



Background paper for the Workshop on Point-based and Area-based statistics
(Work Session on Methodological Issues Involving the Integration of Statistics and Geography
25 – 28 September 2001, Tallinn, Estonia)

In search of a system of small statistical areas

Two Questions for discussion at the ECE conference in Tallinn September 2001

Lars H. Backerⁱ (Statistics Sweden)
And
Marja Tammilehto Luodeⁱⁱ (Statistics Finland)
Philippe Guiblinⁱⁱⁱ (the Office of National Statistics (UK))

Abstract

This note describes the main argument of the paper "Accountants and Engineers¹" submitted to the CES 2001 in Geneva.

The argument presented here is judged to be of central importance in order to improve the NSO's (National statistical office's) contribution to efforts to improve the "Human condition" through multilateral projects.

We believe that statistics has traditionally concentrated excessively on accounting models and methods that may primarily be used for benchmarking purposes and not as a support for engineering efforts to act.

We believe we should argue in favour of projects making statistics more suitable for "engineering" use, especially in emergency situations:

- to provide statistics for areas polluted with radioactive fallout after nuclear emergencies (see Chernobyl disaster April 25th -26th, 1986);
- to provide statistics for areas hit by floods and storms (see Poland in Summer 1997);
- to provide detailed maps describing the distribution of the "Foot and Mouth Disease" (see UK after February 19th 2001).

In order to be useful, the information provided must be considerably improved, particularly for key variables undergoing rapid change:

- harmonisation for more variables over larger areas;
- ever better modelling (from semantic descriptions to standards);
- higher resolution, greater detail (this is the domain of small-area statistics).

¹ See (Backer 2001) *Accountants and Engineers; why the difference?*

Table of contents

In search of a system of small statistical areas	3
Why systems of small statistical areas? (12 Points).....	3
A. Background	3
B. Information systems	4
C. Information for action	5
D. System of small areas for statistical analysis	6
Discussion.....	7
Two Questions	8
References	9

List of figures

Figure 1: Statistics for "Benchmarking" and for "Engineering"	4
---	---

In search of a system of small statistical areas

Among statisticians, it is generally agreed that "small area statistics" is an important field for development. As an indication, this topic was chosen as the "theme" for the annual "Conference of European Statisticians", organised under the auspices of ECE in June 2001, at Geneva.

On this and other similar occasions it has become clear that we all agree that a shared system of small area statistics is necessary. However, we do not seem able to agree on what we mean by this concept nor do we have a clear idea of what strategy to choose for its realisation.

An appendix to the Program² for a Workshop on "Point- and Area- based statistics" (small area statistics) to be held at the ECE conference in Tallinn 25 through 28 September 2001 was prepared in the current note.

Why systems of small statistical areas? (12 Points)

A string of ideas leading from the root of our actions to central questions that we want to focus on during our seminar may hopefully be illustrated by the 12 steps below.

A. Background

1. "Emergencies" and "opportunities"

The Spanish philosopher Ortega y Gasset once remarked that life is shot at us "point blank" and forces us to action. Otherwise put, we are constantly forced to act on "emergencies" to fight, and "opportunities" to exploit.

2. Our actions are not without purpose

John Cage, the famous American composer, once asked the question whether a conductor should only mark the time, or whether he should induce the musicians to ever-greater accomplishments³. Our actions, either as individuals or complex man-environmental systems, are not without purpose. The more "civilised" among us believe that the only worthy goal for human action is to improve the "Human Condition" for all⁴.

3. All actions are based on more or less qualified systems of knowledge

Our actions as individuals are based on more or less qualified "world views" or systems of knowledge that we use for reference when dealing with "life". The statistical systems contribute of course, together with a series of other sources of information, to our efforts to build qualified systems of knowledge based on the scientific method.

² For the program to this workshop the reference is (Backer 2001) *Point- and Area- based Statistics; The Program for a Workshop in search of a system of an infrastructure for small areas for statistics.*

³ See (McLuhan and Fiore 1967) *The Medium is the Message; An inventory of Effects*

⁴ For the idea of the task to improve the "Human Condition" as the overriding task of a civilised society see for example; (Arendt 1971) *The Human Condition*, (Heidegger 1949) *Ueber den Humanismus* and/or (Braudel 1993) *A History of Civilizations*

B. Information systems

4. Two types of information

In the pursuit of this end, there are two general uses for statistics and related information. We would suggest calling them "Information/Knowledge for Development" or "Engineering", and "Information/Knowledge on Development" or "Accounting".

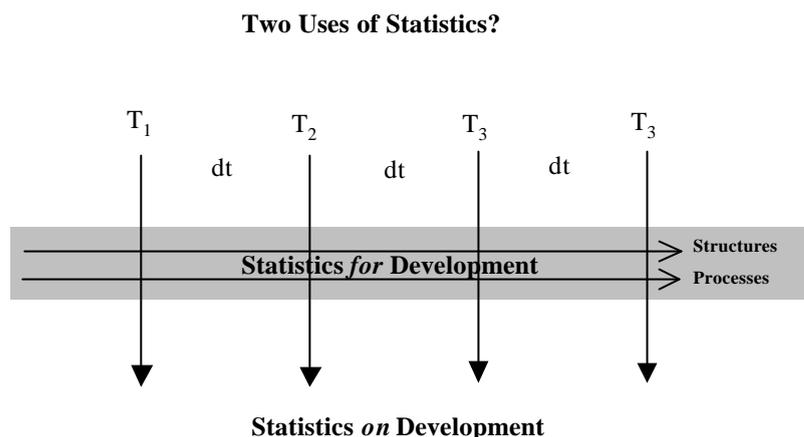


Figure 1: Statistics for "Benchmarking" and for "Engineering"

- **Information for Development/Change**
The first is to describe the human condition in terms of structure and processes (Engineering)(Focus on patterns of distribution)⁵
- **Information on Development/Change**
The other is information designed to judge the "state" of a region by comparing it to other systems using a system of standard benchmarks (Accounting)(Focus on aggregations)

5. Classical "Accountant" Statistics provide information *on* change or "Development".

Traditionally statistics, and especially official statistics, have concentrated primarily on the task of using "accountant" (Benchmarking) methods to describe the state of the system by aggregating statistics on administrative areas.⁶ Inspired by John Cage (above) the purpose of these statistics primarily mark the passing of time.

We have here taken it for granted that it is the task of a good government to not only mark the time by presenting reliable "Accountant" statistics, but also to take action to improve "the human condition" by effectively dealing with emergencies and exploiting opportunities on behalf of us all.

⁵ These distributions seldom (or never) respect administrative borders.

⁶ Evidence to support this argument is partly to be found in historical documents (See (Johannisson 1988) *Det mätbara samhället: Statistik ock samhällsdröm i 1700-talets Europa*) and also by comparing central statistical publications with accountant practices.(See: (Statistiska Centralbyrån 1999) *Statistisk årsbok; Statistical Yearbook of Sweden 2000*)

As a consequence of this principle it is not only our duty to account for change, but also to provide information, statistics or even knowledge to achieve "progress" or even "development".

6. *We have no satisfying system for providing qualified information **for** "Development".*

This information may be useful for benchmarking purposes but is of little use to the engineers as a foundation for taking action to improve structures and processes.⁷ Here, patterns are of great value. For such patterns to be useful they must have a high resolution, be much more "fine grained" than the lower levels of the standard system administrative areas. Statistics on commune level (Nuts5) is (for most countries) useful for policy design only at an extremely high and abstracted level of analysis far removed from "real" problems as experienced by individuals. The main problem is the lack of resolution⁸.

C. Information for action

7. *"Emergencies" and "opportunities" do not respect administrative borders*

The statistical system of single countries and whole regions might be of great value not only for our accountants (for benchmarking purposes) but also for our engineers (for policy implementation purposes).⁹

- See Chernobyl 1986
- See Floods in Poland 1997
- See Foot and Mouth disease in the UK 2001

Here we need information on the areas affected. Datasets should focus on the interior of these areas to be compared with the exterior. Emergencies and opportunities seldom respect administrative borders and have to be addressed according to their own "pattern" of distribution.

8. *Small area statistics are needed to make statistics available for development policies*

The argument for the development of small area statistics is not to be found in the improvement of traditional "Accountant" type statistics, but rather in the improvement of information needed for development policies.

9. *Challenge*

In a unified system of information describing our societies in terms of structures and processes, we need both engineering and accountant type

⁷ See here (Coombes, Green and Oppenshaw 1986) *An efficient algorithm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain and others*

⁸ For the problem statement as formulated by Eurostat/Gisco please see (Rase 2000) *Technical Specifications; Improving comparability of statistical data across EU Member States*

⁹ There is a strong demand for the multi-use of all types of resources according to "cost-benefit" arguments. This applies also to statistics. Accountant methods could be used to deliver aggregations of critical variables on polygons delimiting the affected areas only, and engineering methods to show the distribution of "observations" over the area.

information. The challenge to the statistical system(s) in general is to organise efforts to design and implement a geographical base for doing "small area statistics" that would improve the comparability of statistics across the ECE countries.

This implies

- Better Harmonisation
Trend towards global harmonisation of vital statistics. This is achieved through constant harmonisation efforts; regional datasets might be combined to cover increasingly larger areas.
- Better Models (standards for data and processes)
One major tool to achieve better harmonisation is through better standards based on ever better semantic descriptions (models) of the world. These models should account for both structures and processes.
- Higher resolution, greater detail
Due to the massive diffusion of computer methods, bulk information is not the most important problem, but constitutes an opportunity to provide both information on and for development. This calls for higher resolution. There is a trend towards large amounts of individual observations, all registered with proper reference to the state of their objects (their properties) and location in space and time. Data is aggregated on administrative areas only where it is absolutely necessary for disclosure purposes.

D. System of small areas for statistical analysis

10. An integrated system of small tessellations¹⁰ and a method for clustering regions

A system of small statistical areas consists of two components:

- A system of small statistical regions
A system of small statistical areas that consists of both a system of small irregular statistical "seed" areas¹¹ and a system of small grids¹² that covers the areas under our mutual jurisdiction;
- A clustering method
A program or method for clustering regions according to the statistics analysed.

¹⁰ See the so-called "Tandem project" financed by the Commission through Eurostat/Gisco and realised by a consortium consisting of experts from Statistics Sweden, Statistics Finland and the Office of National Statistics (UK).

¹¹ A system of very small irregular tessellations ideally containing approximately 2000 persons or smaller.

¹² A system of very small regular tessellations of 500X500m or smaller.

11. The Tandem project has offered a method for testing this and other hypotheses related to the development of a common system of geo-statistics.

The Tandem consortium¹³ has produced what it believes to be a hypothetical solution and an effective method to test it. Could we have a critical appraisal of the strategy suggested here?

12. We need to know whether this hypothesis and the method to test it are feasible

We need to know whether this method is feasible. Or do alternative, perhaps better, approaches to this problem exist?

Discussion

Seen in a historical perspective, statistics were primarily used as a tool for providing inventories for states and their varying subdivisions in what are now hierarchies of administrative areas or regions. This was normal at a time when we believed that progress and development equalled the idea of building countries as machines for effective production, distribution and consumption of goods and services.

At present, however, especially with the advent of new information technologies, our ideas of development have changed. We regard development as a system of processes where we progress by constantly observing changes in complex man-environmental systems and act to avoid disasters and exploit opportunities.

With this change in perspective, the production and use of information and statistics must change too. We are no longer satisfied with providing information on societies seen as a “kit of parts” that may be accounted for with simple inventories. We need to be able to map phenomena, their distribution in space and their change over time in order to provide decision-makers with qualified information to be used as a basis for rational decisions. Still, classical inventory-type statistics on administration are still in demand for many purposes (collecting taxes, calculating expenditures for health and education, etc.). But in order to manage environmental disorders or to react effectively to defend and develop complex industrial systems, we need to adopt process oriented techniques to monitor phenomena and map their spatial distribution and change over time.

It is a well-known fact that most things observed from this perspective display patterns of distribution that do not respect administrative borders. It is also a well-known fact that in order to describe patterns of distribution, we need a better tool than that represented by tiers of administrative areas.

In order to cope with this and related problems, the NSOs can be divided into two groups of countries that find themselves in differing situations.

1. The larger group consists of countries that collect the major part of their information as aggregated on tiers of small administrative areas (enumeration areas, etc.) We will refer to these systems as “area-based”.

¹³ The Tandem consortium was organised to realise the project described in (Backer 2000) *Tandem GIS: A feasibility study towards a common base for statistics across EU. A management summary (Revised version)*

Since they generally have less data on point coordinates, they often have good systems of areas.

2. A smaller group of NSOs register their information with point coordinates.. We will refer to these systems as “point-based”. As these countries have data on point coordinates they have little need to develop expensive systems of small statistical areas.

However, the situation is not as clear-cut as it may appear, because most NSOs produce and use both area-based and point-based geo-statistics. So, in order meet the demand for improving the comparability of statistics for describing both “patterns of distribution” and aggregations of statistics on administrative areas, we need to agree on 2 important issues regarding a system of what we call "seed areas" for statistical analysis:

1. We should try to agree on shared ways to describe patterns of distribution using both area- and point-based techniques; and
2. To agree on how point-based and area-based statistics may be used in combination (aggregations and disaggregations from one to the other).

In addition to this issue pertaining to the use of small areas for statistical analysis, there is another less obvious question. If the number of areas used in the analysis is relatively small it becomes very important that the areas are either of comparable size, or that the number of observations are roughly equal in each area.

This fact calls for a method for aggregating (clustering) both regular and irregular tessellations before any analysis can be started.

The Tandem_GIS is an effort to study the feasibility of this ambitious undertaking. It should be clearly stated that this is not a project to extend or change the NUTS hierarchy of administrative areas. It is rather a project aimed at improving the comparability of statistics by providing better means for aggregating statistics on other hierarchies of regular and irregular tessellations that may better describe patterns of distribution than the traditional systems of (small) administrative areas.

Tandem_GIS will report to Eurostat in October 2001.

Two Questions

We will, at the current workshop, try to address both of these questions and hope that the participants of the workshop may help to answer 2 questions:

1. The Tandem consortium suggests that in order to develop a unified (harmonised) system of geo-statistics, it is necessary to see statistical systems as a system of processes rather than as a structure of data. If so, then we need to test new advances against a system of production (sub-) processes that are compiled after continued competition over time.
Do you agree?
2. The Tandem consortium further suggests that, for the construction of a system of small areas for statistics, we need both a system of (regular and irregular) tessellations and a method for clustering them according to the data studied. (Thus posing as a candidate to become a "best practice" in the method described under point 1.)
Do you agree? (Or are there any other better competing methods?)

Lars H. Backer
Stockholm, Monday, 03 September 2001

REFERENCES

- Arendt, Hannah. 1971. *The Human Condition*. Chicago and London: The University of Chicago Press.
- Backer, Lars H. 2000. Tandem GIS: A feasibility study towards a common base for statistics across EU. A management summary (Revised version). Stockholm: Statistics Sweden.
- . 2001. Accountants and Engineers; why the difference? Paper read at Conference of European Statisticians, June 8 2001, at Geneva.
- . 2001. Point- and Area- based Statistics; The Program for a Workshop in search of a system of an infrastructure for small areas for statistics. In *Tandem*. Stockholm.
- Braudel, Ferdinand. 1993. *A History of Civilizations*. Translated by R. Mayne. Harmondsworth, England: Penguin books.
- Coombes, M. G., A. Green, and S. Oppenshaw. 1986. An efficient algorithm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain. *Journal of Operations Research Society* 37:943-953.
- Heidegger, Martin. 1949. *Ueber den Humanismus*. Frankfurt AM: Vittorio Klostermann.
- Johannisson, Karin. 1988. *Det mätbara samhället: Statistik ock samhällsdröm i 1700-talets Europa*. Styockholm: Nordstedts.
- McLuhan, Marshall, and Quentin Fiore. 1967. *The Medium is the Massage; An inventory of Effects*. Harmondsworth, Middlesex, England: Penguin Books.
- Rase, Daniel. 2000. Technical Specifications; Improving comparability of statistical data across EU Member States, April 2000.
- Statistiska Centralbyrån. 1999. *Statistisk årsbok; Statistical Yearbook of Sweden 2000*. Edited by S. Centralbyrån. 1 vols. Vol. 89. Stockholm: Statistiska centralbyrån.

ⁱ Lars H. Backer is a senior GIS- R+D- expert, since 1992 connected to the program for regional planning and natural resources at Statistics Sweden. He holds graduate and postgraduate degrees in architecture, urban- and regional- planning from ETH(Zurich) and Nordic universities(Oslo, Stockholm). Mr. Backer is specialised in the construction and use of static and dynamic digital models for urban and regional development. He has extensive International experience from Teaching, R+D and consultancy in his field.

ⁱⁱ Marja Tammilehto-Luode is a Senior Adviser at Statistics Finland (SF), since 1992. She has been involved in various projects for developing the use of GIS and small area statistics. She holds a M.Sc. in Geography. Her work areas have spanned from GIS policy issues to new information technology application. Mrs. Tammilehto-Luode has extensive international experience in working groups, and consultancy of GIS, regional and urban statistics.

ⁱⁱⁱ Philippe Guiblin is a Statistician Methodologist at the Office for National Statistics (UK), since 2000. He completed a Ph.D. in Geostatistics in 1997 (École des Mines de Paris, France) in which he explored different ways of applying the geostatistical methods to fish stock assessments within a EU funded project. He is now involved in several ONS methodological projects covering spatial modelling issues, spatial statistics and small area statistics.