Società italiana di economia pubblica
XV CONFERENZA

DIRITTI, REGOLE, MERCATO
Economia pubblica ed analisi economica del diritto

Pavia, Università, 3 - 4 ottobre 2003

INTERNATIONAL RESEARCH INTO USING MICRO-DATABASES TO IMPROVE POLICY ANALYSIS LESSONS FROM THE DIECOFIS AND EUROKY-PIA RESEARCH PROJECTS

PAOLO ROBERTI*, RICHARD EASON°, FILIPPO OROPALLO*

*Italian Statistical Office

°Inland Revenue, UK,

Provisional do not quote

The views express in this paper are those of the authors. They cannot and should not be attributed to their respective employers.

DIRITTI, REGOLE, MERCATO
Economia pubblica ed analisi economica del diritto

XV Conferenza SIEP - Pavia, Università, 3 - 4 ottobre 2003

pubblicazione internet realizzata con contributo della

COMPAGNIA di San Paolo

società italiana di economia pubblica

dipartimento di economia pubblica e territoriale – università di Pavia

ABSTRACT

The authors report on two EU funded projects involving various government departments and academic institutes. "Diecofis" assesses the issues facing modelers in developing multi-sourced micro-databases of the business sector in Italy and the UK to improve the measurement and analysis of fiscal and economic issues, particularly those relating to competitiveness, fiscal stability, and comparative taxation. Euroky-Pia addresses the issue of establishing knowledge databases to support policy impact analysis across all areas of government. The paper focuses on the high level issues such as the assessment of user needs, co-ordination of data from administrative and statistical sources, the roles and constraints affecting the organizations contributing to analysis, and the key differences in national circumstances. Different modeling approaches are addressed.

SUMMARY

INTRODUCTION

Paraphrasing the title of a famous book, “no policy is an island”. Ultimately, the impact of a public policy hinges on a host of factors ranging from program design, ways and means, program administration and, substantially on the underlying socio-economic environment. In other words, policies have to be shaped and tailored around the problems that they are designed to tackle, and with a sound understanding of the underlying socio-economic milieu.

Social policy analysts recognized this tenet long ago. Accordingly, good analysts have dovetailed macro, meso and micro research within an integrated and multidimensional and topical analytical framework. Economic policy analysts, instead, have by and large continued to focus on macro and sectorial aggregates and relationships. Micro-analysis has been limited by the databases available. Mapping systemic economic change onto micro-systems remains for the most part an under researched areas.

The social indicator adventure of the late 1960s and 1970s established the limits of summary descriptive statistics for policy impact analysis (PIA). Since then social PIA has made advances which at the time were difficult to imagine. This is witnessed by the swelling availability and extensive use of households’ micro data and microsimulation (static and dynamic) models, and by the development of a host of micro founded (summary and decomposable) indicators. The tools developed for the analysis of inequality and poverty, and their high degree of sophistication provide notable examples of the developments that have occurred.

Now, the focal point of social PIA has moved on from ratios and indicators per se to ratios and indicators that assist in the unravelling of the complex and multi faceted social policy conundrum. This involves dealing with matters relating to severity, relativities, delivery options, competing claims and choice between programs which can differently impact on *persons* and groups.

In other words, the limelight has turned away from aggregate snapshots to distributional analysis within an integrated analytical framework that acknowledges interrelationships (such as between different policies) as well as the existence of multifarious other elements[1]. In the end, sets of “purpose oriented” and “scope fulfilling” indicators have replaced the long lists of crude, area-specific average ratios popular in the 1970s and 1980s. In parallel, and to facilate these developments, national statistical institutes have tailored the supply of social statistics to social policy analysts’ and social PIA’s demands. There is now the increasingly widespread practice of collecting data through multitopic and longitudinal household surveys[2] and the increasingly

[1] Including first and second order effects, changes in individual behaviour, take-up, enforcement and compliance hitches, such as insufficient information and lack of knowledge, fraud and sheer error.

[2] See, among others, Grosh M. and Glewwe P. (eds), **Designing Household Survey Questionnaires for Developping Countries. Lessons from 15 Years od the Living Standards Measurement Study**, The World Bank, Washington D.C., 2000.

unproblematic access to micro households datasets (such as the LIS-Luxembourg Income Study data set).

But no similar development has occurred in the field of economic analysis. Here, the object of analysis remains mostly macro and meso (sectoral and territorial) data. Amazingly, enterprise micro data are still quite difficult to access. Moreover, enterprise data collection is "divided" and uncoordinated. Use of multitopical surveys is not widespread and the ensuing fragmentation of available information compels analysts to research in an "environment with many blind spots, where crucial information may be seen only dimly or not at all". Each survey is shaped by single topic or dimension. Even when datasets can be linked and merged together easily, they normally remain separate. And, when data sets are integrated, it is more often for validation rather than research purposes. In practice, there is a sizeable wedge between the information that is potentially available and the one that is actually accessible. Micro economic research is hindered and research opportunities are foregone. For instance, if economic growth is buoyant and income grows, we draw detailed charts to map income distribution. These reveal the gainers and the losers, and whether income differentials have fallen or widened. No similar information is accessible for the income production side. Individual enterprises can seldom, if ever, be ranked according to their performance. Important questions remain unanswered, such as: has growth been good because performance has improved for all firms, or only for some? Whose performance has increased most? Have performance gaps widened or shrank? What are the profile and traits of best performers?

In practice, social statistics have progressed and have increasingly been customized to tailor users' needs. Instead, developments in economic statistics have lagged behind. Sheer lack of information and little access to micro data are the culprits. The information wedge has hindered the development of serious microeconomic analysis of systemic and policy aspects, such as incentives, drivers, "differential effects" and patterns. This has also hindered the development of micro economic models for enterprise analyses and policy, or greatly reduced their scope. When they have developed, they have had to be based, with very few exceptions, on a small fraction of the information needed or potentially available.

Against this background, it has appeared important to gather research consensus on the scope for the development of enterprise analytical tools, including microsimulation models and the information to support them, including integrated and systematized information systems (EISIS). To this effect, Istat in Italy and Inland Revenue in the UK joined forces and to took the lead to develop two EU FP IST projects, **Diecofis** and **Euroky-pia**, to generate critical research mass to test and pilot problems and methods, develop prototypes and search for the best ways that can eventually lead to fill an increasingly patent gap in economic policy analysis. Istat started from the perspective of a national statistics office with wide ranging data collection responsibilities while the Inland Revenue began from the perspective of a department with policy responsibilities and some data of its own from the

administration of its policies. Moreover, national differences in organisation, law, culture, and practice might also be important factors.

This paper draws lessons from experience from both projects. It has three parts. Part 1 provides a description of the two projects, their background, objectives and accomplishments so far. Part 2 discusses in some detail some main projects results. Finally, Part 3 highlights the lessons that can be drawn and directions for future research.

I. ENTERPRISES INTEGRATED AND SYSTEMATIZED INFORMATION SYSTEMS AND MICROSIMULATION MODELS TO IMPROVE POLICY ANALYSIS IN THE EUROPEAN UNION

a. DIECOFIS[3]: Background, Main Objectives, Overview and Achievements

Background Tax policy remains a critical issue in the EU policy agenda. Member countries have not yet reached a stage at which they can comfortably move in unison. Nor have they developed all the instruments and PIA knowledge needed to properly inform debates.

In spite of extensive discussions, experts' and working groups and a crowd of reports, the "facts" on the impact of taxation on businesses are scanty. Those available have a high degree of approximation and are not necessarily informative. Tax indicators have well-known pitfalls and drawbacks. Clearly, they do not catch the great diversity and wide dispersion (much greater than for individuals and families) which is observed in actual ratios, nor their sensitivity to different "real drivers", that is to choices that can lead companies to pay more or less tax as a proportion of their profits or performance. Analysis of the specimen average wage earner across counties is simple and helpful, but analysis of the average company is neither.

Understanding how taxes affect economic performance, and developing better indicators to gauge their impact - especially in the area of corporate taxation - is central to endow the EU with a set of efficient and fair tax policies. To achieve this, a system of micro-founded indicators based on factuals and counterfactuals estimated through micro-simulation models, is proposed under DIECOFIS. The project is committed to pave the way for the development of national and EU-wide micro-simulation models of enterprise taxation by exploring the problems and issues that need to be tackled to build up needed knowledge and capacity for tax-PIA**.**

Main Objectives The general objective of DIECOFIS is to foster the development of "best practice" policy analysis and evaluation techniques in the field of taxation, to further the Lisbon objectives and EU governance. To this effect, the projects intends (i) to develop an integrated and systematized

[3] DIECOFIS, ***Development of a System of Indicators of Competitiveness and Fiscal Impact on Enterprises Performance***, EU IST 2000-31125, www.istat.it/diecofis

statistical information system[4] of enterprise data that effectively supports policy-making; (ii) to develop national micro-simulation models for the analysis of the impact of taxation on competitiveness and enterprise performance in both static and dynamic contexts, in two European countries, Italy and the UK; (iii) to learn lessons and portray a "common demonstrator" in view of the subsequent expansion of the activity to the whole EU area; (iv) to develop systemic maps and indicators – elementary as well as composite – that permit to capture links with economic performance and with ICT induced changes.

Synoptic Project Overview and Achievements DIECOFIS has offered an opportunity to work together in a trans-national, multi-disciplinary and networked research environment to government officials, statistical services, academics and researchers in five EU member countries[5], in view of developing knowledge and tools for enterprise tax policy and performance analysis.

The project is shaped around four major thematic areas (see Box 1). During its one and a half year's life span, it has (i) pilot tested the development of an EISIS in Italy; a UK companion corporate tax model for Italy covering three major taxes on enterprises (corporate tax, Regional tax on economic activities and social security contributions); composite and decomposable performance indicators; (ii) as well as helped to get a better grasp on methodological and research issues; (iii) charted the road for developing an EU demonstrator and for replicating similar models in the other EU member countries.

Presently, the project is at a stage at which substantial progress can be claimed on both the Italian and UK sides. However, the road ahead is still very long and uphill. Admittedly, work is still in its early stages.

When DIECOFIS started, the UK was far more advanced than Italy in the development and use of microsimulation business models. Tax analysis and research within the Inland Revenue was well established. Microsimulation UK models were already operating on data drawn from the administration of tax and household surveys since the 1970's [6]. These were the main sources of distributional analysis to support policy formulation.

BOX 1 DIECOFIS MAIN RESEARCH THEMATIC AREAS

DIECOFIS is shaped around four major thematic areas. **Theme 1** embraces the issues that fall within the upper section of Diagram 1. They include:

- The systematisation of single data sources into an integrated database, and related data quality and validation issues;
- The development of appropriate statistical and IT tools to integrate and select the data needed for micro-simulation purposes; and
- Sensitivity analysis.

[4] Including data files from economic, tax and social insurance census, survey and administrative sources.

[5] Consortium members include: ISTAT (co-ordinator), Inland Revenue, Informer SA, European Commission Joint Research Centre, CERES Centre Economic and Social Research, University of Cambridge, London School of Economics, University of Florence, Wirtschaftsuniversitaet Wien, University of Tor Vergata.

[6] See 1, Eason 1993.

Once these activities are completed, all the variables required for micro-simulation purposes will be available and stored in a single "hub". This is expected to open up new vistas such as exploring the linkages between policies and performance, developing micro-founded systemic indicators, and permitting access to micro-data through new demand driven methods.

Theme 2 embraces the issues that fall within the central section of Diagram 1. Basically, these concern the development of micro-simulation models and tools for policy impact analysis -in the circumstances, static tax models. Behavioural issues, however, cannot and are not overlooked. At this early stages however, only some tax evasion and avoidance issues are scheduled for modelling.

Theme 3 embraces the issues that fall within the bottom section of Diagram 1. It includes the study of the properties, and the development of indicators, composite as well as elementary, that can be used to describe the links between the dynamics of performance and competitiveness and policy impact. The majority of existing macro indicators are inadequate for capturing the complexities arising from globalisation and technological change, and how these are reflected in growth patterns. The use of systematized and integrated statistical information systems makes it possible to create new micro-founded indicators that are more appropriate to describe different economic systems and their features and to understand their systemic strength and weakness. In turn linking this information with micro-economic counter-factuals can be expected to permit to grasp how policies can hinder or foster growth prospects.

Theme 4 embraces issues that are closely related to those mentioned under Theme 3 above. It deals with the issues associated with the development of an "EU demonstrator". It presents the lessons that can be gathered from the work carried out under the three preceding themes and the knowledge that is needed for tackling the task of developing a pan-European tax micro-simulation model. Diagram 2 presents a visual illustration of how the activities on the development of the two national models will be harnessed to extract the knowledge on the "core features" (as opposed to country specific features) upon which the pan-European model can be built.

Among the various models, the main one was the corporation tax forecasting model[7] This only had a narrow dataset for use in policy analysis. Moreover, the capability in respect of modelling the taxation of the business sector was insufficient to meet the many and varied questions that arise concerning the direct and behavioural impact of taxation on businesses and the wider impacts on business performance and the economy. To progress on these, a new team was established in 2002 to examine the need for a new more comprehensive business model to support policy analysis. In the circumstances, the team welcomed the opportunity to participate in the Diecofis project, both to learn from the experience of other participators and to investigate international issues such as the comparison of corporate tax systems.

A key area of interest and primary task for the new team was to improve understanding of two main topics:

- *"drivers" of business performance and tax payments*, that provide a sound understanding of the relevant economic and business relationships which link aspects such as the macro-economic situation with commercial business performance and business performance with the level of tax liability, taking into account the nature and complexities of the tax system. Based on previous research, establishing and

[7] See 2, Eason 1997.

modelling these relationships was deemed paramount to improve model accuracy, since any policy analysis requires some estimate of the impact of changing tax policy,
- *the availability of relevant micro-data*, including developing a critical assessment of aspects such as reliability, definitions, timeliness, availability, and access.

On the Italian side, analysis and research was in its very early stages. Moreover, the capacity for policy analysis that existed concerned households. In effect, DIECOFIS began with a "vision", a strong interest, a wealth of enterprises data and "pixie dust" sparkled by the British Embassy in Rome. This successfully connected Italian and British researchers and academics. Austria, also believed and joined this endeavour.

After a year and a half DIECOFIS is a fact. Microsimulation models exist for corporate taxation, the regional tax on economic activity, and social security contributions. Progress in developing the EISIS is particularly noteworthy. Istat has successfully integrated and systematized five primary and secondary statistical sources of enterprise data that are routinely collected. Subsequently, statistics from three administrative sources, containing commercial accounts, tax and social security data, have been added. All data, and the macrosimulation models have been validated. Sensitivity analysis is under way.

New vistas have gratifyingly opened up in linking taxation and economic performance and in broadening the number of variables in the EISIS. Further, microsimulation models are being used (i) to simulate the proposed reform of business taxation in Italy; (ii) to study interactions between tax systems and enterprises "systemic" characteristics; and (iii) to analyse the determinants of tax ratios (how much of the difference that is observed between the tax ratios of two countries, notably the UK and Italy, is due to differences in rates, tax base definition, deductions and so on.

In a nutshell, progress has occurred and can be summarized in five points. DIECOFIS has made it possible to:
- Microsimulate existing legislation and "what if" hypotheses;
- Study systemic features, that is factors of strength and weakness of different "economic textures";
- Develop fine-grained "maps" and decomposable/composable indicators;
- Monitor and benchmark performance with systemic indicators shaped and tailored to suit purpose;
- Explore relationships between taxation, drivers and performance.

Results and problems are discussed in a more structured way below. On the whole, research activities have progressed along two main tracks. The UK's, which has proved to be the most robust, when the focus is on tax modelling and tax analysis (impact and forecast). And the Italian track, which has shown to be the most robust when the analytical framework embraces wider and interrelated economic, fiscal and tax issues.

Setting aside the obvious dissimilarities in aims and purposes, the main difference between the two approaches is in the database and their broader/narrower coverage. In the case of the former, analysis can be undertaken from data that are directly available from, in the UK instance, the tax authority's detailed information. A drawback is that the models that can be developed, and the scope of the analysis are "restricted" within narrow limits by the data that are accessible. These are functions-dependent, while databases are purpose built. These features shape and hamper the research that can be supported. They also preclude the study of the broader context. In the Italian approach, the data come from the integration of different sources, which range from census and survey to administrative (including fiscal) data. In this instance, the available information has a "higher potential", is "richer" and much more flexible. Many more issues can be investigated and analyses can stretch from aggregate to very detailed, depending on scope, tools and the availability of software that can be used to "*slice", "dice", "drill up", "drill down", "drill through"* and *"drill across"* the information *hyper* and *microcubes* depending on users' needs

b. EUROKY-PIA[8]: Background and Main Objectives

Background The "vision" behind EUROKY-PIA stems from DIECOFIS. When the latter was thought about, its focus was deliberately narrowed on taxation and, specifically, corporate taxation and other taxes on enterprises. Wide country coverage appeared undesirable. Moreover, it was believed that it was desirable to limit the scope of the project to exploring, testing, piloting and mapping opportunities and ways ahead, before embarking in an ambitious, full EU scale endeavour aimed at developing knowledge and capacity for policy analysis across all areas of government.

As work on DIECOFIS has advancing, it has right away appeared desirable to enlarge the project's perspective and broaden its purpose, scope and coverage due to obvious complementarities and economies of scale.

At the beginning of the 2000s most EU countries appeared to lag significantly behind the USA in policy analysis, though demand for PIA was clearly picking up. PIA knowledge and capacity was needed to support EU policy makers and policymaking at the "federal", national and regional level[9]. To achieve this, it appeared desirable, along the lines of EUROMOD for social PIA, to build support and to foster the highest level of EU-wide cooperation. A EU network was formed and funding was obtained under the EU-IST FP5 to explore issues, carry out feasibility studies, prepare an agenda, a plan and a roadmap.

Main Objectives The end objectives of the project are two. The first is to shape a "vision" and a "roadmap" for developing "Policy Analysis Knowledge" that can serve to prepare the ground for a future FP6 project that can support

[8] EUROKY-PIA, ***Developing European Knowledge for Policy Impact Analysis***, EU IST 2002-38704.

[9] This is witnessed by the EU ongoing effort to develop PIA knowledge and capacity.

emerging national and Eu policy issues, as well as policy coherence and effectiveness.

The second is to involve in a common effort, within a framework of excellence and EU wide cooperation, all different actors, including (i) *National Statistical Institutes*, that collect, hold and provide high-quality information; (ii) *IT enterprises and communities*, that can support NSIs' transition efforts from "hard-statistics and "hard-databases" to *e*-statistics, *e*-data bases and "virtual" wide-ranging data sets resulting from the integration and systematization of information from administrative and statistical sources, by means of appropriate procedures that permit to allow for differences in data quality (e.g. imprecision and errors); (iii) *Academia and research institutions* (national and international), that have a key role in the development of the best methods, tools and models for policy impact analysis; and (iv) *government departments and other organizations of the civil society* that need and require policy impact evaluation to evaluate proposals and choose, support and enact those that are shown to be the most cost-effective and fair.

It is envisaged that within this common framework that the project hopes to "mould", all the actors or "policy agents" shall be enabled to interact. And, thus doing they shall be able to appreciate their different perspectives and understand how best they can complement each other to make policy impact analysis "demand" and "supply" meet.

EUROKY-PIA has just taken off. At this stage activity has centred on a review of the issues and of the status of PIA in different areas and countries. Meanwhile, the core group[10] of the network has begun identifying the EU priorities and policy areas on which a PIA investment will have to focus at the launch, as well as defining a medium-term strategy (plan and sequence) in view of presenting a EU FP6 integrated project.

II. DIECOFIS MAIN ACHIEVEMENTS AND RESULTS

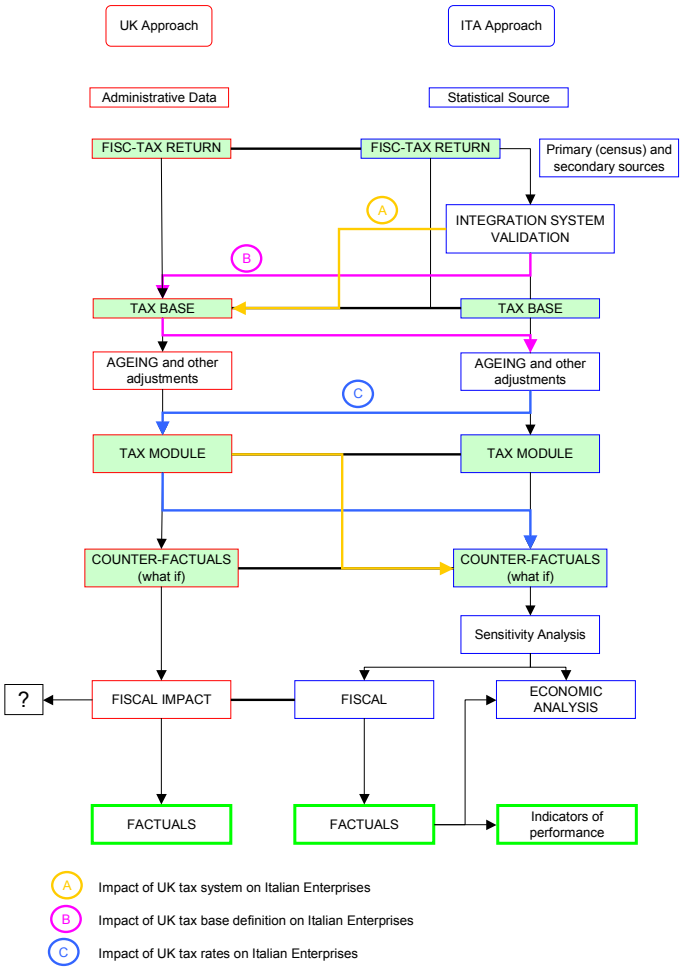
DIECOFIS has fulfilled expectations and permitted to make important steps forward. It has shown great potentiality for PIA. It has displayed the importance and need of investing into the development of EISIS, and the additional value added that National Statistical Offices can create if they put users' needs and demands in the limelight of their mission. As seen above, targets have been met and results have gone beyond consortium members' expectations (such as in the case of purpose-oriented, scope-fulfilling and well-behaved "systemic indicators").

The EISIS and microsimulation models are a fact. The former is the result of, firstly, the integration and systematisation of five primary and secondary enterprises statistical sources which are routinely collected within the EU. And, subsequently, of four administrative sources containing data on commercial accounts, tax returns, social security contributions and custom records. All

[10] Consortium members include: ISTAT (co-ordinator), Inland Revenue, Informer SA, European Commission Joint Research Centre, Mantos UK, Global Insight, University of Tor Vergata.

data have been validated. Sensitivity analysis are underway. Tax modelling work has also been completed and validated. Currently, it is being applied to simulate the new government proposal on the reform of business taxation. A further, on going activity includes "cross-country simulations", that is the simulation of the tax burden imposed on one country (i.e. Italy) enterprises if another country (i.e. the UK) legislation were to be introduced. To this effect, three variants are being simulated and studied. These include, equalization limited to statutory tax rates only (*Variant A*) or to tax base rules only (*Variant B*); and full substitution (*Variant C*). Eventually, this exercise will permit to explain how much of the differences observed between the UK and Italy can be imputed to differences in legislation (and specific provisions) or to structural and behavioural factors. Chart 1 illustrates the various tax modelling activities envisaged under DIECOFIS.

Chart 1. Cross-Country Simulations



Progress under DIECOFIS has included:

a. Business modelling

As witnessed by the few models that are available, business tax microsimulation is tricky and demanding, even when static (tax-calculator) models are used. These may appear fairly straightforward and uncomplicated.

Underneath their apparent simplicity, however, intricate issues are invariably hidden. It is not infrequent, that two apparently similar tax units bear quite different tax burdens. This happens because economic and demographic similarity of enterprises is not necessarily a good proxy for tax liability similarity. In the latter there are more and different dimensions than in the former. Moreover, small, often unremarkable details can make a big difference. If situations are heterogeneous, even more so are the "drivers" that link commercial and business performance with tax liability. Complexity and heterogeneity make business tax microsimulation intricate and "data thirsty", even more so than individuals and families microsimulation.

Unlike the latter, the modelling of behaviours may not only be necessary, but unavoidable to ensure a high degree of analytical accuracy, such as in the case of large multinational groups and small enterprises. In practice, the business universe is fragmented and unalike. Hence, tax models have to be shaped to mirror this fragmentation. This implies, as indicated by the UK experience, that to make empirical research manageable, the development of knowledge for microsimulation purposes is always best combined with the expected users' needs, before the structure of the relevant business model can be determined and modelled. Given that different modelling approaches and different data sources can be used for microsimulation purposes – typically, tax assessments for companies and unincorporated businesses, balance sheets and survey data– the statistical information that can be accessed for modelling purposes may be more or less rich and effective depending on the extent and quality of the integration process of the data that are available from different sources. In general, the greater, the level and quality of the process of integration, the richer and more effective the database available to support tax modelling and analysis can be expected to be.

Against, this background, DIECOFIS has permitted to establish that enterprises data integration is a concern that, first, does not figures high in NSIs' business plans, let alone their agenda and priorities. Second, it is seldom carried out. When it is, it is functional to data production, since it is largely used for validation and remedial purposes. Database fragmentation is still the norm in most countries, which suggests that opportunities are unduly lost by not putting users' needs and demands in the limelight of statistical activities.

An indication that has come from the joint IT-UK DIECOFIS experience is that in transplanting experience or reproducing approaches, a first necessary step, before engaging in the development of microsimulation models, is to take stock on data availability and, on that basis, define a logical model. The research work carried out by the UK hints that, unless you are a national statistical office, it is always desirable to structure the thinking in terms of a business modelling environment, rather than a single database and model; and on the broad area of outputs that a business modelling environment might aim to produce. From this can emerge a better understanding of the data and modelling requirements needed to support policy analysis of the business sector.

Chart 2 helps in navigating the “seas” of possible modelling choices and in mapping best courses once the “tonnage of the boat“, that is the information which is available, is known. Each “sea” is of different “depth” and opens up to different levels and type of analysis:

Chart 2. A Framework for Business Modelling

Level of analysis	**Outputs**	**Type of Model**	**Micro-data sources**
High	Economic indicators and analysis	Determinants of productivity, profitability and competitiveness	• Economic survey data • Consolidated accounts • Other sample surveys of business
Mixed	Fiscal indicators	Quantifying determinants of tax performances and UK effective tax rates	• IR tax assessments & computations (company level) • IR and City ‘knowledge’ • UK and global consolidated group accounts
Low	Specific tax analysis	Estimates of cost of specific tax changes and distributional/behavioural impact	• IR tax details • Details from accounts • Ad-hoc inputs

"*High*” level analysis, which tends to be based on a limited but relatively high level dataset. It might use macro-data on the economy and micro-data on variables such as turnover, profitability, investment, overall tax accrual, but the typically one micro-data source has been used and it has lacked the depth necessary to explore many of the issues of interest. Such analysis may produce statistically significant results but the knowledge gained is limited and the results may be spurious because of the impact of a factors not isolated in the analysis.

“*Low*” level analysis, which uses data directly related to the issue being studied but the narrow limits of the database prevent analysis of the broader context. Much of the current microanalysis on specific tax proposals is undertaken directly from the tax authority’s detailed information and thus the reasons for the changes in income or tax allowances cannot be studied.

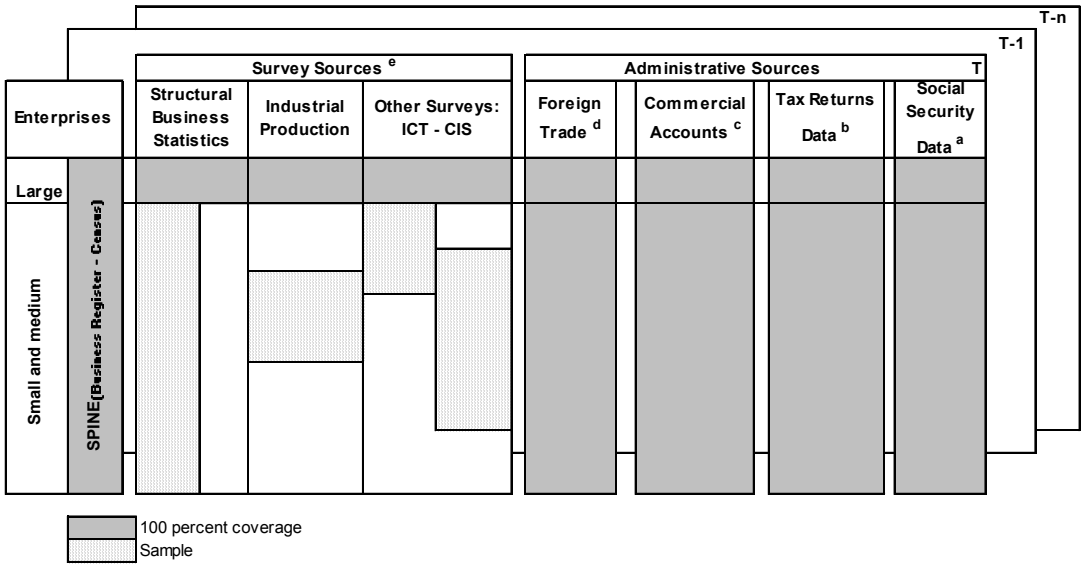
In between, lie intermediate levels of analysis. It is on these richer "mixed" or multi-sourced database that analysis nowadays wants to focus research. Many major issues concerning detailed tax policies and wider fiscal issues can only be addressed through the collation of micro-data from different sources. In the UK, the main area of investigation that led to this conclusion was the study of the Effective (Corporation)Tax Rates (*ETRs*) of major UK companies. But consideration of tax policy for small companies has also led to the need to collate data on the companies with that on their owners.

If access to data is less constrained and problematic, as it has been in the case of ISTAT, information “*oceans*” can be navigated, even though the boats that are available have not been built for this. Information “o*ceans*” are very deep and rich. They open up amazing research vistas. This is the conclusion to which has lead research in Italy. When DIECOFIS began all pieces needed to put together the EISIS jigsaw where available. The successful integration of

census, survey and administrative data has lead to prove that (i) putting the puzzle together is a viable job (though, with methodological challenges that may well have to be solved in better ways in the future); and (ii) data integration can best be carried out by NSIs since they "hub", and are best equipped to retrieve a great quantity of data and to by-pass information-access barriers. EISES give maximum choice and minimize restrictions. They allow to travel the whole spectrum of available information, backwards and forwards, and to link "large", "low" and "very fine" scale levels of research.

All in all, Diecofis has unveiled a new vision for enterprises statistics. In this vision, enterprises data are not handled as "fragments" of a universe, but are assembled as pieces of a mosaic. The model behind this vision is characterized by a "spine" (the businesses register) and is pigeon-holed (survey and administrative data). The spine is constituted by a list of unique "identifiers" ordered according to some criteria. Each identifier singles out a tier (i.e. an enterprise). On each tier, next to each identifier/enterprise there are many horizontally-lined boxes. In each box are arranged statistics, one variable per box. Within each box, statistics are sequenced according to their age, starting from the most recent. When a data is missing, this is signalled by an empty space. If estimates are available for missing variables these may be included or they may be calculated with appropriate methods (and imputed) to fill gaps. The end result is an hypercube which contains all available information. This cube can be *drilled* in all directions: *up, down, across* and so on. What is needed is a *drill* that is a software. Given this configuration, the research potential associated with the development of an EISIS, such as the one built by ISTAT, is augmented to its the highest degree (though various methodological issues have not yet been tackled).

Chart 3. EISIS Hypercube

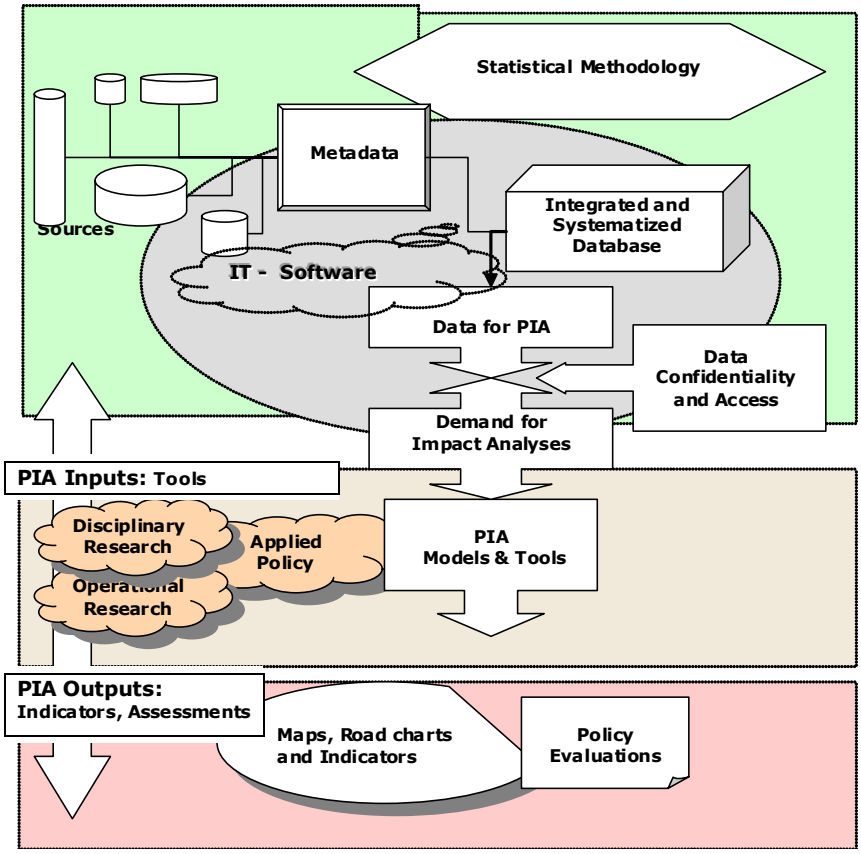


(a) Enterprise with employees only - This database is shaped by the National Social Security Legislation
(b) All enterprises - This database is shaped by the National Tax Legislation
(c) Incorporated enterprises only (Their account system is regulated by UE directives)
(d) Exporting enterprises only
(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;

From a research and analytical viewpoint, the latter proves to be by far superior in terms of potential and opportunities. In particular, it allows to approach microsimulation in ways that are very much model, users' needs and demand driven. All that is needed to "connect" models and data is a good and flexible software. With the traditional approach, instead, models can but be driven by the data, that is supply determined.

Eventually, DIECOFIS has permitted to propose a new integrated vision and approach to statistics for PIA and, specifically, for microsimulation. This vision has lead to its adoption as corner stone for EUROKY-PIA (see Chart 4).

Chart 4. Project Areas and Information, Tools and Methods for Policy Impact Analysis (PIA)



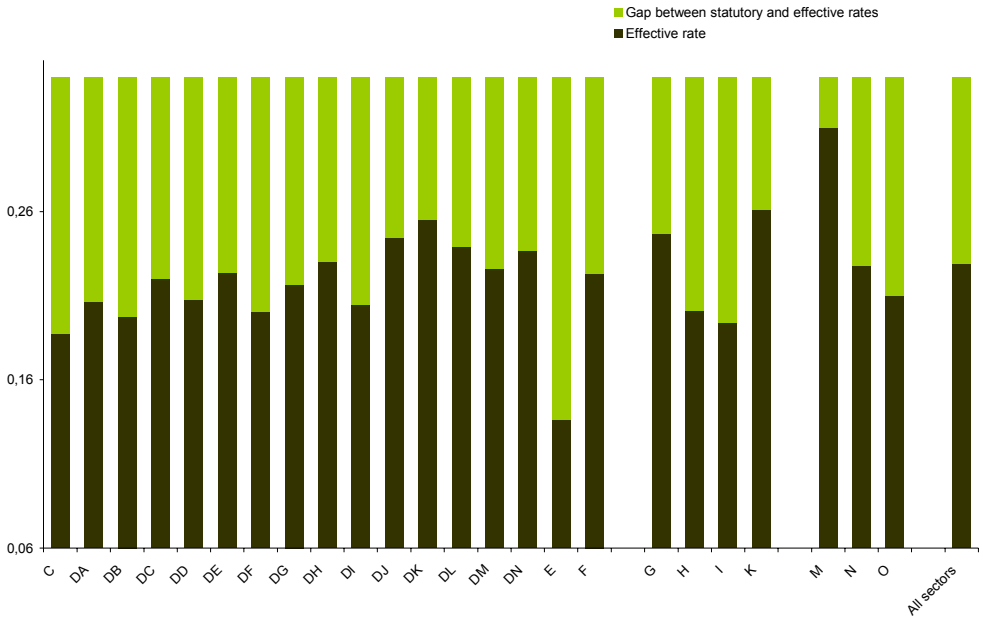
b. Tax Modelling and Analysis

Moving to tax incidence analysis, DIECOFIS has permitted to inject a substantial dose of realism in it.

Statutory and Effective Tax Rates As Graph 1 makes clear, statutory tax rates are not a very good proxy for tax incidence analysis, both within and across countries (see Graph 1). The gap between statutory and effective tax rates can be not only large, but have a pattern characterized by widespread differences in tax incidence across groups and areas, even when rates are

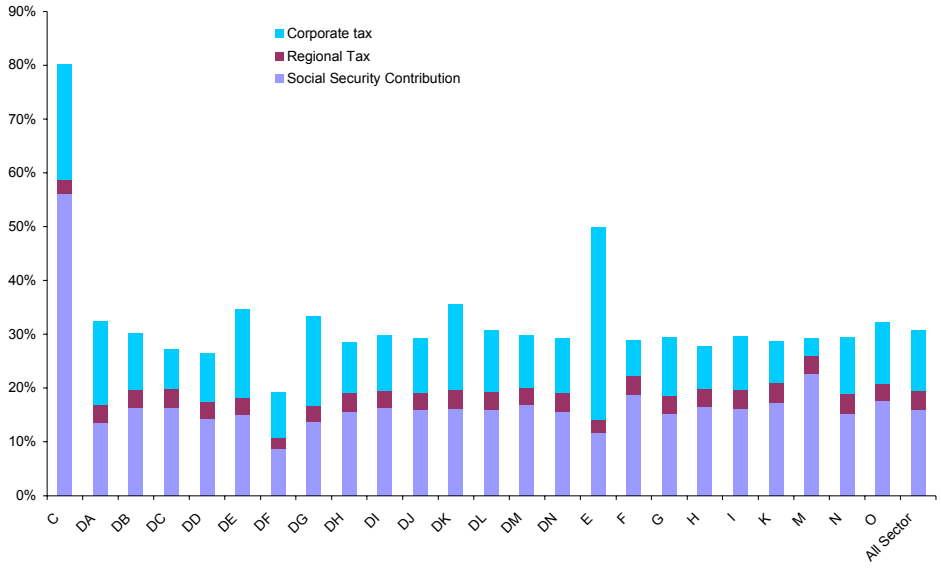
uniform and proportional. Investigating the importance and the factors behind these differences is essential, if analyses and comparisons are to be meaningful and revealing. Understanding these aspects is vital especially for tax convergence and harmonization policies.

Graph 1 - Statutory and Effective Rates by NACE Sectors in Italy



DIECOFIS has made it possible to shed light on these differences. The empirical results seem to suggest that the significance of cross-country comparisons on tax incidence can be of limited relevance, if based on summary indicators of tax burdens. What they show may not only be distorting but flawed.

Graph 2 – Total Tax Takes by NACE Sectors in Italy[11]
(as a percentage of Total Value Added)



[11] Include the Corporate income tax, the regional business tax and employers' social security contribution

Total Tax Takes Another problem with traditional summary tax ratios is that they do not necessarily depict the full tax picture. Governments, can make choices and use different instruments to collect their revenue. Whenever more than one tax instrument is used (e.g. in Italy, the corporate tax, the Regional IRAP tax and employer's social security contributions) or different instruments are used to tax different type of enterprises, looking at one tax dimension only, leads to biased comparisons. Microsimulation models permit to overcome these shortcomings as they make it possible to add up different taxes and to estimate total *net* tax takes. The ensuing tax ratios, as shown in Graph 2 that presents a picture that is by no means similar to the one shown in Graph 1.

Tax Incidence and Large Companies In the UK, a relatively small number of large multi-national company groups comprise a large proportion of business activity, employment, profit, and tax. The same applies to large (even if not necessarily multi-national) companies in Italy. Although they comprise less than one percent of all enterprises, they employ just less than one forth of all workers and produce just about 38 percent of the total value added. Thus, in any study for business tax policy coverage of large groups in general, and multinationals in particular is very important. Recently, Inland Revenue has been able to have for the first time investigated the Effective Corporation Tax Rates paid by the largest 200 or so groups, measured as a proportion of their estimated UK accounting profit. In some countries this would be a straightforward and regularly undertaken part of compliance activity to check whether tax is falling as a proportion on reported profit. In the UK, it is more difficult for four main reasons:

- tax is charged on each company's profits and not at company group level whereas groups report their performance at group level;
- most groups have substantial non-UK activity which generates little UK tax but the segmental details in the accounts are limited;
- accounting details for each member of a company group are not readily available for analytical purposes in the Inland Revenue;
- most groups pay their corporation tax by quarterly instalments under group payment arrangements, but the amount paid may be rather different from the amount finally agreed when all the details for all the group members are settled.

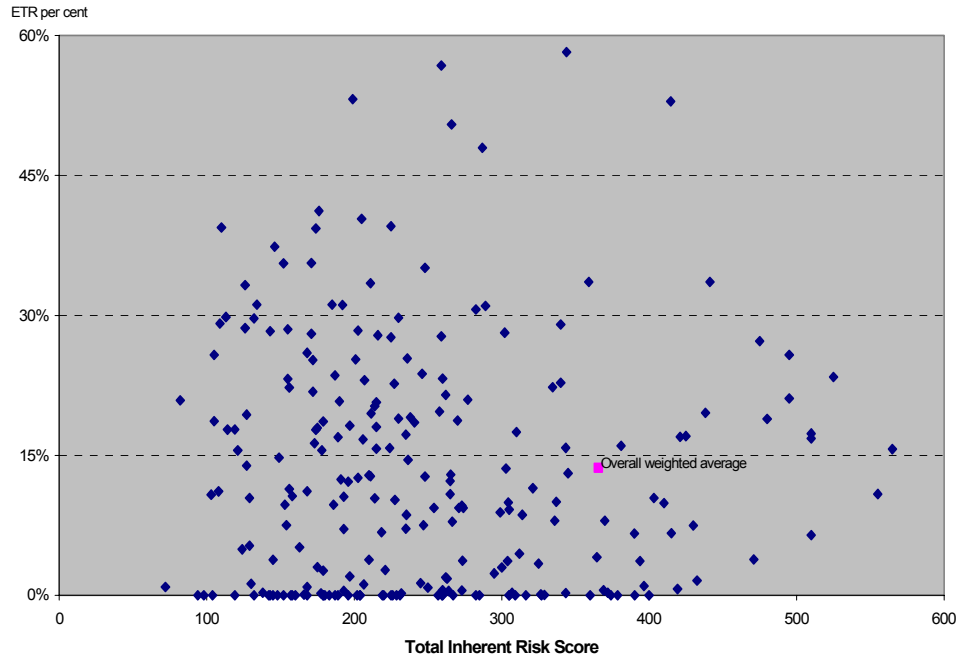
Nonetheless estimates were produced for the largest companies. Compared with the main rate of tax of 30 percent, the analysis of *ETRs* by their Inherent Risk Scores (an indicator of complexity and compliance risk[12]) showed huge variation, see graph 3, and by industrial sector wide variation was again found see graph4.

Despite the inevitable inaccuracies in estimation the wide dispersions proved the need for a rich and broad database to begin to understand the reasons why some companies pay little or no tax while others pay far more as a proportion of their profits. What were the real "drivers" of this diverse position?

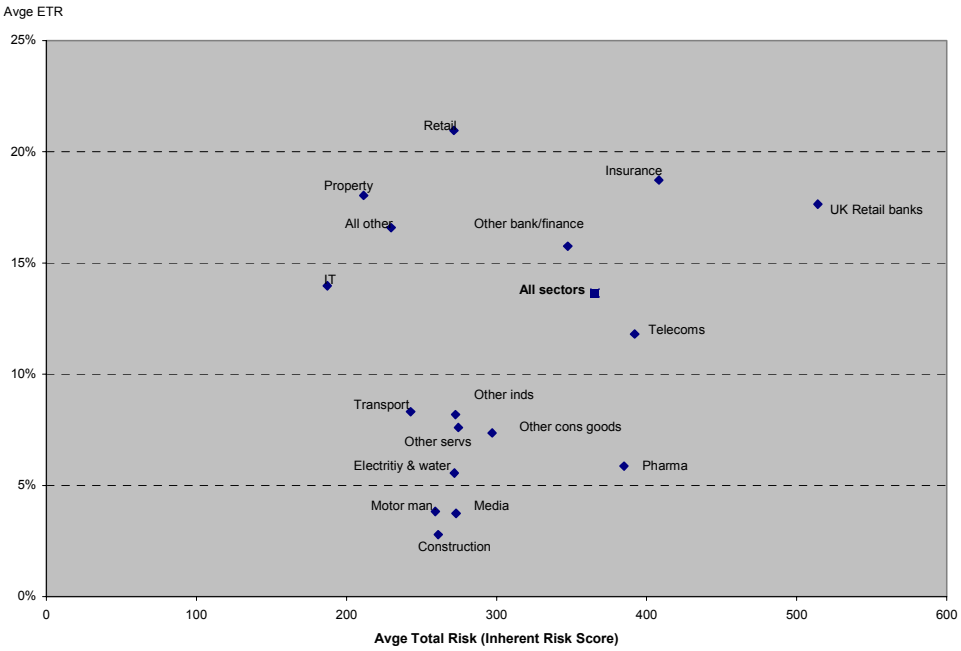
[12] See Yeend, Eason (2003)

Study so far has been limited. However, some conclusions can be safely drawn. Just as business is diverse in its activity, its financing, and its performance, so are the drivers of performance and tax levels. At one extreme, a large bank may reduce its tax because of an increase in bad debt provision or an unsuccessful venture into e-banking while a pharmaceutical company's profits may fall (or surge) as a result of the expiry of patents on important drugs. Substantial changes in corporate structure, such as mergers and acquisitions, have major impacts on the taxation of the companies concerned
and the profits of the financial institutions advising on the changes.

Graph 3 – Effective Tax Rates for Total Companies, UK



Graph 4 – Effective Tax Rates for Top Companies, UK



Tax Incidence, Small Companies and Individual Entrepreneurs

Small companies are by definition rather simpler than large multi-national groups. However their successful modelling for tax policy purposes does involve factors that are not immediately obvious. A small company is usually owned by an individual or just a few people (maybe two or three other family members or just a business partner). These often also work directly for that company. In this position the tax of the company is closely related to the tax of the individuals. Income of the company can be paid as remuneration to the owners (and family members), reported as company profit and then paid as dividends, or profits may be retained in the company and dispersed to the owners later by various means. National insurance, loans, pensions, and share schemes can also affect the owners remuneration strategy. Hence successful modelling of small companies requires collation of details about the owners and their finances.

In principle, the taxation of small companies and individual entrepreneurs may well be considered an issues that belongs to the area of individual income taxation. Their circumstances, however, are quite different, which would suggest their investigation and the development of appropriate modelling techniques. For instance, small companies and individual entrepreneurs in Italy are overwhelming. They represent over 99 percent of all enterprises, employ just over 66 percent of workers and produce more than 62 percent of total value added.

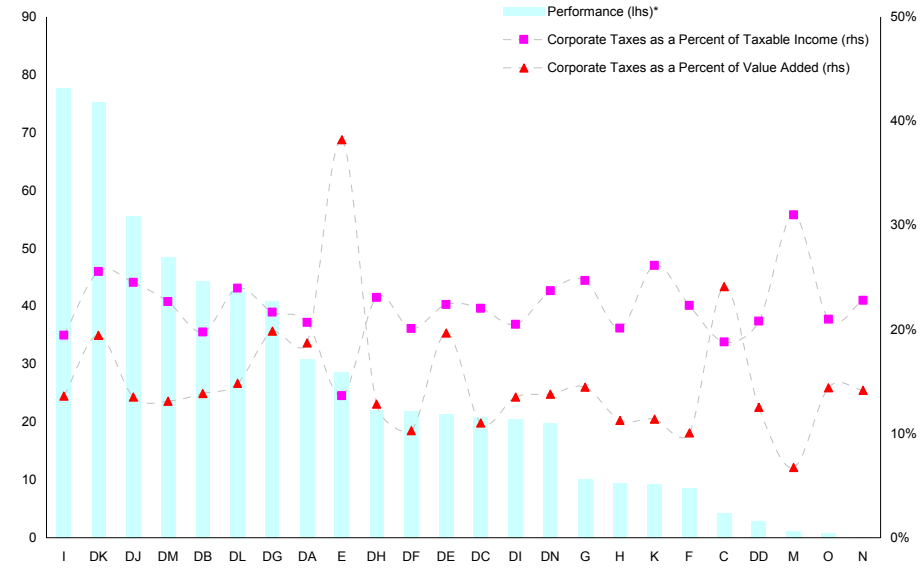
c. Taxation and Economic Performance

The view behind existing summary tax indicators, is that they can be effectively paired with indicators of economic performance and interpreted on the basis of convictions, such as that a negative relationship exists between tax burdens and performance. The high level of aggregation and heterogeneities of various types in the data and, hence, in the indicators that are customarily used, often void of significance the inference that can be drawn from them. Averages always hide the dispersion in the frequency distributions from which they are calculated. Summary indicators are no difference. As with the former, there is a risk that they can serve more to back beliefs than provide strong objective evidence.

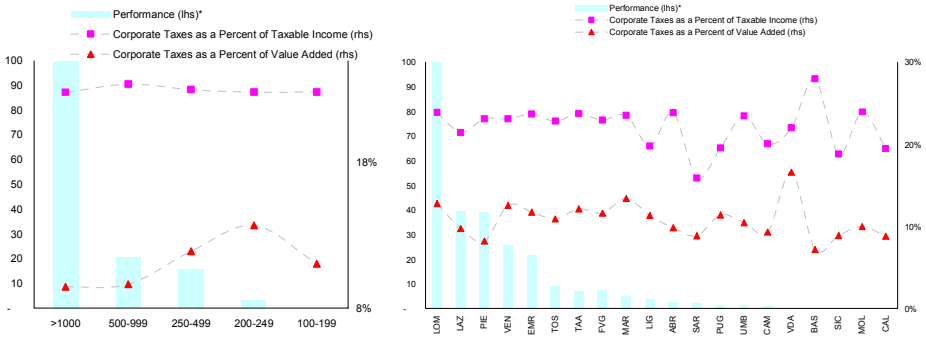
DIECOFIS has allowed to progress on indicators and to throw new hindsight into existing relationships. It has made it possible to calculate sets of micro-founded indicators of impact and performance, characterized by high levels of homogeneity and precision (since they refer to the same populations). In turn, this has permitted to finely map enterprises and their performance by, say, size, regions, sectors and so on, using sets of elementary and multidimensional, composable and decomposable indicators. As it can be gathered from Graphs 5 and 6, DIECOFIS has made possible to draw “parades” displaying “dwarf” and “giant” enterprises (or groups of them), lined up from top to worst performers (with performance measured by means of a

three-dimensional and composite indicator for NACE sector, and enterprises classified by region and size). Since, the tax burden for each enterprise in the parade was known, the analysis has permitted to get rather precise maps and indications as to how the latter may relate to the former.

Graph 5 – Performance[13] and Tax Burden by NACE sector[14] in Italy



Graph 6 – Performance and Tax Burden by Number of Workers and Region in Italy



[13] As measure by the RO Composite Indicator of Performance. See Roberti P. and Oropallo F. (2003).

[14]

NACE	DESCRIPTION	Firms %	Empl. %	NACE	DESCRIPTION	Firms %	Empl. %
C	PRODUCTS FROM MINING AND QUARRYING	0.2	4.6	DK	MACHINERY AND EQUIPMENT N.E.C.	10.7	2.7
DA	FOOD PRODUCTS, BEVERAGES AND TOBACCO	4.5	2.9	DL	ELECTRICAL AND OPTICAL EQUIPMENT	6.6	3.6
DB	TEXTILES AND CLOTHING INDUSTRY PRODUCTS	8.6	2.1	DM	TRANSPORT EQUIPMENT	4.0	6.7
DC	LEATHER AND LEATHER PRODUCTS	2.2	1.7	DN	OTHER MANUFACTURED GOODS N.E.C.	2.9	1.8
DD	WOOD AND PRODUCTS OF WOOD AND CORK (EXCEPT FURNITURE)	0.8	1.4	E	ELECTRICAL ENERGY, GAS, STEAM AND WATER	0.7	20.4
DE	PULP, PAPER AND PAPER PRODUCTS; RECORDED MEDIA; PRINTING SERVICES	3.6	2.4	F	CONSTRUCTION WORK	3.9	2.1
DF	COKE, REFINED PETROLEUM PRODUCTS AND NUCLEAR FUEL	0.4	6.3	G	WHOLESALE AND RETAIL TRADE SERVICES	8.8	3.3
DG	CHEMICALS, CHEMICAL PRODUCTS AND MAN-MADE FIBRES	4.9	3.6	H	HOTEL AND RESTAURANT SERVICES	2.4	3.6
DH	RUBBER AND PLASTIC PRODUCTS	3.5	2.3	I	TRANSPORT, STORAGE AND COMMUNICATION SERVICES	4.7	13.2
DI	OTHER NON METALLIC MINERAL PRODUCTS	3.6	2.5	K	REAL ESTATE, RENTING AND BUSINESS SERVICES	9.4	2.4
DJ	BASIC METALS AND FABRICATED METAL PRODUCTS	9.1	2.3	M - N - O	EDUCATION – HEALTH AND SOCIAL - OTHER SOCIAL SERVICES	4.6	6.0

III. LESSONS THAT CAN BE DRAWN CONCLUSIONS AND NEXT STEPS

This analysis leads to three main conclusions. The first concerns the issues in constructing *EISIS*s. The second, the issues in accessing *EISISs* in ways that are flexible and adaptable. And the third, modelling.

1) Issues in constructing EISISs *EISIS*s have a great potential. They create new "information value added" and permit to remove artificial wedges that presently hinder data access[15].

As Chart 4 above shows, PIA research activities may be aimed at providing "inputs", "means" and "output". They can cut-across the board, as when the purpose is to develop general PIA knowledge and capacity; or be area specific. The former embrace a range of necessary "spine" activities needed to carry out policy analysis and evaluation. These include disciplinary research, operational research, applied policy analysis, IT development and basic input information/treatment. The latter, instead, focus on specific policies and areas.

Three reasons support a systemic approach to the development of PIA knowledge-bases and capacity. First, because of complexity and rapid change, the data and models that are used for PIA are characterized by a process of rapid obsolescence. Continuously, new sources, additional fine-grained information and new tools are needed to successfully confront changing and entirely new problems. Thus, PIA requires anticipatory research, research concerning how and when to develop new data sources, and research on methodologies for new model requirements.

Second, Data has little sense without some theory supporting it. National accounts, the key statistical function in the world, is born from Keynesian macroeconomics. It provides the inputs required for macroeconomic PIA. The theoretical progress in microeconomics and, more generally social science, in fields as important for policy making as market structures, pricing, technological change or simply investment, is still lacking an adequate statistical observational system. Consequently, lack of adequate data hinders not only research on these issues, but also on the theoretical underpinnings which are needed to support data collection activities.

Third, the advent of the Information Society is permanently recording and storing billions of transactions that could potentially provide this microdata required by microeconomic PIA. This information golden mine is to a great extent un-mined. Nor are new technologies being used to store and access appropriately the microdata that are collected. To a great extent, collection and access remain supply driven, not demand pulled, with a strong inclination to "rehearse the past". Besides, a variety of wedges hinder access (including wedges due to old fashioned storing methods), with a consequent loss of

[15] See Roberti P *et al.,*"*Verso un'analisi "sistemica" del tessuto industriale italiano*", in l'**Industria**, 4, 2002, p. 666.

potentially available information and high opportunity costs in terms of both quality and quantity.[16].

One aim of **EUROKY-PIA** is to develop e-accessible statistical information systems of micro-data, including metadata, relational models, software and statistical methodologies to support PIA in various policy areas. The objective is to make them accessible to Government Agencies, the research community and stakeholders. This will require reconciling (data-linking) statistics from multiple sources, taking into account differences in definitions and adjusting for inconsistencies between sources (metadata matrix), as well as knowledge of the quality and reliability (statistical properties) of linked datasets. In parallel an investment in IT software (including extract, transformation and Loading procedures; implementation of integration algorithms; on line analytical processing implementation) is needed.

Difficulties cannot be underestimated. Of course, the task is not easy for data providers, who necessarily have to operate in an economic environment that constrain their capacity for action. But in many cases it will be more a matter of applying appropriate information technologies. One advantage with the new data is that it does not need to be developed by expensive surveys; it is already there in the network, and the question is how to find where it is and how to get access to it.

In the construction of Multi-Sourced Databases for Companies, the richness of data required means that

- access to a variety of micro-data sources is essential, that the data sources need to be matched successfully (either by record linking or statistical matching or imputation), and that the variables used need to be well understood because of the recognised dangers of trying to draw conclusions from data drawn from different sources;
- that further data capture may be necessary; in particular that the data drawn from the tax system for analytical purposes needs to be extended and centrally collated. For companies the limited range of information on the tax assessment covering the different types of income and the main allowances and reliefs is already available. The initiative to encourage companies to e-file the whole of their tax return, including the accounts and the computations that derive the assessment details from the accounts, should make more information available for the self-selecting group of companies which e-file. Comparable data capture for others remains an
- outstanding problem. Similarly the collation of details for small companies and their directors needs development;
- it is unlikely that one multi-sourced database will suffice for the range of purposes envisaged to support business tax policy. The tasks of updating and maintenance and the inevitable problems of size and complex structure suggest that trying to hold all the data as one entity will be expensive and probably ineffective. It is expected that the data will need to be organised as a substantial core with prompt and easy

[16]., *Ibidem*

access to other databases. Furthermore, the arrangements for access, record linkage, imputation, and perhaps modification of variables must be solved in advance for effective use.

2) Issues in accessing EISISs in ways that are flexible and adaptable
This is about the analyst's task. Successful analysis requires data, IT facilities, and an analyst capable of merging the former components successfully or, if *EISIS*s and IT facilities are available, of digging out the right statistics that are needed. For business sector research, the complexity of the data and the extensive modelling needs indicate that the analysts must be extremely capable. Their role will not be to use a few menu-driven screens on a well-established and user-friendly model using a fixed clean dataset. Their role will be to first establish the key variables for their research from previous research and discussion. Then, they will need to form a relevant dataset from an *EISIS* or, if this is not available, from available core data, other databases, and by new data collection if necessary. Almost certainly they will use a subset of businesses of which some will have missing information; they will therefore have to reconsider aspects of non-response, sampling and grossing to population levels. Next they will analyse their data before considering the precise modelling to be undertaken which will probably involve coding specific changes, incorporating various assumptions and estimated relationships. Sensitivity analysis will be necessary. And the whole process will probably have to be repeated several times as various aspects are refined. The IT model behind DIECOFIS *EISIS* (see Annex 1) provides a prototype of the software that may be applied for managing and accessing the data needed for microsimulation.

3) Modelling and International Comparisons DIECOFIS has confirmed that micro-data analysis, possibly linked with some macro-data, and microsimulation are certain to be fundamental for effective policy analysis in general, and for taxation and fiscal indicators in paricular. For the business sector, the variation over the population is much greater than that for individuals and households. The Inland Revenue has several micro-simulation models for the business sector (CT forecasting, oil and gas tax forecasting and costing, insurance companies tax, capital gains of companies), but these do not have the wide ranging capability that is sought.

Experience so far suggests that the best strategy for the future is to develop a modelling "environment" that focuses on facilitating research studies, rather than providing one model that delivers the required output from a few menu-driven screens. At the national level, this environment would have to provide the following facilities:

- a core database and prompt access to a range of other databases with aspects of access, record linkage, consistency of variables already solved;
- the addition of further micro-data by the researcher, with suitable record linkage facilities;
- a library of macro-economic data that can be used in analysis;
- core modelling facilities including the capacity to design and establish new datasets for analysis;

- standard software for example tax calculators, standard reports, graphics, and access to already established models;
- some advanced modelling facilities, for example the ability to model using different hierarchies of data (global performance, UK and IT group, individual company), expected behaviours based on previous research, some feedback facilities for use in simulating changes, iterative solutions, and projection facilities for ageing the population taking account of the economic conditions and the diversity of the population;
- bespoke software development so that the researcher can modify and extend the core modelling and standard software to suit the specific needs.

The overall objective is to develop a highly flexible modelling environment to meet the wide range of requirements. Modelling must develop from the relatively static limited data models that served yesterday's needs to dynamic models using multi-sourced micro-data combined with relevant micro-data.

At the EU level, what one will want to ensure is that the possibility for "plugging in" a common European tax module is envisaged in each national microsimulation model. The Italian team is piloting the issues and problems that may have to be expected. Currently, the work looks feasible. Yet, more countries will have to be modelled before feasibility can be truly gauged
One of the aims of the Diecofis project was to facilitate international comparsions of the business sector and its taxation. The UK contribution from the Inland Revenue has approached this issue from the perspective of the needs of the tax authority which needs a strong evidence base for its policy decisions. UK research has demonstrated the need for substantial development of data and modelling facilities to achieve that objective. The Italian contribution from the ISTAT has approached the issue from the perspective of a producer of data. Its research work has relied on both statistical surveys and administrative sources, including tax authority information. It has shown great potential. At the same time it has confirmed the difficulty of making direct comparisons across countries, since approaches are heterogeneous and *EISIS*s are far from common. In addition, the composition of the business sectors is quite different and tax systems are not easily comparable. Comparisons of different tax systems on a typical business basis approach are possible. At this stage, however, they are believed to be uninformative, since many tax heterogeneities may be hidden behind them. Further work is needed to develop the multi-sourced data and modelling described above in several countries so that, for example, detailed analysis of *ETRs* for large companies can be undertaken or, for small companies, the relationships between tax on directors remuneration and company profits can be compared. The analysis of the sources of heterogeneity across groups and countries is important and should be factored in in developing *EISIS*s and models.

ANNEX
Main Features of the IT Model for DIECOFIS - EISIS

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.

Chart A1 –Diecofis Conceptual Scheme

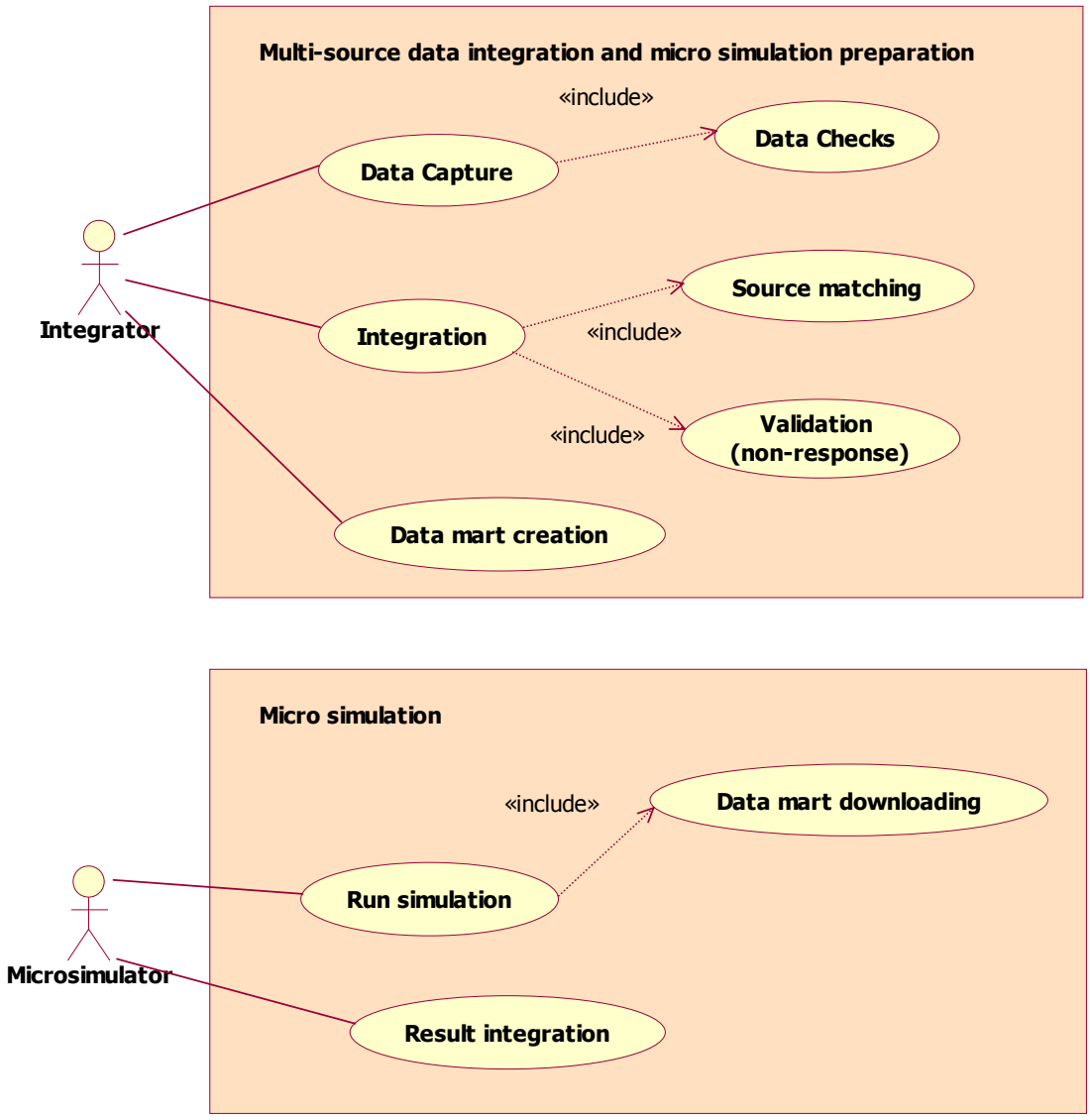
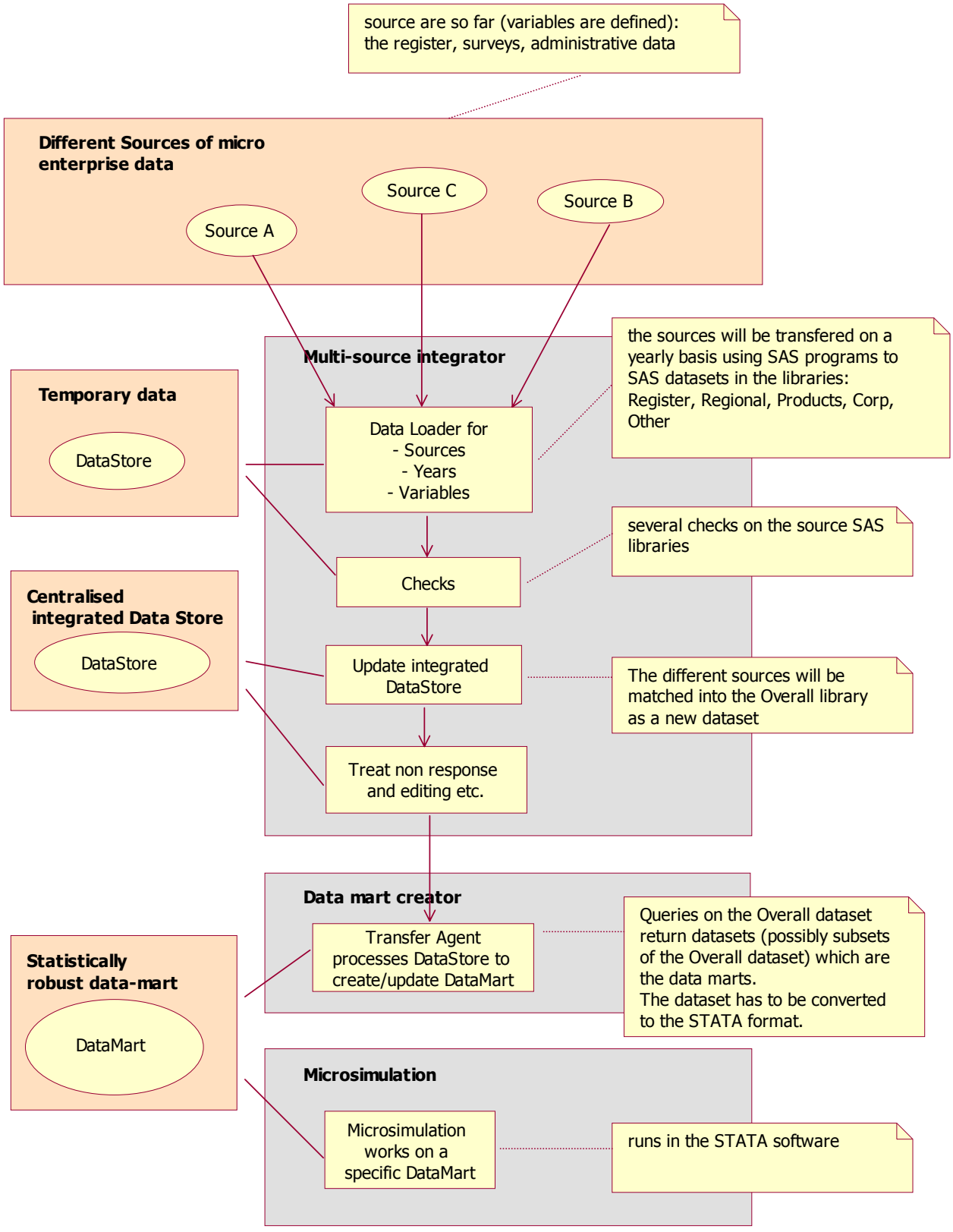
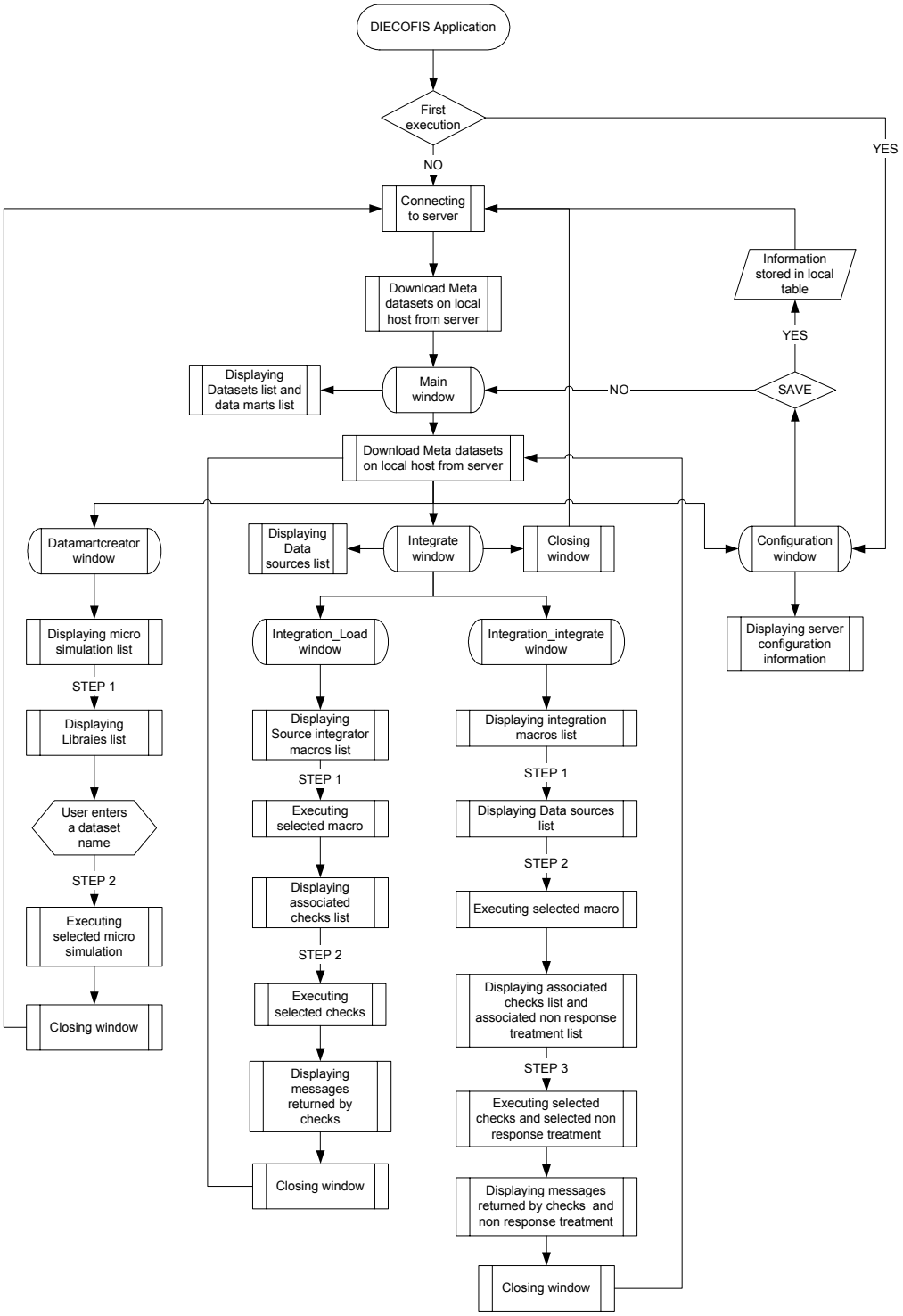


Chart A2 –Diecofis IT formalization of the Integration process



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.

Chart A3 –Diecofis Interface Flow Cart



REFERENCES

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

Eason, R., (1997), "*Modelling Corporation Tax in the United Kingdom*" – Statistics and Economics Division - Inland Revenue, UK.

Eason, R., (1993), "*Microsimulation for direct taxes and fiscal policy in the United Kingdom*" prepared for IARIW conference in Canberra, December 1993 - in: A Harding, "Microsimulation and Public Policy" North Holland Ed. (1996).

Grosh, M., Glewwe, P., (2000), Designing Household Survey Questionnaires for Developping Countries - Lessons from 15 Years of the Living Standards Measurement Study, The World Bank, Washington D.C.

Milanovic, B., (1999), True World Income Distribution, 1988 and 1993, in «Policy Research Working Paper 2244» - World Bank - Washington D.C.

Oropallo, F., Lo Cascio, L., (2003), The Development of a Multisource and a Systematized Database for Economic and Policy Impact Analysis – Diecofis project - CAED London – September 2003.

Roberti, P., Oropallo, F., Inglese, F., Lo Cascio, L., De Martinis, G., (2003), "*Towards a Systemic Analysis of Italian Industrial Texture Review*" - Industria 4/2002 – Il Mulino – November 2002.

Roberti, P., Oropallo, F., (*forthcoming*), Composite Indicators for the Measurement of Economic Performance - mimeo – ISTAT - NESIS, Rome – 2003

Yeend, C., Eason, R.J., (2003), Database Construction for Tax Indicators Building Purposes - ISSUES OF DATABASE DEVELOPMENT IN THE INLAND REVENUE, UK – Diecofis project - http://www.istat.it/diecofis