Statistics Production in the 90´s - Decentralization without Chaos

Bo Sundgren

R&D Report
Statistics Sweden
Research-Methods-Development
1989:11

Från trycket Augusti 1989
Producent Statistiska centralbyrån, utvecklingsavdelningen
Ansvarig utgivare Staffan Wahlström
Förfrågningar Bo Sundgren, tel. 08-783 41 48

ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1989

INLEDNING

TILL

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.
Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Efterföljare:

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report 1989:11. Statistics production in the 90's - decentralization without chaos / Bo Sundgren.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1989-X101OP8911

1989-06-20

STATISTICS PRODUCTION IN THE 90'S - DECENTRALIZATION WITHOUT CHAOS

Bo Sundgren

To be presented at the International
Meeting of Official Statistics
in Bilbao, Spain
September 7-10, 1989

STATISTICS PRODUCTION IN THE 90'S - DECENTRALIZATION WITHOUT CHAOS

Bo Sundgren
Statistics Sweden
S-115 81 STOCKHOLM

1 Centralization and decentralization in statistics production

Today official statistics production is by and large centralized in many countries, including my own, Sweden. Also if we look at a statistical office as an organization, we shall often find that it is highly centralized in many respects. Historically there have been several good reasons for centralizing official statistics production, on the micro level as well as on the macro level, but are these reasons still valid, and will they be valid during the 1990's, or will the next decade be a decade of decentralization? These are some of the questions that I will discuss in this paper.

1.1 A systems approach to statistics production

According to systems thinking, everything is a system, consisting of parts that are also systems, subsystems, and all systems are themselves parts of larger systems, supersystems. Figure 1 shows an application of this principle to official statistics production.

A basic building-block in all official statistics production is the individual statistical survey. On the first level, we have the operational parts of a statistical survey. We all recognize the traditional, serial flow of tasks that have to be performed, when we conduct a survey: data collection, coding, editing, data transformation, aggregation, tabulation, graphical presentation, analysis, publication, distribution.

On the second level, we control the different steps in the survey, and the survey as a whole, by means of design and planning, administration, and evaluation. The statistical design includes the establishment of a frame and a sampling strategy, if any, and the EDP design includes systems analysis, data modelling, and programming.

On the next level, we look upon a statistical system as a whole as the object of control. The statistical system under consideration may be the statistical information system of a country, or a part of such a system that is managed by a particular statistical agency.

In accordance with the systems principle, a statistical system is itself a component of a larger system. Like other information systems, a statistical information system is a part of a decision and control system, whose objective is to govern the behaviour of some 'piece of reality', the object system, in accordance with certain goals. The role of the information system is to help to describe, explain, and prognosticate the behaviour of the object system, and to assist in the evaluation of the effects of decisions and actions visavi the object system.

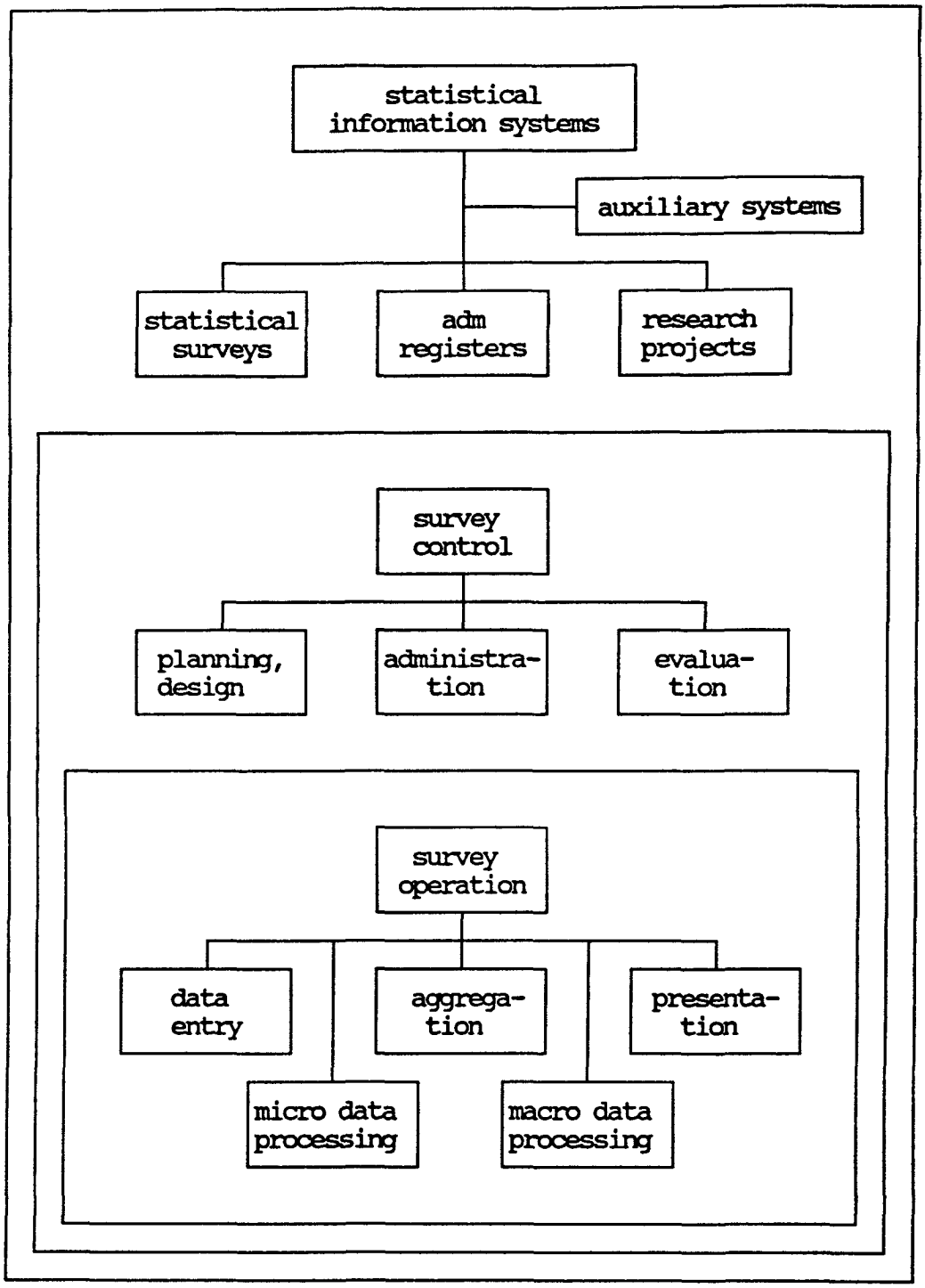

Figure 1. A systems approach to statistics production.

1.2 The meaning of centralization and decentralization

The systems and subsystems of official statistics production can be centralized and decentralized in different respects. On the national level centralization may mean that all major statistical information systems and their subsystems are (a) controlled, and (b) operated by a single organization, the central statistical office of the country. Decentralization would then mean that (part of) the control and/or operation of major statistical information systems is performed by some other organization(s) than the central statistical office.

Decentralization on the national level can take place in several dimensions. One type of decentralization is regionalization. For example, in federal states like the Federal Republic of Germany it is common that regional statistical offices have major responsibilities in official statistics production. Another variation is departmental decentralization, meaning that there are different statistical offices for different areas of social and political interest, or that statistics production is an integrated part of the governmental agency that is responsible for the administration of a particular area of interest (object system); for example, an agency for health administration could be responsible for producing official health statistics. Countries like the United States and the United Kingdom have basically a departmentally decentralized official statistics production.

Different functions that are performed in the control and operation of statistical surveys need not necessarily be uniformly organized from centralization/-decentralization point of view. For example, even in a highly centralized system for statistics production, data collection and other input-oriented functions may be decentralized to regional or local agencies, or (like when statistics production is based on administrative data) departmentalized to administrative organs. On the other hand, basically centralized systems for statistics production may also be more or less decentralized as regards the output-oriented functions; specialized research institutes may have the responsibility for analyzing the official statistics, private companies may take care of the distribution of statistical publications and other output from the statistical databases, and so on.

Inside a statistical office the work may be organized in a more or less centralized (functionalized) or decentralized (survey-oriented) fashion. For example, until recently it has been common for statistical offices (like other organizations) to have a highly centralized organization of computer-related work (systems analysis, programming, data entry, computer operation etc).

Thus if we move through the different subsystems and functions of the statistical system of a country, centralization/decentralization considerations come in on all levels and in different dimensions; due to combinatorial explosion the total number of organization alternatives for the statistical system of a country will be very large indeed.

2 The impact of new technology on the organization of statistics production

2.1 The role of computers in the centralization of statistics production

In order to study the role of computers and computer-related resources in the centralization of statistics production, we may look at the case of Sweden. Statistics production in Sweden was centralized into its present form in the early 1960's. The need to rationalize efficiently by means of centralized computer technology was then a major reason for centralization. But there were also others. One was the belief that only a strong, central statistical office could afford to maintain a powerful methodological development of high quality and enough quantity to form a "critical mass". Another reason for centralization was the need for coordination and integration of individual surveys into statistical information systems, based on unified conceptual models like the system of national accounts and socio-demographical and socio-economical models.

Thus in theory there have been several important and good reasons for a relatively high degree of centralization in official statistics production. However, judging from the Swedish experience, I would say that at least those arguments which have to do with the needs for coordination and integration of

statistics from contents point of view have not in practice played such an important role as one could have expected and desired. Instead the need to mobilize and share relatively scarce and indivisible computers and computer-related resources (programmers, systems analysts and other types of EDP specialists) seems to have been the most concrete and practically important reason for having a centralized organization.

2.2 Different types and levels of computer support

Since computers and computer-related resources seem to have a great impact on the organization of statistics production, it should be important to look at present trends in computer hardware, software, and methodology development, and to try to anticipate the application of the results of these developments in statistics production, if we want to foresee changes in the organization of statistics production. In the next chapter I will give a brief description of three important trends: decentralization, integration, and standardization. However, before we delve into those topics, we shall take note of another circumstance of great significance: the very scope of computers and computerization in statistics production is likely to expand dramatically during the next decade; a lot of new tasks and functions will be automated or at least computer-supported to a greater extent than before. Let us take a closer look at the possibilities by studying figure 2.

	ROUTINE TASKS	"INTELLIGENT" TASKS	
SURVEY SYSTEM DESIGN, CONTROL, AND OPERATION	dissemination of statistics from many surveys through statistical databases	helping university researchers to use and combine data from many sources	SURVEY + SURVEY ENVIRONMENT = SURVEY SYSTEM, OR STATISTICAL INFORMATION SYSTEM
SURVEY DESIGN AND CONTROL	documentation; partial computations and analyses	selection of sampling strategy; complete EDP systems design	STATISTICAL DESIGN AND EDP SYSTEMS DESIGN
SURVEY OPERATION	data entry, coding, editing, tabulation	coding, editing, statistical analysis	PERFORMING THE STEPS IN THE PROCESSING OF A SURVEY
	THE COMPUTER AS AN EFFICIENT CALCULATOR AND PEDANTIC BOOK-KEEPER	THE COMPUTER AS AN AMPLIFIER OF THE HUMAN INTELLECT	

Figure 2. A cross-classification of the tasks of a statistics production system.

Figure 2 shows a cross-classification of the tasks of statistics production. In one dimension the tasks have been classified in accordance with the system/-subsystem structure of figure 1. In the other dimension we distinguish between those tasks which are of a more or less routine character, and those which require more "intelligence". Many tasks in a statistical office are of a rather routine nature. Data are entered, edited, sorted, counted, and presented in a fairly straightforward way. The computations are not always very complicated, but the volumes of data are sometimes quite large. Under these circumstances the computer is little more than a pedantic, incredibly efficient book-keeper, who makes no errors. Nevertheless, it has turned out to be good enough to save large amounts of money for statistical agencies.

Thus so far we have been very successful in multiplying the human being's capability to move and sort data, and to count them, and to eliminate the human tendency to commit errors in those operations. But can we also use the computer as an amplifier of the human intellect in statistical work? Without exaggerating the possibilities of disciplines with fancy names like "artificial intelligence" and "expert systems", I think that there are many good opportunities of using knowledge-based methods in statistics production, and I think that such methods could actively support a sensible and efficient decentralization of tasks within the framework of an integrated statistical system.

Generally speaking, contemporary computer applications in statistics production mostly fall into the bottom-left category of the classification in figure 2. Thus in order to locate and explore the unused potentials we should move upwards and to the right in the diagram.

3 Decentralization, integration, and standardization

No discussion today about computerization will last very long before the words 'decentralization', 'integration', and 'standardization' have been mentioned. Naturally these words are buzz-words. They are sometimes used to disguise truisms, sometimes for propagandistic purposes, and sometimes even to confuse and mislead. However, if we take them seriously, they also represent some important trends in the present development of computer technology and its applications. I will briefly discuss how these trends may affect the organization of statistics production, and how the three trends are actually related to each other.

3.1 Decentralization of computers and computer-related resources

The computer technological development has surpassed all expectations. Computers with a capacity of yesterday's mainframes are now available on everyone's desk, and they cost little more than a typewriter. Thus we are now able to buy computer resources in small pieces and at a very low cost per piece. We hardly any longer need a centralized organization for the reason of sharing expensive, indivisible, and scarce resources. Instead we can integrate the decisions concerning computer resources with other important decisions in the statistical office and try to develop the same type of "balanced decentralization" of decision-making as in other areas, letting those responsible for a statistical survey take as full responsibility as possible for all types of resources needed in the design and operation of the survey.

In Sweden we have taken some important steps in this direction. On the national level those governmental agencies which have enough competence and experience in EDP are now allowed to take more computer-related decisions without having to ask any other agency or the minstry of finance, provided, of course, that the agency is able to handle all decisions within its given budget.

Within Statistics Sweden we are also trying to treat EDP-related decisions,

not separately, but integrated with other decisions. We have established an EDP policy, which will of course be updated from time to time, and within this policy, each department is authorized to take its own decisions, as long as they are within the budget of the department. Thus, as far as possible, all types of costs - for hardware, software, and personnel, for mainframe-related resources and for micros - are measured in "the same kind of money".

Already five years ago Statistics Sweden decentralized the personnel resources for application systems development and maintenance to the subject matter divisions. On the central level a new division for research and development was formed, which retained the responsibility for such functions as development and maintenance of generalized software, research and development in the area of statistical data processing, and EDP training.

One good effect of this decentralization is that the manager of a particular statistical survey has now much more complete overview, knowledge, and control of his/her product and all types of resources that are needed, assuming of course that the manager has the capacity and willingness to make use of these opportunities. On the other hand there is naturally a risk that the statistical office as a whole will fall apart into a large number of small, uncoordinated survey based organizations. In order to prevent this, a number of specialized "councils" (among others one for EDP) have been created for giving specialized advice in policy matters etc to the top management and the Director General of the office.

3.2 Integration of different types of tasks, skills, and competences

Integration of different tasks, skills, and competences is the other side of the "decentralization coin". The effects of this integration are becoming visible throughout the organization. Managers are losing their personal secretaries and administrative assistants, who find themselves replaced by personal computers, networks, and office information systems. Subject matter statisticians are taking care of application development and maintenance without the assistance of systems analysts and programmers, and they do some of their own publishing without having to rely on typographers. Interviewers take over data entry and data editing tasks.

All these types of integration have been enabled and facilitated by the decentralization of computers and computer-related resources. The integration has a number of good effects, including job enlargement, shorter communication and decision paths, less administrative overhead, and a more clear division of responsibilities. On the negative side there is a risk of "happy amateurism" replacing competent and efficient professionalism, and of isolationism and self-conceit in the relatively small and independent organizational units. However, on the whole the positive effects seem to outweigh the negative ones, and there seem to be more staff members who feel they have gained from the development than who feel they have lost. Even some of those who have lost responsibilities and empires welcome or accept the development as being basically sound and find new roles in the organization relatively quickly.

Naturally, there will always be a need for good specialists in several fields of competence in a statistical office. The on-going technological development only eliminates a need for centralization and functionalization that was based on the indivisibility of large, expensive computers, and on the relative scarcity of systems analysts and programmers.

3.3 Standardization of technology, software, and methodology

Policies and standards are wellknown tools for achieving and maintaining a desirable degree of uniformity and integration in decentralized systems.

However, the effort to decentralize is not the only good reason for standardized tools and procedures in statistics production. Computers are now so cheap, and people so expensive, that it is very rarely worthwhile to aim at maximum technical efficiency in the design of a computerized information system. Most statistical surveys are small or modest in size, they are processed rather infrequently, and response time requirements are often quite moderate in comparison with those of many commercial on-line systems of administrative character. Furthermore, even in those few cases where technical optimization is really optimal, from an executive point of view, it is usually not executively optimal to technically optimize the system as a whole, but only some limited part or aspect of it.

4 The future architecture of statistical information systems

4.1 Is there a need for a statistical office any longer?

As noted earlier the technological development has alleviated one restriction on statistics production that used to exist: the necessity to share scarce and expensive computers and computer-related resources. This has started a decentralization process. Are there any natural limits to this development, or will the decentralization stop only when the statistical office has been dissolved into a number of separate statistical surveys? Will there be a role to play for a statistical office as an independent organizational entity in the future, or will the statistical surveys be taken care of by other governmental agencies?

Personally I do believe that statistical offices have an important role to play in society, quite regardless of the decentralization possibilities that the technological development is now offering, but I think that we need to ask those critical questions indicated above. Others will do it.

4.2 User needs

The needs for coordination and integration are deeply and directly founded in some strongly felt user needs. Users of statistics are rightfully irritated when they have difficulties to locate and interpret the statistical data they are looking for, and they are not happy when they have to go to several places in order to get all the data they need, instead of getting everything in one place, including some advice about how to combine data from different sources.

Furthermore there is a growing number of rather advanced users of statistics, with more or less sophisticated models and hypotheses that they would like to try on official statistical data, and sometimes combine with their own data. Due to the technological development these users will always have access to powerful computer equipment of their own, and they have a good understanding of the possibilities offered by modern technology. If they are not well served by the statistical office, they will exercise all the rights that they may have to obtain statistical data in rather "raw" form from the statistical office, and use them together with their own data, software, and models in data laboratories that they build and run independently of the statistical office. If a statistical office wants to be successful in this competition it must be active, imaginative, and flexible, and it must use its relative advantages in methodological competence, and coordination and integration possibilities.

4.3 Needs for rethinking?

The statistical survey is the basic building block in statistical organizations. The on-going decentralization will further strengthen the power and control of individual statistical surveys. From a managerial point of view,

this development has its advantages. It clarifies responsibilities within the organization, and the person in charge of a statistical survey will not have so many others to blame, if something goes wrong.

On the other hand the user needs discussed above call for other organizational solutions. In order to make it easy to locate and interpret data, all statistical data of any importance must be well documented, and they must be documented in the same way from survey to survey. Thus all statistical data must be accompanied by appropriate metadata. The metadata must be computerized, and like the data themselves they must be organized in accordance with uniform database principles. Many statistical offices have tried to implement these ideas in different ways, but the results are not always encouraging. Better results will be required.

But not even well documented databases is enough for the users. They want to combine data emanating from different surveys and other sources, like administrative registers. Traditionally, statistical offices have been able to hide behind their publications. A statistical survey is responsible for the contents of the reports and publications that it publishes, but it takes no responsibility for how the user may possibly combine data from different publications. This strategy will not be maintainable in the era of new technology.

Obviously statistical offices must activate themselves in the area of standardized concepts and classifications. This is not always fully appreciated by those in charge of individual surveys, and sometimes not even by their managers. On the other hand people outside a statistical office seem to accept the office's responsibility and welcomes its competence in this area.

Unified concepts and classifications is an excellent basis for combining data and putting them into models. However, there will always be cases where complete standardization is not possible. For example, due to the different purposes of an administrative register and a statistical survey, it is inevitable that all definitions cannot be harmonized between them. Experts in statistical offices should assist in finding ways around the problems, exploiting the constructive power of statistical method.

These examples show that there is a need for managerial action and control that is global in relation to the individual surveys. One may go one step further and say that the new problems and possibilities call for a new survey concept. One effect of modern technology is that the ties between input and output will be weakened, both physically, logically, and in time. The statistical end-products and typical usages of statistics will be based on combinations of input data from many different sources, and the data collected by one statistical survey will be used for many different purposes, by different users, and at different points of time.

Thus, if we look upon a statistical survey as a basic building block of statistical organizations and statistical information systems, it may be more adequate to think in terms of three different types of surveys:

* input-oriented surveys, collecting and editing the data, performing some routine tabulations and analyses, and preparing the data for future use by putting them, with their accompanying metadata, into common databases;

* common databases, taking care of data from different input-oriented surveys, and forming the basis for output-oriented surveys;

* output-oriented surveys, making use of existing data in common databases and other sources, inside and outside the statistical office.

R & D Reports är en för U/ADB och U/STM gemensam publikationsserie som fr o m 1988-01-01 ersätter de tidigare "gula" och "gröna" serierna. I serien ingår även **Abstracts** (sammanfattning av metodrapporter från SCB).

R & D Reports Statistics Sweden, are published by the Department of Research & Development within Statistics Sweden. Reports dealing with statistical methods have green (grön) covers. Reports dealing with EDP methods have yellow (gul) covers. In addition, abstracts are published three times a year (light brown /beige/ covers).

Reports published earlier during 1989 are:

1989:1 (grön) Går det att mäta produktivitetsutvecklingen för SCB? **(Rune Sandström)**

1989:2 (grön) Slutrapporter från U-avdelningens översyn av HINK och KPI **(flera författare)**

1989:3 (grön) A Cohort Model for Analyzing and Projecting Fertility by Birth Order **(Sten Martinelle)**

1989:4 **Abstracts I** - Sammanfattningar av metodrapporter från SCB

1989:5 (gul) On the use of Semantic Models for specifying Information Needs **(Erik Malmborg)**

1989:6 (grön) On Testing for Symmetry in Business Cycles **(Anders Westlund and Sven Öhlén)**

1989:7 (grön) Design and quality of the Swedish Family Expenditure Survey **(Håkan L Lindström, Hans Lindkvist and Hans Näsholm)**

1989:8 (grön) Om utnyttjande av urvalsdesignen vid regressionsanalys av surveydata **(Lennart Nordberg)**

1989:9 (grön) Variations in the Age-Pattern of Fertility in Sweden Around 1986 **(Michael Hartmann)**

1989:10 (grön) An Application of Generalized Precision Functions in the 1985 Swedish Family Expenditure Survey **(Håkan L Lindström and Peter Lundquist)**

Kvarvarande beige och gröna exemplar av ovanstående promemorior kan rekvireras från Elisabet Klingberg, U/STM, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 41 78.

Kvarvarande gula exemplar kan rekvireras från Ingvar Andersson, U/ADB, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 41 47.