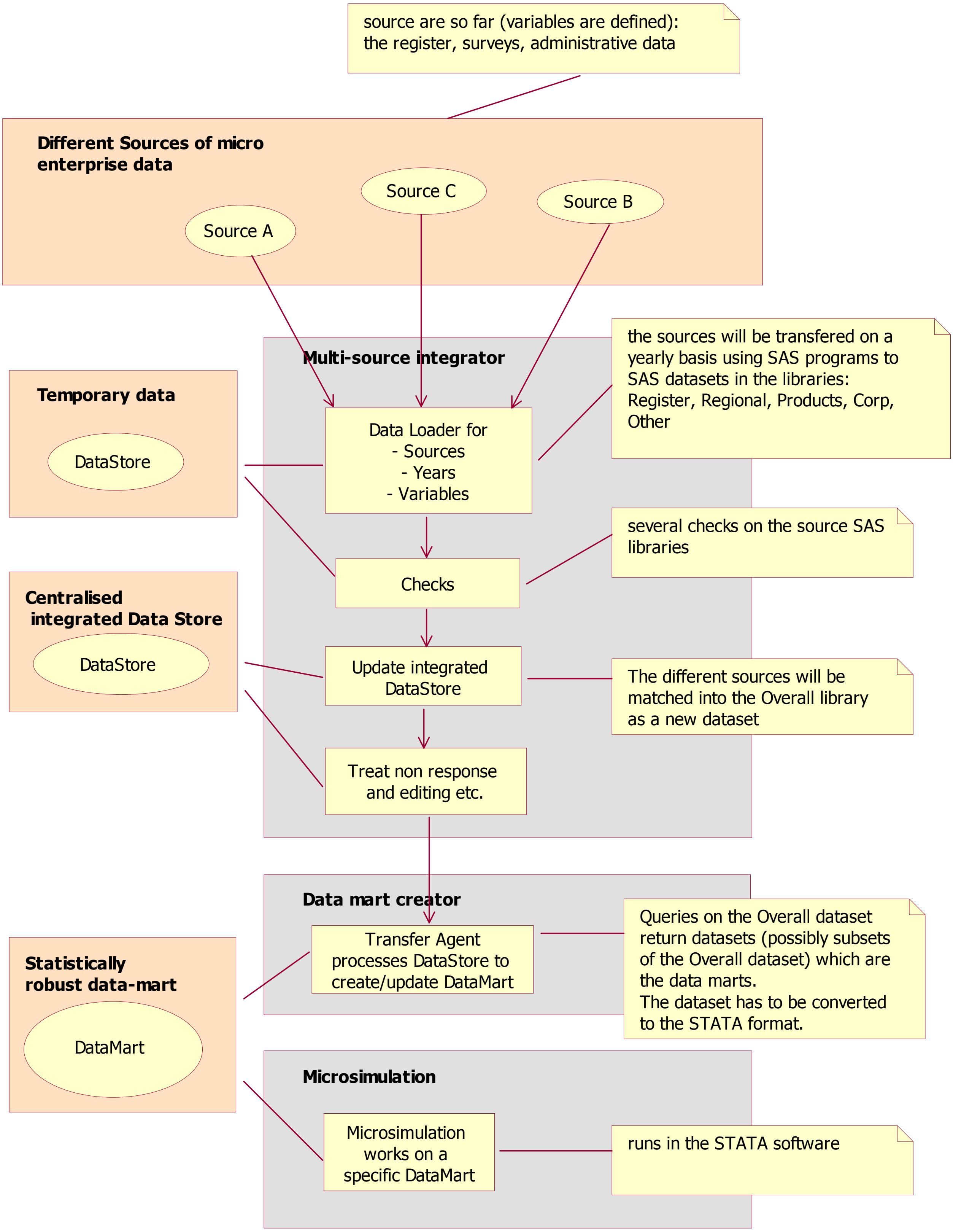
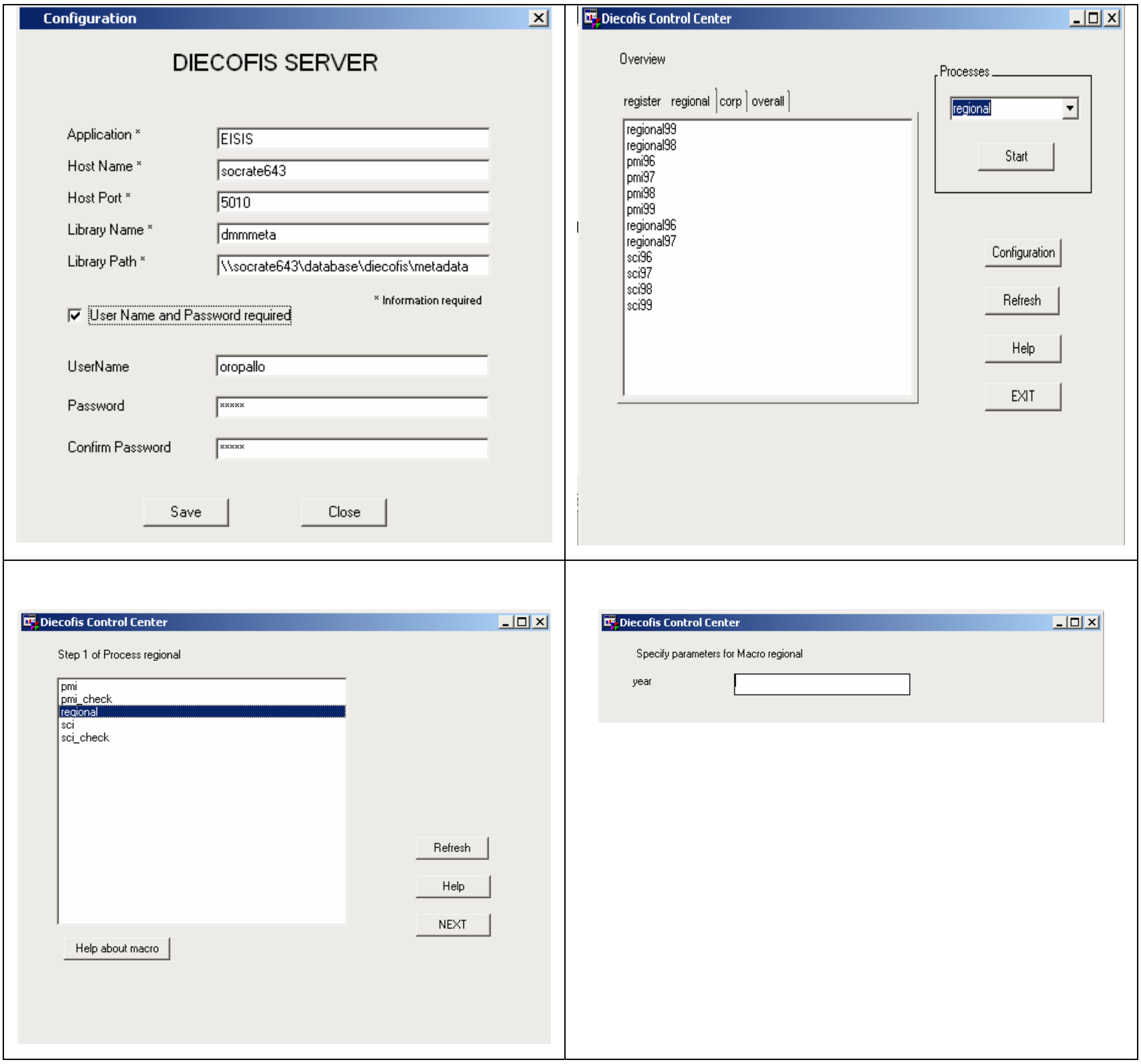

source are so far (variables are defined):
the register, surveys, administrative data
Different Sources of micro enterprise data
Source C
Source B
Source A
Multi-source integrator
the sources will be transfered on a yearly basis using SAS programs to SAS datasets in the libraries: Register, Regional, Products, Corp, Other
Temporary data
DataStore
Data Loader for
- Sources
- Years
- Variables
several checks on the source SAS libraries
Checks
Centralised integrated Data Store
DataStore
Update integrated DataStore
The different sources will be matched into the Overall library as a new dataset
Treat non response and editing etc.
Data mart creator
Transfer Agent processes DataStore to create/update DataMart
Queries on the Overall dataset return datasets (possibly subsets of the Overall dataset) which are the data marts.
The dataset has to be converted to the STATA format.
Statistically robust data-mart
DataMart
Microsimulation
Microsimulation works on a specific DataMart
runs in the STATA software

Figure 2b: Diecofis IT formalization of the Integration process

The Diecofis User Interface for the multi source data integration and data mart creation has been realised inside the SAS Software allowing the user to interact with the SAS macros that have been developed on a user-friendly and transparent way (cf. Informer SA 2003a, 2003b). The flexibility and extensibility of the Diecofis system with regard to new source integration and new micro simulation extension is guaranteed.
All SAS Macros have been implemented in the Diecofis Software and all the specific processes have run under the Software. The chart below shows the Configuration window and the Control Centre window. If a Process is selected then in the next step one relative macro has to be selected with the opportune parameter.



Configuration

DIECOFIS SERVER

Application * EISIS
Host Name * socrate643
Host Port * 5010
Library Name * dmmmeta
Library Path * \\socrate643\database\diecofis\metadata

* Information required

☑ User Name and Password required

UserName oropallo
Password *****
Confirm Password *****

Diecofis Control Center

Overview

register regional corp overall

regional99
regional98
pmi96
pmi97
pmi98
pmi99
regional96
regional97
sci96
sci97
sci98
sci99

Processes: regional

Diecofis Control Center

Step 1 of Process regional

pmi
pmi_check
regional
sci
sci_check

Diecofis Control Center

Specify parameters for Macro regional

year

Figure 3b: Diecofis Interface

SAS Macros developed have been adapted to be used in the framework of the final release of the Diecofis Software. SAS macros are divided in 4 macro types. They refer to 4 kinds of processes and 4 types of datasets:

1. *register*
2. *regional*
3. *corp*
4. *overall*

The *register* macro loads and transforms the register data.

The *regional* macros load, check and merge survey data:

- pmi.sas (This macro loads all statistical information on Structural SME Surveys).
- sci.sas (This macro loads all statistical information on Large Enterprise Surveys).
- pmi_check.sas and sci_check.sas (These two macros perform check on survey data).
- regional.sas (This macro refers to the merging activities regarding survey data).

The *corp* macros load and check administrative data:

- corp.sas (This macro loads commercial accounts data).
- corp_check.sas (This macro performs check on administrative data).

The overall macros perform the final integration:

- overall.sas (This macro merges survey and administrative data).
- overall_check.sas, overall_imp.sas, overall_imp2.sas (These macros regard the imputation activity and the final check).

The microsimulation model run on STATA software and uses the overall dataset opportunely stored in STATA format by using the STATA TRANSFER software.
The folder is the follow:
The input datasets are: overall"year"
The output datasets are: final"year"
The final datasets contains fiscal variables of simulated scenarios. This final dataset is stored also in SAS format and is used to produce tax burden and performance indicators.
The macros on indicators are the following:

- indicators.sas
- tax_analysis.sas

They contain the calculation of decomposable indicators and fiscal burden rates.

References

C. Abbate. Completeness of Information and Imputation from Donor with Minimum Mixed Distance*, Quaderni di Ricerca ISTAT*, n. 4/1997, pp. 68-102. 1997

R.F. Bardazzi, M.G. Pazienza, V. Parisi. *The Effects of the Italian Tax Reform on Corporations: a Microsimulation Approach*.http://www.istat.it/diecofis. 2003

J. Black. Changes in Sampling Units in Surveys of Businesses. In: *2001 FCSM Research Conference Papers*, US Census Bureau. 2001

J.M. Brick and G. Kalton. Handling Missing Data in Survey Research, *Statistical Methods in Medical Research*, vol. 5, pp. 215-238. 1996.

G. Brackstone. Managing Data Quality in a Statistical Agency, *Survey Methodology*, December 1999, vol. 25, pp. 139-149. 1999.

M. Denk and F. Oropallo. *Overview of the Issues in Longitudinal and Cross-Sectional Multi-Source Databases*. http://www.istat.it/diecofis/deliverable_list.htm. 2002.

M. Denk, F. Inglese and M.G. Calza. *Assessment of different approaches for the integration of ample surveys*, http://www.istat.it/diecofis/deliverable_list.htm. 2003

J.C. Deville and C.E. Särndal C. E. Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, vol. 87, pp. 376-382. 1992.

EC-JRC, ISTAT - Software analysis - Development of methodologies and of a software for the measurement of statistical quality, and for comparing the robustness of alternative multi-source, integrated databases - DIECOFIS http://www.istat.it/diecofis/deliverable_list.htm. 2003.

Eurostat. Use of Administrative Sources for Business Statistic Purposes: Handbook on Good Practices – Theme 4 (Industry, Trade and Services), Eurostat Edition. 1999.

Eurostat. Model quality report in business statistics, Eurostat Working Group "Assessment of quality in business statistics". 1999.

P.D. Falorsi and S. Falorsi. *Un metodo di stima generalizzato per le indagini sulle famiglie e sulle imprese (A Generalized Estimation Method for Surveys of Families and Firms)*. Quaderni CON PRI, University of Bologna. 1995.

FCSM – Federal Committee on Statistical Methodology. *Report on Exact and Statistical Matching Techniques, Statistical Policy Working Paper 5*, Washington, DC: U.S. Department of Commerce. 1980.

E. Giovannini and A. Sorce. Integration of Statistical (survey) data with registers (administrative) data. *Paper contributed to the Meeting on the Management of Statistical Information Technology*. 2001.

W.H. Inmon. – Data Marts and Data Warehouse: Information Architecture for the Millenium – Informix Corporation.

ISTAT. L’innovazione tecnologica nelle imprese (Firms’ Technological Innovation) – Note Rapide – July 1999 (LE Survey). http://www.istat.it/Imprese/Ricerca-e-/index.htm.

ISTAT. I risultati economici delle medio-grandi imprese Anni 1998-99 (Economic Outcomes of Medium-Large Size Enterprises) - Statistiche in breve - July 2000 (LE Survey) - http://www.istat.it/Imprese/Struttura-/index.htm. 2000.

ISTAT. Indagine Prodcom (prodcom Survey) – Indagine sulla struttura dei Costi (Cost structure Survey) http://www.istat.it/Imprese-e-/index.htm. 2001a.

ISTAT. Struttura e competitività del sistema delle imprese industriali e dei servizi nel 1998 (Structure and competitiveness of industrial and service enterprise system in 1998). - Statistiche in breve - Luglio 2001 (LE & PMI Survey) http://www.istat.it/Imprese/Struttura-/index.htm. 2001b.

ISTAT. Indagine sul Commercio Estero (Foreign Trade Survey). Current version available at http://www.coeweb.istat.it/. 2002a.

ISTAT. L’uso delle tecnologie dell’informazione e della comunicazione nelle imprese (The use of ICT in Italian firms) – Statistiche in breve http://www.istat.it/ Imprese/Ricerca-e-/index.htm. 2002b.

ISTAT. CONCORD v1.0 - (Generalized Data Editing Software) SOFTWARE GENERALIZZATO PER IL CONTROLLO E LA CORREZIONE DEI DATI RILEVATI NELLE INDAGINI STATISTICHE – MPS – ISTAT 2002 - http://www.istat.it/Metodologi/index.htm. 2002c.

ISTAT. GENESEES v1.0 - (GENEralised software for Sampling Estimates and Errors in Surveys) SOFTWARE PER IL CALCOLO DELLE STIME E DEGLI ERRORI CAMPIONARI– MPS – ISTAT 2002 - http://www.istat.it/Metodologi/ index.htm. 2002d.

T.B. Jabine and F.J. Scheuren. Record Linkages for Statistical Purposes: Methodological Issues. *Journal of Official Statistics* 2 (3), 255-277. 1986.

J.B. Kadane. Some Statistical Problems in Merging Data Files. In: 1978 Compendium of Tax Research, US Dept. of the Treasury, 159–171. Reprinted in *Journal of Official Statistics* 17 (3), 423–433. 1978.

G. Kalton and D. Kasprzik. The treatment of missing survey data, *Survey Methodology*, vol. 12, n. 1, pp. 1-16. 1986

G. Kalton and D. Kasprzik. *Imputing for missing survey responses, Proceedings of the Section on Survey Research Methods*, American Statistical Association. 1982.

W.A. Kamakura and M. Wedel. Statistical Data Fusion for Cross-Tabulation, *Journal of Marketing Research*, vol. 34, pp. 485-498. 1997.

J.G. Kovar and P.J. Whitridge. Imputation of Business Survey Data. In: Cox et al. (eds.), *Business Survey Methods*, New York: J. Wiley. 1995.

H.J. Lenz. Multi-Data Sources and Data Fusion. In: *Proc. New Techniques and Technologies for Statistics (NTTS) 1998*, EUROSTAT, 139–146. 1998.

R.J.A. Little and D.B. Rubin. *Statistical Analysis with Missing Data*, Wiley & Sons, New York . 1987.

F.M. Malvestuto. Data Integration in Statistical Databases. In: Michalewicz (ed.), *Statistical and Scientific Databases*, Chichester: Ellis Horwood, 201–232. 1991.

F. Oropallo and D. Skalbania. Concept of IT framework issues and development of software for the creation of a multi-source data base http://www.istat.it/diecofis/deliverable_list.htm. 2003.

G. Paass. Statistical Record Linkage Methodology: State of the Art and Future Prospects. In: *Proc. International Statistical Institute*, 45th Session, Amsterdam. 1985.

G. Paass. Statistical match: Evaluation of existing procedures and improvements by using additional information. G.H. Orcutt and H. Quinke (eds). *Microanalytic Simulation Models to Support Social and Financial Policy*. Amsterdam: Elsevier Science, pp. 401-422. 1986.

P. Roberti, F. Oropallo, F. Inglese, L. Lo Cascio and G. de Martinis G. Towards a Systemic Analysis of Italian Industrial Texture Review, *Industria* 4/2002, Il Mulino. 2003.

R.H. Renssen. Use of Statistical Matching Techniques in Calibration Estimation, *Survey Methodology*, vol. 24, n. 2, pp. 171-183.1998.

W.L. Rodgers. An Evaluation of Statistical Matching. *Journal of Business and Economic Statistics* 2, 91-102. 1984.

D.B. Rubin. *Multiple Imputation for Non-response in Surveys*, Wiley & Sons, New York. 1987.

N. Ruggles and R. Ruggles. A Strategy for Merging and Matching Microdata Sets. *Annals of Economic and Social Measurement* 3 (2), 353–372. 1974.

F. Scheuren and W.E. Winkler. Regression Analysis of Data Files that are Computer Matched II. Survey Methodology 23, 157–165. 1997.

J.L. Schafer and M.K. Olsen. Multiple imputation for multivariate missing-data problems: a data analyst's perspective, *Multivariate Behavioural Research*, vol. 33, pp. 545-571. 1998.

A.C. Singh, C.A. Mohl. Understanding Calibration Estimators in Survey Sampling", *Survey Methodology*, vol. 22, n.2, pp. 107-115. 1996.

A.C. Singh, H.J. Mantel, M.D. Kinack and G. Rowe. Statistical Matching: Use of Auxiliary Information as an Alternative to the Conditional Independence Assumption, *Survey Methodology*, June 1993- vol. 19, No.1 pp. 59-79 - Statistics Canada. 1993.

W.E. Winkler. Matching and Record Linkage. In B. G. Cox et al. (ed.), *Business Survey Methods*, Wiley & Sons, New York, pp. 920-935 (355-384). 1995.

Author's adresse

Filippo Oropallo
Istat - Istituto Nazionale di Statistica
Via Magenta, 2
00185 - Roma
Italy

Tel. +39 06 46733632
Fax +39 06 46733706
Elec. Mail: oropallo@istat.it
http://www.istat.it