DIVERSINESS AND COMPLEMENTARINESS OF STATISTICAL SOURCES

Àlex Costa and Jordi Galter Pareja

Institut d'Estadística de Catalunya

Barcelona, Spain

Address:

Via Laietana, 58

Barcelona 08005

Tel. 934120088

Fax. 934123145

40th CONGRESS OF THE EUROPEAN REGIONAL SCIENCE ASSOCIATION
29 august-1 September 2000 in Barcelona

DIVERSINESS AND COMPLEMENTARINESS OF STATISTICAL SOURCES

ÀLEX COSTA/JORDI GALTER
Institut d'Estadística de Catalunya (Idescat)

Introduction

The first part of this paper presents the diverseness and complementariness of statistical forms of productions and of general statistical sources. The reader will be able to evaluate that the diversity identified is not useless, instead it reflects the different capabilities and limitations of each source. Therefore, it's possible to affirm that the diverseness of the statistical sources is justified by the complementariness between diverse forms of data production.

In the second part of the paper regional sources are taken into account. The diverseness and complementariness of sources are once more considered. At this point, the convergence of the priorities and activities of diverse statistical organizations such as the central and regional governments is crucial. Complamentariness is analyzed from two points of view: production process and final statistical products.

Diverseness of general statistical sources

In official statistics diverse methods of information production exist together: census operations, survey investigations, the use of administrative registers and the estimation of synthetic accounts. All these different methods of production are useful, however, and they clearly have different drawbacks and features. We will examine this diverseness in a simple and empirical way, generating statistics about statistics that are included in the publication catalogue of Idescat.

In Idescat's catalogue, we can find summaries such as the Statistics Yearbook of Catalonia, books about instrumental statistics (methodological or classifications), or publications on subjects such as: demographics, social statistics, structural economic statistics or short-term economic statistics. In order to determine the diverseness of the production methods already mentioned we can elaborate statistics on statistics (Table 1).

Table 1: Statistics about the methods of statistical production. Results in percentage. Idescat publications.

	Census	Surveys	Ad. Registers	Synthetical	Total
Demography	50,0	--	33,3	16,6	100
Social	33,3	33,3	16,6	16,6	
Economic structure	8,3	58,3	16,6	16,6	100
Short-term economics	--	66,6	16,6	16,6	100
Total	16,6	47,2	19,4	16,6	100

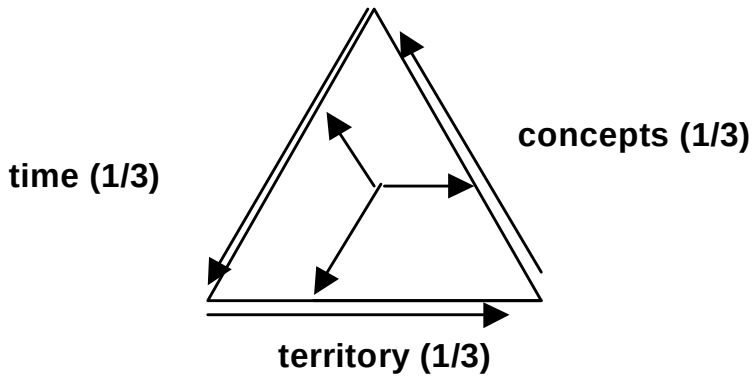
In these results, it can be shown that even if sample surveys are clearly the most important form of production (they alone represent 50% of the total), the other methods are not marginal, but instead have an important presence. Therefore, the information made available confirms the importance of sample surveys but it also illustrates and confirms the idea of the diverseness that was mentioned at the beginning of this paper.

Complementariness of general statistical sources: quantity and promptness of the information.

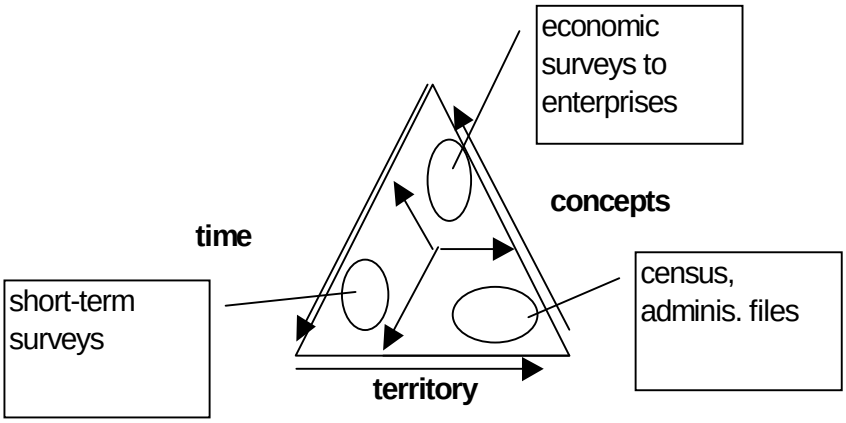
The largest amount of information and its promptness are natural and legitimate goals. The amount of information can be expressed as a function of an optimum conceptual detail and a great territorial distribution. Daily experience shows us that these three dimensions: promptness, conceptual detail and territorial distribution, are rarely found in acceptable levels in one statistical product. Generally, a survey with large conceptual detail (for example, an enterprise survey), will present some delay and will rarely make an important approximation to the territory. Also, promptness, so important with regard to short-term indicators, does not provide a satisfactory territorial approximation and limited conceptual detail. Finally, there are some statistical operations, such as the population census, that offer a large territorial distribution, but are not conceptualy detailed and are not satisfactory with respect to their quickness in data availability.

This does not mean that we must not make an effort to improve the promptness of enterprise surveys or that they can not be used at the territorial level. Neither is it pretended that short-term statistics should necessarily be deficient from the conceptual detail point of view. What has to be pointed out is that, in reality, some trade-off exists between these three characteristics of a statistic. This argument can best be visualized by a graphic in form of a triangle with projections at each side. A statistic that is perfectly centered in the triangular space would have a value of 1/3 at each side, with axis that measure its value in promptness, conceptual detail, and territorial distribution. This idea is shown in graphic 1. The rule that this triangle shows can be explained the following way: for a fixed amount of statistical resources and technology (methodology and organization), the significant improvement one of the three aspects will result in some loss in the remaining two. Another way of explaining this idea: each statistical operation has a place in this space, giving more or less priority to each feature. For example, as graphic 2 shows, the Industrial Enterprise Survey, the Population Census and the Large Shopping Centers Sales Index (LSCI) have in each case different priorities on each side of the triangle.

Graphic 1. Three presentations of a statistic: the rule of the triangle



Graphic 2. Three primary operations and three features: example of the rule of the triangle



This complementariness of the features regarding the amount of information and its promptness is the first of the justifications about the diverseness of sources. Regarding the forms of productions, we cans see in table 2 the complementariness of primary forms of production.

Table 2: Features of the primary forms of statistical production

	Timely promptness	Territorial information	Conceptual detail
Census	--	++	+-
Structural surveys	+-	--	++
Short-term surveys	++	--	+-
Administration registers (i)	+-	++	--
Administration registers (ii)	++	+-	--

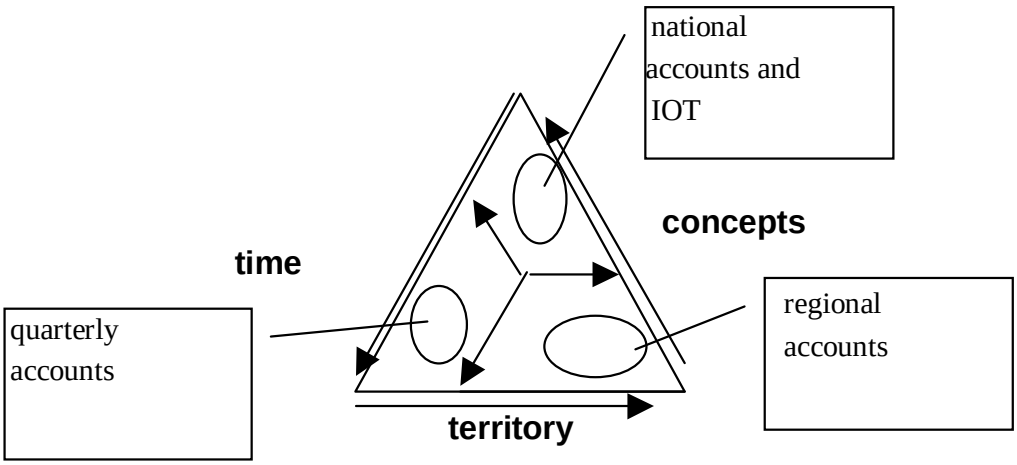
The rule of the triangle is not a fixed rule without exceptions. International trade data from customs are punctual and also have a great detail in their nomenclatures. The Labour Force Survey is quite good regarding its conceptual detail as well as its

promptness, although its territorial distribution is affected by the rule of the triangle. Despite the exceptions that can be found, the rule of the triangle is sufficiently valid in general and it also is useful in a didactful manner for statistics users to establish marked priorities as is the case of regional statistics.

In the area of synthetical accounts the rule regarding the contraposition between promptness, conceptual detail and territorial distribution also applies as can be expected if we consider that the synthesis feeds itself from statistical information. In graphic 3, this situation regarding accounting statistics is shown.

The rule of the triangle is the first expression that allows us to affirm that the diverseness of sources and of statistical production processes, if it translates in complementary information, is a positive diversity. This diversity allows us to have strategies that approximate the reality adapted to the preferences of our investigation. This is important since punctual information at the concept and territory level is difficult to obtain.

Graphic 3. Three account operations and three features: a sample of the rule of the triangle



Diverseness of regional statistics sources.

All that has been said up to now about the diverseness of statistics sources and about their complementariness can be applied to all statistical systems. However, in a statistical system such as ours, where regional statistics is the subject of interest and of responsibility of central statistical organizations, such as INE in Spain, as well as of regional organization of the autonomic communities, a new diverse and complementary axis is added related to the plurality of production organizations. At this point, we must ask ourselves what is the nature of the diverseness found in regional statistics sources produced by the convergence of statistical administrations and we must also ask ourselves about their complementariness.

One first basic element is the legal framework. This framework determines the exclusive competence of statistics regarded as those of state interest at the central administration and those regarded to be of exclusive competence at the regional level. However, official statistics is not a free investigation field but rather is determined by an international consent (in our case basically European) that defines a rather closed catalogue about which statistic that are useful for the following of the demographic and economic reality. This creates a general coincidence of interests between the central and regional administration, with some differences regarding the emphasis and focus. In front of this situation and considering the fact that the central statistics was already at work at the moment of developing the official statistics at the autonomous community level, we can identify three basic strategies: 1) minimize production activities in favor of dissemination, 2) duplicate operations, 3) complementary operations.

Most of the autonomic communities have opted basically for the first or third strategy and only in specific cases has the second option to duplicate been considered. This second option is possible, although it's expensive and complex, and from the cost/benefit point of view it's difficult to justify. In the case of Idescat, we have given priority to the first and third options.

Complementariness of regional statistics sources: process collaboration and making the most of each activity

One first line of complementariness appears in the activity field, in the production of large statistical operations that, as a survey or as a census, are carried out by the central administration's statistical organization, INE. It's clear that these operations are of most interest to regional statistics. For example, the population census, agriculture census, the annual survey to industrial enterprises, and the surveys to service enterprises. Idescat's strategy was from the beginning one of collaboration and support to INE's operations and, at the same time, to make most use of them. This strategy of collaboration and making the most of existing statistics generates an important and diverse number of performance formulas. Table 3 summarizes different possibilities of intervention and making most of data.

Table 3. Idescat's areas of activities in relation to INE's programs.

Questionnaire	Edition of Spanish/Catalan questionnaires with INE/Idescat anagram	Addition of a module with questions that are of interest to Idescat
Directory	Exhaustive Directory Inspection	Continuos updatting of the directories
Sample	Without participation (within the limits of general interest, a general interest for Idescat could be defined)	
Field work	Questionnaire presentation with INE/Idescat anagram and signatures	Idescat's field work execution
Purge, imputation and validation of microdata	Purge, imputation and validation of microdata of Idescat	
Elevation and estimation of sample data	Idescat elevation is complementary to INE's elevation	Idescat elevation is autonomous to INE's elevation
Tabulation from microdata	Idescat's tabulation is complementary to INE's elevation	Idescat's tabulation is autonomous to INE's elevation

At the last table, five of Idescat's activities have been identified that are associated to INE's procedures, some of which can be considered as: 1) coproductions with process collaboration (with or without fieldwork), 2) dissemination from microdata and 3) autonomous production while taking advantage of INE's microdata.

All these formula relationships with INE are related to the survey and census fields. Specifically, the five formulas are found in the surveys while census represent only two formulas: strong production collaboration and dissemination.

By their nature, administrative registers are an activity that is attained in collaboration, since it's an activity that takes into account the use of external organization data (customs, commercial registers, etc.). In any case, we can differentiate between simple and direct procedures or more complex procedures. Finally, the secondary production or synthesis of accounts, is carried out at the moment without collaboration in the process, although there is some counseling relationship.

Considering all that has already been explained, if we want to evaluate the diverse forms of production (and also the dissemination from microdata), we must add new forms of production to those that are already known.

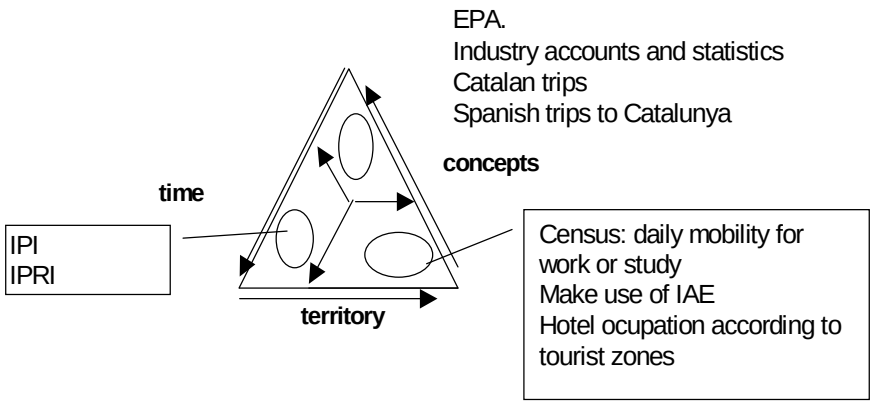
The last recount allows us to renovate the statistics of table 1 about production methods in Idescat's case, as a regional statistic organization. We can see how the different forms of cooperation and making use of existing data on official regional statistics, does not create a negative panorama but instead produces more diverse forms of procedures, in a way that diverse general sources are multiplied by regional sources according to complementariness strategy in the production process.

The complementariness of regional statistics sources in final products: quantity, promptness and comparison of results

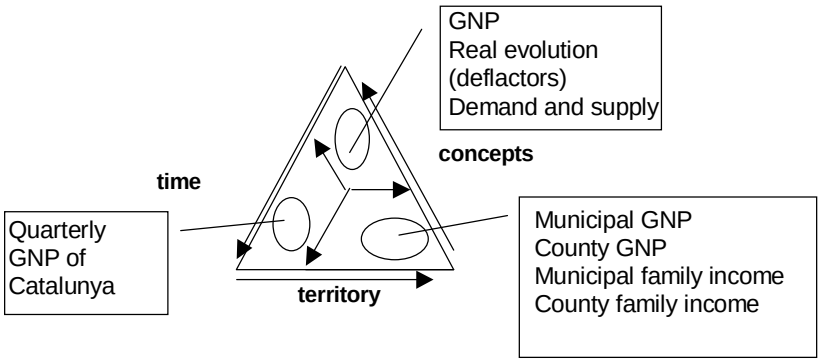
The second type of complementariness refers to the field of results. This complementariness derives from the acceptance of the regional information provided by the central administration's statistical office, normally the INE. This way, the best operative activity and most justifiable socially tries to compliment this body of data in those points where the data available does no cover our fields of interests. Following this line of work, the next step would be to analyze those general type of products that include regional information, detect deficiencies and consider improving them. From this point of view, and considering the three viewpoints presented in the first part of this work, we can identify a series of statistical results that complete the information on

Catalunya and that emphasize more conceptual detail, a better knowledge of the territory, or some statistical products more specific or necessary to follow short-term indicators. In the field of primary production (with all diverse forms of production that we have shown) and of the synthesis of accounts, examples of these different statistical results are shown in graphic 4 and 5.

Graphic 4. Idescat's primary statistical products



Graphic 5. Idescat's accounts statistical products



Each one of the products of these two graphics has a clear justification since they improve the information offered by INE.

These statistics are produced following different strategies: sometimes with autonomic surveys, as is the case of the LSCI, of the tourist trips of Catalans, or the trips made by residents in other communities from the rest of Spain or Catalonia. In other cases, it's necessary to obtain a specific module of interest to Catalonia's government as is the case of imposed mobility in case of work or study. Sometimes, it's very useful to exploit INE's files as is the case of the Industrial Production Index or the Industrial Price Production Index. Finally, in operations so important as the Labour Force Survey , the Industry Survey, or the Agriculture Census, an additional complementary tabulation is carried out. This tabulation can be accompanied (or not) by field work cooperation in the production process. These are the cases of the industrial survey and of the agriculture census, respectively.

The possibility to adopt the central administration's results and complement them is clearly effective in the case of macromagnituds. From the first moment, Idescat studied the suitability to adopt the final results of INE's Regional Accounts. It must be pointed out that once these results were adopted, it became necessary to extend them conceptually offering annual GDP results not only from the supply side but also from the demand side as well as the need to deflate all the macromagnituds in order to determine the annual evolution of the Catalan economy in real terms.

In the territorial field, it was necessary to supply data from the family home income accounts and GDP at the county and municipality level. Finally, regarding time series, it was an important goal to arrive to Catalonia's quarterly accounts, obtained at the same time and comparable to Spain's quarterly accounts. At the moment, these results, all compatible with the final estimations of INE's Gross value added and family income estimations for Catalonia, are greatly appreciated by our users who by this method find in Idescat's activities a source of enrichment and complentariness with INE's information, and it does not create a contradiction of choosing between available data.

The key to this planning consists, as has already been mentioned, in assuming INE's results as interesting and available to be extended and complemented. To make this

possible, as has been explained before, there are institutional and legal reasons. There are also efficiency reasons when it is clear that official statistics are considered to be a public service.