UNITED NATIONS

ECONOMIC AND SOCIAL COUNCIL

Distr.
GENERAL

E/CN.3/406
14 October 1969

ORIGINAL: ENGLISH

STATISTICAL COMMISSION
Sixteenth session
Item 12 of the provisional agenda

DRAFT RECOMMENDATIONS FOR THE 1973 WORLD PROGRAMME OF INDUSTRIAL STATISTICS

PART III. ORGANIZATION AND CONDUCT OF INDUSTRIAL CENSUSES

Report of the Secretary-General

.....1. The attached document (ST/STAT/46) has been prepared in response to a request by the Commission at its fifteenth session. It constitutes the third and final document (Part III) in a series intended to define the objectives of the 1973 World Programme of Industrial Statistics and to indicate appropriate methods of carrying out the related inquiries in each country. Parts I and II of the series (E/CN.3/404 and E/CN.3/405) deal respectively with the general statistical objectives of the Programme and the individual commodity detail which is proposed for collection.

2. The three documents in the series were distributed to countries and to interested regional and international agencies for comments. Summaries of the comments received and of the proposals of the Secretary-General for the preparation of definitive editions of the documents are before the Commission as E/CN.3/404/Add.1, E/CN.3/405/Add.1 and E/CN.3/406/Add.1.

UNITED NATIONS

SECRETARIAT

Distr.
LIMITED

ST/STAT/46
14 October 1969

ORIGINAL: ENGLISH

STATISTICAL OFFICE OF THE
UNITED NATIONS

DRAFT RECOMMENDATIONS

FOR THE

1973 WORLD PROGRAMME OF INDUSTRIAL STATISTICS

PART III. ORGANIZATION AND CONDUCT OF INDUSTRIAL CENSUSES

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

I. INTRODUCTION

General remarks

1. This Manual is intended as a practical guide to the planning and management of industrial censuses, with primary emphasis on the problems faced by countries at an early stage of statistical development. The authority for the preparation of the Manual is derived from resolution 7(XV) of the United Nations Statistical Commission which asks the Secretary-General, inter alia, "to prepare, with expert help and following regional consultation, a manual on the practical problems of conducting industrial inquiries as a supplement to the recommendations for the 1973 World Programme of Industrial Statistics..." This resolution, together with the observations of the Commission on the proposed Programme as a whole, will be found in the report of the fifteenth session of the Statistical Commission (26 February - 8 March 1968) (E/4471).

2. The Manual is the third in a series of documents designed specifically to present the United Nations recommendations for the 1973 Programme. The first (ST/STAT/44) defines the statistical objectives of the Programme and is based on the general recommendations for infrequent industrial inquiries contained in International Recommendations for Industrial Statistics [1]. The second (ST/STAT/45) establishes a standard list of individual commodities for which, it is recommended, production and consumption data should be compiled for 1973. Other publications which provide reference material of direct or indirect value to the inquiry are listed in the bibliography to the present document (chapter VIII).

3. The Manual may be regarded to some extent as a revised and up-dated version of certain parts of the study Industrial Censuses and Related Enquiries published in October 1953 [2]. The scope of the Manual is, however, more restricted than the scope of the earlier study which encompassed repetitive monthly, quarterly, and annual surveys as well as industrial censuses, and which included detailed recommendations on the content of census inquiries.

1/ Chapter VIII: Reference 1.
2/ Chapter VIII: Reference 2.

4. For the purposes of the Manual, an industrial census is conceived as a comprehensive survey conducted at infrequent intervals, usually ten years. The survey encompasses the activities of mining, manufacturing, and the supply of electricity, gas and water as defined by ISIC [3] Major Divisions 2, 3 and 4 [4]. The basic units for which statistics are compiled are establishments such as mines, quarries, factories, mills, works and plants. Enterprises are not the basic statistical units although in some cases they may be the basic reporting units and certain statistics may be compiled on an enterprise as well as an establishment basis. The inquiries usually cover a period of one year. The principal kinds of information collected concerning establishments are generally: [5]

Kind of activity (industry)

Location

Type of ownership

Type of legal organization

Type of economic organization

Number of employees and other persons engaged in industrial activity

Compensation of employees

Cost of goods received and consumed, and payments for industrial services rendered

Electricity purchased, generated, consumed and sold

Acquisitions of fixed assets (structures and equipment)

Capacity of power equipment

Inventories

Products shipped or produced, and receipts for industrial services rendered to others.

5. The first industrial census resembling the modern type was conducted in the United States for the year 1810 though, by modern standards, this census

---

3/ Chapter VIII: Reference 3.

4/ Construction activity, which in earlier recommendations was considered part of industrial activity, is now the subject of separate recommendations. See International Recommendations for Construction Statistics, United Nations publication, Sales No.: E.68.XVII.11.

5/ The conceptual framework recommended for the 1973 inquiries is outlined in ST/STAT/44: General Statistical Objectives of the 1973 World Programme of Industrial Statistics.

was not a success [6]. Since this first attempt, many countries have conducted industrial censuses, and much of the experience of the government agencies managing them has been recorded. Today, a country contemplating its first industrial census can take advantage of a considerable body of technical advice to assist it in planning and managing the inquiry. Some of the more important precepts based on the experience of many countries are summarized in the following paragraphs.

Planning and use of funds

6. The importance of advance planning of the collection, editing, coding, tabulation and publication of the data cannot be over-emphasized. Much money and effort have been expended in trying to correct errors which could have been avoided by careful planning.

7. Because of poor planning, some countries have found that their funds were largely expended in the collection phase and no resources were left to complete the processing of questionnaires and publication of results.

8. The development of a unique, well-thought-out plan adapted to the needs of the particular country is important. Statistical systems which work well in one country may be quite inadequate to meet the needs of another.

Co-operation of respondents

9. The cost per questionnaire of collecting accurate figures is considerably greater than the cost of compiling and publishing the resulting statistics. The co-operation of respondents is an important factor in the cost of collection: for example, several field calls may be required to obtain information from a reluctant respondent, while a friendly and co-operative respondent will supply the information on the first visit. The provision of advance publicity and the training of enumerators to explain the merits of the census convincingly are important ways of increasing the co-operation of respondents. Legal compulsion to report is also important but does not prevent unco-operative respondents from reporting carelessly estimated or false figures.

6/ There were earlier inquiries relating to particular industries and commodities but the 1810 Census was the first in which an attempt was made at a comprehensive coverage of establishments.

Limiting requests for information

10. There is widespread awareness today of the need for industrial statistics in economic and social planning. The potential data users, however, are often unaware of the problems attending the collection and compilation of statistics, and may suggest collecting information which is difficult or impossible to obtain. The number of queries on a questionnaire also may make it so long and detailed that it arouses opposition among respondents merely because of its length and the time required to complete it. A long questionnaire may add unduly to the cost of editing and tabulation. Pilot studies or pre-tests are usually effective in determining whether these faults exist. Such studies need not be costly or elaborate. Often the information obtained from interviewing a small number of respondents under actual field conditions will be of immense value in deciding finally on the content and wording of questionnaires.

Checking as operations proceed

11. The development of a feasible plan is not enough to ensure the success of an operation. Good management requires continued checking as the plan is executed. This applies to field enumeration as well as to all subsequent phases of the industrial census operations.

Industrial directory

12. The establishment of an industrial directory in advance of the actual enumeration is discussed in chapter V. Such a directory, even though it may not be possible to include all industrial establishments, can be of great assistance in conducting the census. It will facilitate the checking of field enumeration, the control of coverage in the collection phase, and will aid in ensuring completeness at all stages. If sampling is employed, prior establishment of a directory is essential.

Sampling

13. Although the collection of data from a well-designed sample of establishments can reduce over-all costs, sampling should not be regarded as a panacea. With a relatively inexperienced staff, complex sampling plans should be avoided: they are difficult to manage and will produce unacceptable results if poorly executed. A simple plan, such as the coverage of all establishments above a

certain size and a random sample of the smaller establishments, may be preferable even though it is not an optimum sample in a theoretical sense.

How the Manual should be used

14. In the brief compass of this document, it has not been possible to discuss every subject in adequate detail. The recommendations included are intended to suggest topics which will need further consideration and exploration in relation to the specific inquiries planned by each country. Those interested in the more detailed discussion of certain topics or in the experience of various countries in this work, will find the bibliographical references in chapter VIII useful.7/ Advice on particular problems may also be obtained by writing to the Director, Statistical Office, United Nations, New York.

7/ A comprehensive bibliography will appear in the definitive edition of this document.

II. NEED FOR INDUSTRIAL CENSUS STATISTICS

General remarks

15. This chapter describes some of the more important uses of the information sought in industrial censuses. Consideration of these uses will assist industrial census planners not only in selecting the items to be covered in the census and the emphasis to be given to each, but also in preparing publicity and in answering inquiries about the purposes and usefulness of the census. Particularly in the case of small industrial enterprises, there is little awareness that information of this kind can be of great value in formulating the national economic policies of a country and that, in many instances, it can directly assist industry managers to maximize the effectiveness of their decisions.

16. The establishment and growth of activities such as mining, manufacturing and electric power supply has nearly always been accompanied by a demand for facts about the structure of these activities as well as the nature of their current operations. The role of governments in collecting and disseminating such facts has varied, but in all countries where periodic comprehensive censuses have been conducted, government organizations have assumed a dominant role. Private groups such as manufacturer's associations, which have usually been limited in their scope and authority, have not in general been able to conduct censuses.

17. As products, processes and organizations have grown more complex, the demand for statistics has increased. In some industrialized countries, statistical collection has grown to such an extent that much attention must be paid to the value of the information collected relative to the cost of providing it. In the case of government-sponsored statistical projects, it is important to keep in mind that the cost to industry management of supplying information to enumerators or of filling out mail questionnaires may be as great or greater than the cost to the government of collection, compilation and publication. Careful evaluation of the uses to be made of the data thus can be of very great importance. From the point of view of the statistical agency, it is also desirable to improve its relations with industry management.

A well-thought-out statement on the usefulness of the statistics to be compiled will go far towards convincing industry management of the need for co-operation.

18. It has long been recognized that small enterprises tend to keep less detailed records than do large ones and thus have more difficulty in supplying census-type information. It may also be observed that large firms are better equipped to use statistics and thus are better able to recognize the usefulness of a census and more apt to co-operate with the census enumerator, or to fill out a mail questionnaire completely and conscientiously. Imposing severe limits on the questions to be asked of small enterprises is, therefore, doubly important.

Characteristics of establishments

19. The structure of an industry is usually described in terms of the characteristics of its establishments. Their number, size, location, integration of processes, product mix, and equipment capacity are aspects of structure, as are other characteristics such as type of ownership or control, legal organization, and methods of distributing products.

20. Knowledge of the structure of industry is an essential part of the factual background needed for understanding a particular economic situation. As such, it is a guide leading to intelligent and realistic government decisions affecting industry when establishing measures of potential markets in terms of areas, kinds of business, and kinds of products. Since industrial structure changes slowly, quinquennial or decennial censuses usually provide this type of information with sufficient frequency for general use.

Employment, man-hours and earnings

21. Governments are interested in programmes which promote economic and social welfare, and current facts about employment and earnings are often essential indicators of the progress of such programmes. Recognizing this importance, the International Recommendations for Industrial Statistics [1/] includes recommendations for the collection of employment and earnings data in current industrial inquiries in all countries. A major role of the industrial census in this connexion is to provide benchmark data and mailing lists for use in inquiries conducted on a more frequent basis.

1/ Chapter VIII: Reference 1.

22. A further role of the industrial census in this area is in the development of attempts to relate labour input to value added by manufacture, value of products, and other measures of industrial activity. Productivity studies usually originate with data from industrial censuses, and such studies are needed to estimate the additional workers demanded by contemplated expansions in industry, and to judge the feasibility of such expansions. In addition, data on labour productivity provide clues to the extent of automation or mechanization among industries.

23. Data on age, sex and other characteristics of the industrial labour force are often useful in connexion with social programmes. The association of labour force data with the characteristics of establishments contributes importantly to an understanding of industrial structure. If a single measure of size is used in structural distributions, the number of persons engaged is generally to be preferred.

24. Other requirements for data on the characteristics of the labour force arise from the formulation and administration of labour laws, and negotiations concerning wage and salary rates. As a basis for labour legislation, data may be required on numbers engaged and earnings for special groups, such as women, children, unpaid family workers, homeworkers and ethnic groups. Though current figures are usually needed for collective bargaining and other efforts to set wage rates, these may be based on sample surveys relying on industrial census coverage for their validity.

Production or shipments; value added

25. Industrial census data on the quantity and value of particular commodities are needed in analyses of supply and demand. Such analyses are useful in determining national policy in foreign trade matters, in studying the relation of industrial activities to the national economy as a whole, and in dealing with surpluses or deficits in the supply of commodities and related price problems.

26. Measurement of gross output, or the total value of goods and services produced, is an essential component in the calculation of census value added. Value added in its elementary form is the difference between the gross output of producers and the cost of the related materials and industrial services

consumed. It provides a useful way of measuring without duplication the economic importance of an industry or an industrial sector. When comparisons are made among industries or among geographic areas, value added provides a convenient assessment of their relative significance.

27. It may be noted that the national accounts definition of value added differs from value added as generally calculated from industrial census data.[2/] In the national accounts sense, the four components of value added are compensation of employees, operating surplus, consumption of fixed capital, and net indirect taxes.[3/] In order to derive value added in the national accounts sense from census value added, it is usually necessary to have data on the non-industrial services purchased or sold by establishments. A further adjustment with respect to indirect taxes and operating subsidies may be necessary.

28. Data on the value of the goods and services produced as well as on the cost of the goods and services consumed in their production are essential components of input-output tables. Such tables, which normally measure in some detail the flow of goods and services into, out of, and among the various branches of industry, are of value in economic planning and programming; they are closely related to the more advanced systems of national accounts. In an economy of even moderate complexity, the effect of a possible change in demand such as the diversion of certain types of production from military to civilian use, can most effectively be measured by use of input-output tables.

Materials and industrial services

29. The uses of these data in calculating value added by manufacture and input-output analysis have been mentioned in the preceding paragraphs. In addition, statistics on the consumption of specific materials distributed by industry or by geographic area are of value to industry management in the sectors where the materials originate. Reliable market information of this type is not usually obtainable except from industrial censuses.

2/ A discussion of this point appears in paras. 155-159 of International Recommendations for Industrial Statistics (chapter VIII: Reference 1).

3/ A System of National Accounts, United Nations publication, Sales No.: 69.XVII.3, para. 2.40.

Fixed assets

30. Since the level of fixed capital formation, particularly in the mining and manufacturing sectors, is a sensitive indicator of changes in the level of economic activity in general, current information on investment in plant and equipment attracts great interest. The net change in the total value of fixed assets holds similar interest. Industrial censuses normally provide benchmarks for both these measures and mailing lists for use in sample surveys designed to collect related current information.

Inventories

31. In addition to use in connexion with studies of national wealth, data on inventory levels provided by industrial censuses are useful as benchmarks and as mailing-list sources for more frequent surveys. Such surveys, usually covering specific products or materials of considerable importance to a country's economy, are valuable in connexion with economic analysis, particularly short-term forecasting. Data on changes in inventory levels between the beginning and end of the census year are necessary for the calculation of gross output and value added, when information on products and materials is collected on a shipped or received basis. 4/

Capacity of power equipment

32. Since power equipment (that is, prime movers, electric motors and generators) is a key industrial resource, many governments need to have information on its capacity and availability by industry. Industrial censuses are usually the only source for statistics on power equipment, since changes are gradual and frequent inquiries are not ordinarily needed.

4/ See chapter VIII: Reference 1, paras. 151-159.

III. MAJOR PROBLEMS

General remarks

33. In all countries with substantial numbers of industrial establishments,, industrial censuses are large-scale, complex and difficult undertakings. Countries with little prior experience in conducting such comprehensive surveys face many problems in planning and carrying them through successfully. The statistical agency, if one exists, will have only limited knowledge of the problems involved and of the industries from which information is to be collected. Industrial enterprises will have had little or no experience in furnishing statistics and will frequently have little appreciation of the value of economic data. The structure of the industrial sectors further complicates the task in many countries. In industrialized countries, much of the production is centred in large establishments. This tends to facilitate data collection because the statistical agency can concentrate its efforts on securing adequate reports from the relatively few large establishments. In developing countries, on the other hand, small establishments and households may make important contributions to certain industries. Not only does information need to be gathered from great numbers of scattered respondents, but it is also difficult to collect, because small-scale operations can often be managed with very little in the way of systematic records and consequently the owner must be persuaded to estimate a substantial part of the data. The situation is further complicated by the fact that industrial production in small establishments is often intermittent or carried on simultaneously with other activities.

34. Some of the more important problems likely to be encountered by countries with little industrial survey experience, and suggestions which may aid in their solution, are described in this chapter. The discussion of these problems is largely based on the experience of countries in carrying out the 1963 World Programme of Basic Industrial Statistics. 1/

1/ See International Recommendations on the 1963 World Programme of Basic Industrial Statistics, United Nations publication, Sales No.: 60.XVII.8.Add.1.

Problems involving small establishments

35. In the 1963 Programme, several countries reported that respondents were reluctant to provide the requested data for fear they would be used for tax purposes. This was particularly evident in one country where an investigation for taxation purposes was being carried out at the same time as the census, and in another country where an announcement by the government that it intended to increase tax revenues was made at about the time of the census enumeration. Other objections arose because the respondents were under the impression that the census was a useless undertaking, designed only to create government jobs, and that the reports they furnished would be filed away with no further use being made of them. The concept of economic data being used to benefit the economy in any way was foreign to many respondents. Most of these objections came from the smaller enterprises, since the managers of larger firms tended to be more knowledgeable about the value of economic data.

36. These can be very serious problems and should not be treated lightly. Even though response to census inquiries is mandatory under the law, the opposition of respondents can greatly increase the costs of enumeration and reduce the quality of the responses. Well in advance of census enumeration, the statistical agency should try to evaluate the nature and extent of the response problems by discussing them with industry representatives and with government officials who have expert knowledge of industries, and even by sending its own representatives to talk to potential respondents selected at random.

37. In overcoming the reluctance of respondents, the principal methods at the disposal of the statistical agency are advance publicity and the training of enumerators. The planning of a publicity campaign is discussed in chapter IV. The training of enumerators should prepare them to create a favourable atmosphere for the interview by a convincing exposition of the merits of the industrial census. They should be prepared to answer the most usual questions about the census frankly and completely, to help gain the confidence of the respondents.

38. If it is judged in advance that the response problems among small respondents will be very troublesome, consideration should be given to the compilation

of an industrial directory and to the selection of a sample to represent this group instead of complete enumeration.[2/] The difficulty of obtaining data to represent a group of small respondents would obviously be greatly reduced if answers to the full set of inquiries were required of only one out of every ten, while the entire group would be approached to obtain only the very few items of information required for the industrial directory, such as name, location, activity and number of persons engaged.

Use of population census experience

39. In some of the countries undertaking industrial censuses in the 1963 Programme, the previous census experience was mainly in relation to population censuses. The organization was usually pyramidal in form, with a chain of command extending from the central office through provincial and local supervisors to the enumerators. The training of personnel in such a scheme filters down from the vertex to the base and the data obtained are transmitted from the base to the vertex, in successive stages.

40. There are, however, marked differences between population and industrial censuses. In the population census, the statistical units (the individual persons) are all similar, while in an industrial census there are great differences among the statistical units. In the population census, the statistical agency is organized to handle great numbers of similar questionnaires with fairly rudimentary data; in the industrial census, different questionnaires are frequently used, some complex and others relatively simple. The enumerator of household industries need not have the same qualifications as the enumerator of a large manufacturing establishment. Enumeration of the population must be completed in a very short period; in an industrial census, enumeration can extend over several months with no great detriment to the result. While some of the experience gained in conducting population censuses can be useful in industrial censuses, there are many differences, and adaptability of the organizational structure should be considered carefully.

2/ Sampling and the industrial directory are discussed in chapter IV. More specific recommendations on the application of sampling to the 1973 inquiries are given in ST/STAT/44.

41. The foregoing discussion should not be construed to mean that population censuses are entirely unrelated to industrial censuses. As mentioned in later chapters, two possible conceptual links are (1) a question asked in the population census can aid in locating household industries for the industrial directory and (2) the industry classification system used in the population census should be compatible with that used in the industrial censuses. Sampling methods, clerical processing and tabulation may also have many similarities.

Over-ambitious programmes

42. There are undoubted advantages in conducting an industrial census in conjunction with inquiries into other sectors, for instance, the distributive trades. On the other hand, experience tends to show that a relatively inexperienced statistical agency is not in a position to cope with all of the resulting problems within the limits of the time and resources available. Consequently, the agency's experience and other assets should be weighed carefully in relation to the advantages and disadvantages of combining censuses of several economic sectors in one programme. Feasible alternatives might be to carry out the programme in successive years or to collect all the data at one time but divide the responsibility for the post-enumeration stages among the government agencies directly concerned with the various sectors. Such a division of responsibility is more likely to succeed if carefully planned in advance of the enumeration, with the participation of all the agencies involved.

43. Some countries have tried to use their first comprehensive industrial census to obtain as much information as possible from every establishment. While this reflects legitimate needs for economic data, the problems of locating establishments and obtaining co-operation are often so serious that the added burden of a long and complicated set of questions may lead to failure. This tendency to length and complexity has in some instances been made worse by questions that are unrealistic or badly formulated. To expect large numbers of small establishments to report in detail such information as the kinds of materials used is apt to be unrealistic, may cause resentment, and will not usually produce meaningful statistics. Long and complex questionnaires add to the cost of all subsequent stages of processing also, and unrealistic inquiries can absorb much energy before they are finally abandoned. Adequate

pre-testing of questionnaires in the field will go a long way towards solving such problems. In industrial censuses, such pre-tests are usually much less formal than the types conducted in connexion with population censuses, and consequently are less costly.

Problems of measurement

44. The international recommendations for infrequent industrial inquiries in the statistically less developed countries suggest that such a country's first census, or the first in many years, should collect information for small establishments relating to kind of activity, number of persons engaged, wages and salaries, total cost of materials and industrial services, expenditures for fixed assets, and total value of products. It is not easy, however, for small establishments to furnish satisfactory statistics for even those few items. This applies especially to household industries where separation of data between manufacturing activities on the one hand, and household maintenance on the other, offers problems. The distinction between investment in fixed assets (capital expenditures) and current costs such as those for maintenance and repair is also frequently difficult to determine in practice. Other data which may be hard to define include the value of output of manufactured goods and the cost of raw materials and fuels. Where the same enterprise includes both agricultural and manufacturing activity, such as the production of grain and the production of flour or meal from grain, the cost of raw materials used (grain) is usually not a matter of record and must be estimated for the census inquiry.

45. The principal problems of measurement peculiar to the country's economy, as well as problems associated with the wording of the questionnaires and instructions, may be brought to light in advance of the census by means of small-scale pilot studies or informal pre-tests. These should be followed by adequate training of enumerators to deal effectively with these problems in the field.

IV. ADVANCE PLANNING FOR AN INDUSTRIAL CENSUS

General remarks

46. The success or failure of an industrial census depends in large part on the care taken in formulating its broad objectives and in the detailed planning of each phase of data collection, review and correction of the data reported, and the tabulation and publication of the results. The experience of countries with a long tradition of industrial censuses has shown that advance planning for each census is essential, even though several successful censuses may have been conducted previously. Countries with little experience should plan with even greater care, because their resources are usually limited and there may not be sufficient funds available to correct the effect of a serious misjudgement. Even if funds are obtainable, the necessary additional manpower may not be, and as a consequence, there may be unacceptable delays in the completion of the work. Advance planning is especially important for a country conducting its first industrial census, because the effect of an unsuccessful attempt may be the refusal of the government to sponsor similar projects in the future.

47. The experience of the Latin American countries in carrying out the 1963 Programme has demonstrated that planning is of first importance to avoid inquiries that are unrealistic or too ambitious. In the 1963 Latin American inquiries, some countries tried to obtain the full range of information from all statistical units, not differentiating between large and small establishments. The result was that a high percentage of the replies from small establishments were either missing or grossly inaccurate. This situation could have been avoided if the agency had planned to use a short form with very few questions for the small establishments.

48. Many problems will arise in conducting an industrial census, and even if the inquiries are restricted to the bare essentials, the planning staff will not be able to anticipate all of them. Statistical agencies with considerable experience in conducting censuses sometimes encounter unexpected difficulties. To conduct a census without a plan, however, means that the problems would be greatly increased. The effect of good planning is to minimize the cost of correcting minor errors and to hold the more serious mistakes to a manageable few.

Fundamental decisions

49. There are several basic matters which must be decided, at least tentatively, before detailed plans can be formulated. These are the legislative requirements, the statistical and reporting units to be used, the industries and establishments to be covered, and the assignment within the government of responsibility for conducting the census. They are briefly examined in the following paragraphs.

Legislative requirements

50. The legal authority for the industrial census is fundamental because it gives the project the stability and permanence needed to carry it to a successful conclusion. It will also help to assure the co-operation of respondents. The kind of legal authority required depends to a large extent on the legal traditions of the country. Industrial censuses have generally been conducted under three types of authority:

> Executive decrees. Where wide powers to gather economic information are already vested in a government official, a decree or executive order would carry sufficient authority. The decrees issued have usually specified the nature of the inquiries to be made, the obligations of respondents to furnish information, and the responsibilities of the government in connexion with the individual returns and in compiling and publishing statistics.
>
> General statistical acts. Such acts pertain to statistical systems rather than to censuses alone, and may encompass population surveys as well as economic surveys, and current data as well as censuses. The acts usually have clauses on the obligations of respondents and the government with respect to them. To provide for specific industrial censuses, further action in the form of a decree or other executive action may be necessary.
>
> Industrial census acts. These deal only with industrial censuses but in other respects are similar in their provisions to the more general acts. They differ as to the detail in which the provisions appear; for those that are not sufficiently detailed, further executive action may be necessary.

If feasible, the legislative authority should not be so narrow in its provisions as to preclude or inhibit subsequent progress towards a complete industrial statistics system.

51. Other points that should be considered in formulating legislative authority include the degree of flexibility to be allowed the statistical

organization, or the government agency of which it is a part, in conducting the census. Some legislation is very detailed as to the content of questionnaires, the exact time period to be covered, the types of statistical reporting, and other more or less technical matters. Such legislation tends to hamper subsequent planning, and nullify the advice of advisory committees (see paras. 63 - 66). Even if the provisions are adequate for the immediate project, they tend to become obsolete because of changes in the economy or in the needs for data. Laws which extend brpader powers of decision in technical matters to the agencies conducting the census are less apt to hamper the successful completion of the census or to interfere with the subsequent establishment of a general system of inquiries.

Statistical unit

52. The recommendations in this Manual are based on the assumption that the establishment is the statistical unit, that is, the unit to which the information collected relates. The reports of multi-establishment enterprises may originate with the headquarters of the enterprise, but a separate report will be made for each establishment and it will be treated in subsequent operations as the basic statistical unit.

Industries to be covered

53. A decision on the industries to be covered is fundamental because it determines the funds and personnel required to conduct the census and the burden on respondents, and may also affect the decision on which government organization shall be responsible for the various phases of the work.

54. The international recommendations for an infrequent industrial inquiry [1] provide for coverage of establishments engaged primarily in mining, manufacturing, and electricity, gas and water supply.[2] Those planning the census may well be confronted with proposals, from advisory committee members or other organizations, for coverage of additional activities as well, such as the distributive trades (ISIC Major Division 6). The census planners should

---

1/ Chapter VIII: Reference 1.
2/ Major Divisions 2, 3 and 4 of the revised ISIC.

consider the feasibility of such proposals very carefully, not only in relation to the funds available but to the ability of the staff to carry out a broader project which would present many additional problems. The planners may find that the success of the entire census would be endangered or as a minimum that the publication of results would be unduly delayed because of excessive time consumed in completing earlier processing stages.

55. On the other hand, there may be compelling reasons for enumerating wholesale and retail trade establishments in conjunction with the industrial census. In some eccnomies, the activities in many establishments are mixed, for example, manufacturing may be combined with trade. Where mixed activities prevail, the enumeration of all business establishments may entail very little more cost than the enumeration of those primarily engaged in manufacturing or mining, particularly if all establishments must be interviewed to identify those within the scope of the industrial census.

Establishments to be covered

56. It would be desirable at this stage to make some broad decisions on the coverage of establishments, because many aspects of the detailed planning to follow would be affected. Estimates of the relative importance of such segments of the economy as household industries should be obtained, and a decision made on whether reports will be obtained from all of these establishments or from a sample of a size commensurate with their importance. Care should be taken not to inadvertently under-represent segments accounting for a substantial part of an industry's total volume of production.

57. If all of the establishments in operation at any time during the census year are to be covered, as opposed to only those in operation at the time of enumeration or at the end of the year, plans for the industrial directory (discussed later in this chapter and in chapter V) would be affected. 3/

Organizational requirements

58. A decision on which government organization is to be responsible for conducting the industrial census should be made at an early stage. There are

---

3/ Estimates of the turnover in enterprise management (number of businesses discontinued and new businesses started) would be helpful in arriving at a decision on this question.

decided advantages in concentrating practically all authority and functions in connexion with the inquiry in a single agency of government. This should not, of course, exclude officials in other agencies from participation in the planning of the census through committees, or the review and inspection of the work of the statistical agency by other agencies with related responsibilities. The advantages of concentrating the census inquiry in a single agency are the speed with which changes in plans can be made effective, the central co-ordination of all detailed plans, and the economies in training full-time personnel rather than a large number of part-time personnel whose other tasks relate to the regular work of their agencies. As one phase of the work is completed, personnel can be transferred to a new phase without loss of time. In sum, concentration of all phases in one organization contributes to prompt publication of the results, because efforts can be more readily co-ordinated to fulfil the planned objectives.

59. Despite these advantages, it may be preferable in particular situations to delegate work to other government agencies. Local or provincial governments may sometimes be called on to provide field interviewers if the national organization is unable to recruit a sufficient number. Alternatively, the services of agencies which maintain contacts with particular industries or groups of establishments, such as an agency which regulates and supervises the activities of power plants, may be used. Such agencies often employ personnel with technical knowledge of the industry, which can be helpful in the collection, editing and general processing of the related questionnaires.

Establishing time schedules and budgets

60. One method of developing estimates of time requirements and setting up a budget for the census is to prepare a time schedule for each operation. Besides a description of the operation, the schedule should show the starting and finishing dates and estimates of the cost and personnel requirements during each week or other time period. Breaking down the operation into sub-operations contributes to more realistic estimates. A sample of such a time schedule is:

1973 INDUSTRIAL CENSUS

BASIC TIME SCHEDULE

Operation No.:_____ Date:_____

Description of operation: Collection of sample schedules for household industries

Item	January				February			
	Week 1	Week 2	Week 3	Week 4	Week 1	Week 2	Week 3	(etc., schedule will extend through entire time period)
Total operation:								
Personnel								
Cost								
Sub-operation 1:								
Province A								
Personnel								
Cost								
Province B								
Personnel								
Cost								
Sub-operation 2:								
Personnel								
Cost								
(etc.)								

61. The same type of schedule can be used for operations performed by other organizations, such as for the printing of questionnaires. For such uses, the figures for personnel would be omitted.

62. To the extent feasible, the prospective supervisors should participate in the development of time schedules for the operations they will be responsible for. After time schedules have been prepared for all operations that can be anticipated, a consolidated schedule should be drawn up. This would be similar to the basic time schedule above except that all of the operations, but not sub-operations, for the census would be listed in the stub. The totals of the consolidated schedule then would indicate, week by week, what the personnel requirements and costs are estimated to be for the entire census. The total cost would constitute the fund requirements for the census.

63. A review of the consolidated time schedule at this point could bring to light unforeseen problems such as unreasonable personnel requirements for some time periods.

Use of advisory committees

64. Many countries have found that the formation of a high-level committee of government officials, industrialists, and other public figures can be very helpful in utilizing the collective judgement and knowledge of the government and of the public in working out the objectives and the legal and organizational basis of an industrial census. Besides aiding in formulating the broad objectives of the census, these committees can also materially assist in launching the census by helping to secure needed legislation, funds and public confidence.

65. To be most effective, the committee should consist of high-ranking and experienced government officials who are users or potential users of data on the industrial sector, of trade association officials representing the industries to be surveyed, of accountants and other industrial executives, of labour leaders, of economists and statisticians, and of legislators. For the legal aspects of its work, the committee should have at least one lawyer. Government departments that might be represented include Ministries of Labour, Social Insurance and Industry; Tariff Commissions; and other agencies charged with regulating or promoting industry, or with economic planning and analysis.

66. After preliminary discussions and general agreement that a census is feasible and should be undertaken, the committee should consider the fundamental questions of legislation, industries to be covered and other matters discussed in the preceding sections of this chapter. The group should consider the needs of the govermment and the public for industrial census statistics, and weigh the needs against the problems of collecting the data and the resources available for such work. For this, industry and government personnel not on the committee could be called on to advise and assist the group. As a result of its deliberations, plans for the census should be improved and interest in it should be stimulated among government agencies, industry officials, and the general public.

67. As plans for the census progress, it may be found desirable to form other committees to advise on specific aspects of the census, such as industry and commodity classifications, reportability of information and editing, coding and tabulation problems.

Publicity

68. Recognition of the merits of the industrial census, particularly among respondents, leads to better co-operation and the filing of more accurate reports. This in turn leads to more efficient processing and possibly earlier publication of results. Publicity, therefore, can contribute significantly to the success of the census.

69. A complete publicity programme usually consists of three distinct phases:

1. Advance publicity to inform the general public that a census has been proposed and to seek advice on the content of the census in its formative state. After questionnaires have been tentatively drafted, comments and suggestions may be solicited from industrial associations, industry periodicals, and directly from a cross-section of the businesses which will later supply the data.

2. Data collection publicity, undertaken shortly before the final questionnaires are distributed, to stimulate goodwill among respondents. Articles can be written about the benefits of census data and the importance of timely and accurate reporting. Descriptions of the statistical organization, its equipment and processing methods may also be newsworthy.

3. Census results coverage, to inform the public about the availability of reports and to stimulate the use of the data.

Classification systems

70. The planning of classification systems for industries, products and materials, size of establishments, and other characteristics of the census data which can be coded need not be given the same priority as the fundamental decisions described above, but certain classification systems such as those for products and materials, if they affect the questions to be included in the questionnaire, should be developed at an early stage. While the design of classification systems is a technical matter and will not attract wide government or public interest, there may be other government agencies which use such systems for related purposes and whose interests should be fully considered. The comparability of the classification of imports and exports, for example, to that to be used in the census for products and materials, should be carefully considered. The classifications by geographic areas which may be in use in other government agencies should also be reviewed for comparability.

71. There are international recommendations for classifying industries (ISIC) [4] and imports and exports (SITC) [5] and wide use has been made both nationally and internationally of these recommendations. The systems used by many countries which have recently developed their own industrial classifications have been based on the ISIC, and other countries have modified their existing systems to make them comparable. Classification systems for products and materials have in many cases been developed from the SITC or the ISIC. [6]

72. To be both useful and practical, the classification schemes should result from considering the requirements for classified data, the information that can be collected for classification purposes, and the coding problems that would arise under different systems. While the classification systems should make the distinctions that are important to most users of the basic figures, these distinctions should not be so precise that the information needed for coding is very difficult or impossible to collect. At the same time, the

---

4/ Chapter VIII: Reference 3.
5/ Chapter VIII: Reference 5.
6/ Specific recommendations on these classifications for the 1973 inquiries will be found in ST/STAT/44 and ST/STAT/45.

classification schemes should encompass all of the industrial activities carried on in the country, and since they may be used for many years, they should anticipate future changes in the economy. In the interest of efficiency, they should make only the distinctions which are or will be economically significant and that exist in the industry structure. For example, the industrial classification should not specify different categories for activities commonly carried on in the same establishment. Conversely, two activities of significance that are ordinarily performed in different establishments should not be combined in one category. If it does the former, coding may be difficult or impossible and it may be necessary in the end to combine the two groups for publication. If the latter, statistics for establishments in essentially different groups will be combined in one group, and whatever economic distinctions between the groups are significant will not be ascertainable from the census tabulations.

73. The classifications must provide an unambiguous class for each unit. For example, in a commodity coding structure, the definition of each category must clearly distinguish between the commodities included there and those included in other categories. Similar considerations apply to all coding structures; if this principle is not followed, the coding operation may be much more difficult and time consuming, because questionable cases will frequently arise which cannot be resolved by coding clerks.

74. Considerable attention should be paid to the comparability of the classification systems used in the industrial census to those employed for other economic data. If the country has a minimal amount of economic data at present, thought should be given to the future development of series and to their potential uses such as in a system of national accounts. Comparability to international systems is also important if national economic planning and analysis is to benefit from the experience of other countries.

Industrial directories

75. The experience of countries with a long history of industrial censuses is that a directory of establishments is an indispensable preparatory step for enumeration, whatever the method selected. In its most elementary form, a directory consists of a set of cards with the name, address, industrial activity and size code of each establishment.

76. A directory is essential if a mail canvass is planned or a sample is to be selected. Even when a complete field enumeration is planned, the cost of compiling a directory may well be more than offset by savings in subsequent operations, such as the verification and control of enumeration. If available early enough, data compiled from the directory will also assist greatly in the development of efficient and effective plans for the census.

Methods of enumeration

77. Once the industries and establishments to be covered have been determined, an important part of planning is to decide on the method of enumeration. There are two basic methods employed in censuses: self-enumeration, in which questionnaires are distributed and received from respondents by mail; and enumeration by interview or personal visit. In practice, various combinations of these methods are normally employed, although personal visits and enumeration by interview usually predominate. Some examples of enumeration methods are:

1. Enumerators visit the industrial establishments and fill out the questionnaire at once, based on answers given by the establishment owners or managers. This method is suitable only if the establishments are small and the questionnaire very simple so that answers to the questions are not difficult to obtain.

2. Enumerators visit the establishments, explain the purposes and importance of the census, and leave the questionnaire for the respondent to complete. At a later date reached by agreement, the enumerators return and pick up the completed questionnaires. They may also review them for completeness at that time.

3. Respondents are told to report to a local statistical office where they explain the nature of their business and are given the appropriate questionnaire to take back with them. After completion, the questionnaire is returned to the local office.

4. Questionnaires are mailed to the respondents and field representatives call for the completed forms at a later date.

5. Questionnaires are mailed to the respondents, who complete and return them to the statistical agency at a later date. This approach is the most common one for countries with developed statistical programmes.

Use of samples

78. Sampling, that is, collecting and processing reports from only part of the universe, may sometimes be advantageously employed. The principal advantages of sampling lie in the potential reduction in costs and manpower

requirements. Since there will be fewer questionnaires to be collected, edited and tabulated, earlier publication of results may also be possible. On the debit side, samples are more difficult to administer and control 7/ than are complete coverage surveys. Basically, this is because the concepts behind complete coverage are relatively simple. Employees without much background or training in statistics can understand these concepts readily and apply them in practice. The difficulties of administering a sample increase in proportion to the complexity of the sampling plan: thus, if there is a choice between two plans, one more efficient but also more complex than the other, selection of the more efficient plan may often be the wrong choice from the administrative point of view.

79. Another drawback of sampling is that the amount of reliable detail which can be published will inevitably be reduced. One way of looking at this problem is in relation to the average number of sample cases per publication cell. If this average is small, as for example where publication cells represent industries within small geographic sub-divisions, sampling may not be feasible. Where cost considerations are of overriding importance, on the other hand, it may be necessary to sacrifice the geographic detail and confine the collection effort to a sample, because sufficient funds for complete enumeration cannot be obtained.

80. Sampling is not a magic formula for success. Collection, editing and tabulation problems which arise with complete coverage will most likely occur also when a sample survey is conducted. Sampling will not eliminate the problems, but they will be fewer in number because there will be fewer respondents and publications cells.

81. Two methods of sampling have been used extensively for industrial inquiries. These are (1) probability sampling where the entire universe is represented by giving every establishment in the universe a known probability (greater than zero) of being selected, and (2) cut-off sampling, where coverage is restricted to all establishments having a given characteristic, most commonly

---

7/ One disadvantage of probability sampling is that it requires careful control over the selection of the sample units, and the application of sampling weights. Otherwise, errors can be introduced which would not occur in a complete canvass.

those above a specified size. No satisfactory sample surveys can be conducted without some kind of "frame" or universe listing. In the discussion which follows, it is assumed that a universe of industrial establishments has been identified by means of a directory or list which indicates the size of each establishment. 8/ If the available funds are not sufficient to finance both a directory and a sample survey, it may be better to establish the directory and postpone the sample survey to a later date. Some useful structural data can often be compiled and published based only on the information contained in the directory, such as size, location and industrial classification of the statistical units.

Stratified probability samples

82. Ideally, stratification should create homogeneous groups of establishments, each group differing in its major characteristics from all other groups. It is not possible to realize this objective completely in practice, and the usual procedure in industrial censuses is to subdivide the universe into two or more groups defined by size of establishment. For example, a frequently used method has been to distinguish between establishments with less than five persons engaged and those with five or more. To achieve greater homogeneity, the size groups are sometimes further subdivided by industrial classification, but a more common practice is to adopt the same sample rate for all establishments within each size group. This has the great advantage of simplicity. In the example cited, the sample might consist of all cases in the top size class (complete coverage) and a 10 per cent sample of the smaller establishments. 9/

83. The 10 per cent sample might be selected by arranging the establishment cards in some systematic way, such as alphabetically, numbering them serially, and then beginning with a random number from 1 to 9, selecting every 10th card for the sample. This method is called random systematic sampling.

8/ See chapter V for a discussion of the establishment of an industrial directory.

9/ Whatever the stratification plan, it is absolutely essential for a useful probability sample in the industrial field to include all of the very large establishments.

Selection of a two-group stratified sample of the type described might be appropriate where (1) there is a large number of small establishments and (2) they account collectively for a substantial proportion of the universe in terms of output or persons engaged. If it is known that the small establishments account for only a small percentage of output or employment, say less than 5 per cent, a cut-off sample may be more appropriate. It should be recognized, however, that the reliability of sample estimates in terms of standard errors cannot be computed from the results of cut-off samples.

84. Stratified samples are most appropriate when enumeration is conducted by mail. If field enumeration is employed, area or cluster sampling is nearly always more efficient, the principal reason being the resultant saving in cost of travel.

Area or cluster probability samples

85. Samples of this type are appropriate in industrial censuses where there are large numbers of small establishments to be covered by field enumeration. Area samples can yield adequate national totals, but publication of statistics by areas will usually not be possible except perhaps for the areas included in the sample. Area sampling should be employed only to cover small or medium-sized establishments. The largest establishments, including those located in non-sample areas, should be enumerated completely.

86. The theoretical criteria for effective cluster sampling are just the reverse of those for stratified sampling. In cluster sampling, the groups or clusters should be an non-homogeneous internally as possible, and each group should resemble all other groups in its major characteristics. In most practical situations, however, the clusters are geographic areas or political subdivisions and their internal composition is accepted as it exists. A simple cluster plan could start with a list of townships, counties or other political divisions. These are divided into two groups, small and large, according to their importance industrially. By random methods a sample of the "small" areas is selected, while the "large" areas are automatically

included in the enumeration.[10/]

87. In the simplest form of such a plan, all of the establishments within each of the sample areas are enumerated. Second-stage sampling, or sampling within clusters, is also feasible and can be employed if some of the sample areas contain unusually large numbers of small establishments, or of smaller defined areas such as villages. The secondary sampling units could be either villages or establishments, selected by random methods.[10/]

Cut-off samples

88. By this method, coverage is restricted to establishments above a minimum size, or "cut-off". These establishments constitute the universe, which is usually enumerated completely. Since cut-off sampling is not a probability-type method, the reliability of the published aggregates or estimates cannot be measured precisely in terms of standard errors. Even though the over-all percentage accounted for by small establishments may be known accurately, the percentage applicable to individual cells is usually unknown and can be expected to vary widely.

89. Cut-off samples should be used only when a small proportion of total output (usually less than 5 per cent) is accounted for by the small establishments excluded from the sample. In recent censuses of manufacturing in the United States, for example, all establishments with no paid employees have been excluded. A special study made in 1958 revealed that although there were 52,000 such establishments, about 15 per cent of the total number, they accounted for less than one-fourth of one per cent of total output. The proportion was nearly two per cent for lumber and wood products, however, and over one per cent for printing and publishing.

90. Where the importance of the small establishments below the cut-off level is known only approximately, it is reassuring to know that a large error in estimating their output may have only a small effect on the total. For example:

	Actual output	Estimated output	Per cent error
Large establishments	980,000	980,000	-
Small establishments	20,000	30,000	+50
Total	1,000,000	1,010,000	+1

10/ See the discussion of random systematic sampling under "stratified probability samples", paras. 82-84.

In this sample, even though the output of the small establishments was overestimated by 50 per cent, the effect on the total was only one per cent. This percentage is probably within the tolerance limits expected even in complete enumeration.

Treatment of small establishments

91. As pointed out in chapter III, many difficulties will be encountered in the collection of industrial census data from small establishments. They are generally hard to locate and identify as industrial establishments. They are unlikely to keep records or have systematic information regarding their activities.[11/] Income and expenditure are seldom segregated by type. The owners may find it difficult to comprehend such concepts as expenditures for fixed assets as distinguished from purchases of supplies, or the distinction between manufacturing and trading activities. They may not be able to supply aggregate figures covering a whole year, particularly value figures. In most countries, especially countries with large numbers of small mining or manufacturing enterprises, or the widespread manufacturing of goods in households, these are serious problems and the success or failure of the census may depend on their solution. While adequate publicity and careful training of enumerators will help to reduce the number of problem cases and hence the over-all magnitude of the problem, there are other steps which can be considered:

Sampling As discussed above, the census of small establishments may be based on a sample rather than on complete coverage. A satisfactory sample can hardly be selected without a directory, and this means that some information will have to be obtained from all establishments. But the few items of information required for a directory are much easier to obtain than detailed answers to typical census inquiries.

The manpower required to conduct the census may be considerably reduced if coverage of small respondents is confined to a sample of moderate size, such as 10 per cent or 20 per cent. In some countries, funds may be available but the lack of qualified manpower may rule out the collection of data from the entire universe.

Short forms Short questionnaires for small establishments have been used very successfully in many industrial censuses. Such forms might include only the items basic to the census such as

11/ This is particularly true of industrial production in households.

location of the establishment, a description of its industrial activity, number of persons engaged, wages and salaries, cost of materials and purchased industrial services, amount expended for machinery and other fixed assets, and value of products and services sold. Although it should be possible to obtain adequate answers to these questions from most small establishments, it may be necessary to modify or omit some of them if advance pilot studies or discussions with small respondents point to serious reporting problems.

The use of the short form, however, raises the problem of what should be done with statistics appearing only on the regular questionnaire. There are various possibilities. If the aggregate contribution of small establishments to the statistics in question is not large, their share may be imputed prior to publication. In computerized tabulation, it has been found possible to allocate such imputations to individual establishment records, thus making them readily available for geographic and other cross-classifications.

The contribution of small establishments to some aggregates such as that dealing with electricity sold to others, can probably be imputed as zero. In other cases, it can be imputed at a fixed ratio to total output, or at different ratios depending on the industrial activity of the establishment.

If the contribution of small establishments to particular items omitted from the short-form questionnaires is large, a sample of small establishments can be selected for long-form coverage. The collection of long-form data from the small respondents included in the sample might then be entrusted to specially trained enumerators or to field supervisors. Further refinement of the results might be obtained by imputing values for these items for all other small establishments using ratios based on the sample results.

Another type of imputation arises when the short form calls for the same information as the long form, but in less detail. As an example, the short form might request the total amount expended for new fixed assets, while the long form might require separate figures on expenditure on (1) building, improvements to land and other construction work, (2) transport equipment and (3) machinery and other equipment. While the distribution between the three items could be estimated for small establishments, based on the proportions reported by large establishments, a better procedure might be to publish the short-form data separately. The captions for this purpose might read:

1. Total expenditure on new fixed assets
   a. Building, improvements to land and other construction work
   b. Transport equipment
   c. Machinery and other equipment

2. Expenditures on new fixed assets, kind not reported

Item 2 would consist largely of short-form data, but might also include data from large establishments which were able to report the total but not the detail.

Differential processing Coverage of large numbers of small establishments creates a heavy workload in editing, coding, and tabulation. The cost of these operations should be considered relative to the importance of the small establishment aggregates, and also in relation to the effect of the errors likely to be discovered and corrected. There is a certain amount of safety in numbers: if errors are of a random type, they will tend to cancel each other out provided they are individually small. This leads to the conclusion that a high degree of accuracy and rigid standards are unnecessary refinements when dealing with small establishments. The amount of attention paid to individual reports during editing and coding should vary in direct proportion to the size of the establishments represented. The funds and human resources available for these operations will be more efficiently used if this general rule can be observed.

In stratified sample surveys, it should be noted, the importance of individual small establishments reports is increased greatly and the foregoing suggestions are not appropriate.

Tabulation plans

92. Before the census questions are made final and the questionnaires are printed, the content of the statistical tables should be outlined, and some consideration given to coding and tabulation procedures. Some countries with limited experience have prepared tabulation plans when the collection of data has neared completion; at this stage they have discovered that certain questions or topics should never have been included in the census, simply because the information obtained could not be coded and tabulated.

93. Members of advisory committees may suggest the inclusion of items of data which may have merit in that they would develop information useful for policy guidance or for other purposes, but which are essentially inappropriate for inclusion in an industrial census. An example is the type of question which calls for the opinion of respondents rather than facts or figures. While often very useful in their own right, opinion surveys have special problems and should not ordinarily be conducted in conjunction with industrial censuses.

94. A detailed consideration of tabulation plans may also reveal weaknesses

occasionally found in census questions, such as categories which are not mutually exclusive, making it impossible to interpret the replies or inaccurately phrased questions, leading to ambiguous or meaningless results.

95. Discovery of such defects while the census is being planned will help to avoid much wasted effort in the collection, editing and coding phases. The process of "thinking things through", that is, visualizing a proposed inquiry in terms of the presentation of its final results to data-users and setting down the actual format of the presentation in outline form will often aid in bringing conceptual and other errors to light.

Staff requirements

96. Because of the broad economic significance of the basic objectives of the industrial census, the director of the organization responsible for the project should be a person of the highest qualifications with administrative ability and training in both economics and statistics. Staff requirements for the rest of the organization will vary from country to country, but the organization might initially be comprised of the heads of the following prospective divisions, who together with the director would form the advance planning staff:

A planning division, whose functions would encompass the determination of the scope and content of the census, the development of industry and commodity classification procedures, the design of sampling techniques and editing and tabulating specifications.

A field division, whose functions would include the compilation of an industrial directory, enumeration by field interview or by mail, coverage control, and any field pre-test of the questionnaires.

A processing division, to edit and code the data on the completed questionnaires, to assemble information on the progress of the project and prepare progress reports, and to prepare tables for publication.

A tabulation division, to operate computing equipment and in the case of electronic equipment. to write computer programmes.

97. The planning staff should participate in preparing agendas and reports for committees, and in acting on their recommendations; plan required research and special studies; prepare or supervise the preparation of time schedules for the planned operations; and make any inter-agency arrangements

necessary to the plans. At least initially, the planning staff should meet frequently to review progress and receive assignments. A time schedule for planning should be set up so that the advance planning will proceed smoothly and on schedule.

98. No attempt will be made to lay down specific qualifications for each prospective division head, or member of the initial planning staff. The following types of training, experience or skills will normally be found useful:

> Educational background in statistics, economics, accounting, business administration, and related subjects. Training in the operation of industrial censuses in a country offering such assistance.
>
> Managerial or accounting experience in mining, manufacturing, electricity, gas or water supply. Experience as a member of the secretariat of a trade association representing these industries.
>
> Experience as a statistician, planner or supervisor in other statistical projects, not necessarily in the field of economics or industry.
>
> Background in setting up classifications systems or in applying them to coding operations.
>
> Experience in supervising field collection of data of any kind; or in other field work such as door-to-door selling.
>
> Experience in budget planning, in purchasing of equipment and supplies, and related management functions.
>
> Experience in managing of general office and clerical functions.
>
> Experience in supervising the operation of accounting equipment, punch-card equipment, and electronic computers. (The latter is essential if such computers are to be employed in processing the census results.)

99. The persons assigned to planning specific phases of the census should later have the responsibility for carrying them out. This helps to ensure continuity of planning and completion of the planned operations. It permits the training and experience acquired in the planning stages to be utilized to the fullest extent in the later operations of the census.

Equipment requirements

100. There are often considerable delays in procuring office machines, tabulation equipment, and specialized supplies such as punch cards. It is desirable, therefore, to begin to arrange to purchase, rent or borrow

equipment as soon as the kinds and numbers of machines required can be approximately determined.

101. As described earlier in the chapter, time schedules for the operations involving equipment will provide information on when and for how long they will be needed. The number required is more complicated because for each operation it will be necessary to estimate the total number of physical work units such as questionnaires to be processed through the equipment, and the normal capacity of the equipment in terms of number of work units per hour or day. The latter figure should be obtainable from the equipment manufacturer or can be estimated with his assistance. Allowance should be made for time lost due to breakdown of the equipment or delays such as those due to failure of the preceding operation to deliver a sufficient number of work units ready for processing. The length of time the equipment will be needed may not be a critical factor if it is purchased, or rented on a week-to-week basis with the right to return it at any time.

102. The following hypothetical schedule illustrates the calculation of requirements for punching equipment:

Date:_____

1973 INDUSTRIAL CENSUS

Requirements for #X-11 Numerical Punch Machines

(Punch-card rate per week, 5000)

	Total, all weeks	May				June	
		Week 1	Week 2	Week 3	Week 4	Week 1	(Etc. - schedule will extend through entire processing period)
Operation No. 243:							
No. of questionnaires							
No. of cards							
No. of machines							
Rental cost							
Operation No. 244:							
(etc.)							
Totals							
Questionnaires							
Cards							
Machines							
Rental cost							

If the equipment is purchased rather than rented, the cost could be the purchase price less salvage value divided proportionally over the weeks of operation. This cost should be part of the consolidated personnel and cost schedule described at the beginning of the chapter.

V. COLLECTION OF DATA

General remarks

103. In chapter IV, some alternatives that should be considered in laying plans for data collection were outlined. These included the possible preparation of an industrial directory, enumeration by mail versus field interview, and sampling versus complete coverage. Chapter V provides more detailed recommendations for the collection phase, with emphasis on the methods generally considered most appropriate for countries at an early stage of statistical development. It is assumed that the census will cover the industries included in Major Divisions 2, 3 and 4 of the revised ISIC, that is, mining and quarrying, manufacturing, and electricity, gas and water supply.

104. The data collection procedures described here would be essentially similar whether complete enumeration or sample coverage were planned. When sampling is used, however, an effort should be made to increase the accuracy and reliability of the basic data by training enumerators more thoroughly, and exercising more care in the subsequent processing phases to detect and correct response errors. Careful attention to these aspects can offset the errors of estimate introduced by the sampling method, and in some circumstances make the results of a sample survey more reliable than the data obtained by complete coverage.

An industrial directory

105. Countries with much experience in conducting industrial censuses have found that a directory of establishments is an indispensable preliminary step, whatever the method of enumeration planned for the census. A directory is basically a set of cards or a listing with the following items of information as a minimum:

1. Name and physical location of the establishment.
2. Mailing address, which may differ from the physical location. This is the address to which mail inquiries would be directed.
3. For establishments which are part of multi-unit enterprises, name and address of the central office or headquarters. This may be the same as the mailing address.
4. Industrial activity description and code.

5. Size code, which should normally be based on the number of persons engaged.

6. Source and date of this information.

106. Owing to the difficulties sometimes encountered in obtaining these items of information for large numbers of small establishments, some countries have compiled directories which cover only establishments above a specific size cut-off, such as those with five or more persons engaged. The completeness of such a directory is usually open to question, since in the absence of a complete enumeration, there will always be uncertainty as to whether all the establishments above the cut-off size are included. A decision to exclude small establishments from the directory should be reached only after consideration of their economic importance, their probable numbers, the cost of collecting directory-type information from them, and the resources available. If it is planned to maintain the industrial directory after the census for future use, the turnover among small enterprises should also be considered. If there is likely to be a high turnover rate, which is the usual case, the planning staff may conclude that the cost of maintaining a directory of small establishments is prohibitive.

Sources of information for the directory

107. The usefulness of a directory depends in great measure on the kinds of data it contains, its completeness and its accuracy. In considering possible sources for the directory, their quality is important because the directory can be no better than its source material. The following are some of the sources which have been used for industrial directories:

1. A preliminary enumeration conducted for the express purpose of establishing a complete directory, or of supplementing directory listings obtained from other sources. An abbreviated questionnaire should be used for recording the data.

2. A census of population conducted some time before the industrial census. In such a case, the population census questionnaire would include a brief inquiry about the industrial activities carried on in the household, and this would form the basis for a directory of household industries. In this connexion, it will be important to know whether the list of households with industrial activity can be made available in time for use in the industrial census. If there will be a considerable delay (due perhaps to the fact that publication of the population census

results will have greater priority than the preparation of this special list), the list may be so out of date as to be of little value. Population census enumerators might also be instructed to fill out directory cards for regular industrial establishments (not households) encountered in their districts. This procedure has usually not resulted in a satisfactory list, however, and it may unduly complicate and slow down the enumeration of the population.

3. Government records maintained for taxation purposes, in connexion with social benefit programmes, licensing, regulation, or other government functions.

4. Industrial directories maintained by trade associations or trade publications. Such sources tend to be incomplete.

5. Telephone directories, residential directories and other local lists.

108. Except for a complete enumeration planned and conducted by the industrial census staff, the use of these sources requires careful examination of the listings for completeness, suitability for census purposes, and accuracy. Administrative records or registers are hardly ever assembled with industrial censuses in mind, and while they may be adequate for their original purposes, they usually will need correction or supplementation before they can be used for the census. The records may have no indication of the size of the establishments and may thus be inadequate for the selection of samples. They may relate to enterprises rather than establishments. The locations given may be those of central offices or sales offices rather than of industrial establishments. They may be out of date, in the sense that they include obsolete listings and exclude enterprises formed recently. They may be incomplete simply because insufficient effort was expended in compiling them. They may contain an unacceptable number of erroneous addresses or descriptions of industrial activity. They may include establishments which are outside the scope of the census but are not so identified on the records. They may contain duplication: for example, the same establishment may be listed two or more times, perhaps under different names.

109. Information on the extent and nature of the deficiencies in a given set of records may often be obtained from the organization which compiles and maintains it. The comparison of one register with another which is presumed to cover the same area or industry should also reveal defects. And finally, it

may be feasible for the industrial census staff to conduct a complete enumeration of a small area or two and use the results as a standard for gauging the adequacy of the various sources.

110. The importance of obtaining the co-operation of trade publications and of industrial organizations such as manufacturers' associations is discussed in the previous chapter. These groups can also sometimes provide expert advice and assistance in the compilation of directories for their individual industries.

Maintaining the directory

111. Countries which plan to follow the industrial census with annual and more frequent inquiries should consider the feasibility of keeping the directory up to date in subsequent years. This task will be simplified if maintenance of the file is planned from the beginning. A provision for registration of new firms and reports on changes in the industrial activity of existing firms may, for example, be included in the legislation authorizing the census. Other possibilities include enlisting the aid of a regulatory agency, which may be able to provide the information needed to maintain the directory by minor changes in its procedures. A system used by some countries consists of transcribing the information in the directory to special forms which are mailed to the enterprises. These listings are corrected and up-dated by the enterprises and returned by mail to the statistical agency. This procedure may be adequate to keep the directory current with respect to changes in the activity of existing establishments and with respect to the acquisition of new establishments by multi-establishment enterprises. It would be necessary, however, to find other sources for establishments of newly formed enterprises, or for the establishments of enterprises which change from an out of scope activity such as wholesaling to an in-scope activity such as manufacturing.

Cards or computer tapes

112. Unless the number of establishments is small, the directory listings should be recorded on punch cards or computer tapes. These methods make it possible to re-sort the listings quickly for special purposes: for example, it may be desired for some special purpose to obtain the names of all establishments engaged in a specific activity. A master list should be printed out

so that the directory can be reconstructed if any of the cards or tapes are lost or destroyed.

Uses of a directory

113. Once the directory is compiled, there are many ways in which it can be made useful, depending on the enumeration plan. If a sample of small industrial establishments is to be selected, the directory cards for the small establishments would be segregated, and every nth card selected beginning with a random start. If the directory is on tape, a computer programme can be written which will select every nth small-establishment record without prior sorting.

114. If the plan calls for a direct mail canvass of certain types of establishments, such as those operated by multi-establishment enterprises, and field enumeration of the rest, the appropriate records can be printed out to form a mailing list for the multi-unit group. At least four copies of this list should be made: (1) to be transferred to the out-going mailing pieces by labels or other means; (2) to be given to the field supervisors, who will instruct enumerators not to call on the listed establishments, except possibly for follow-up calls at a later date; (3) to be transferred to a set of check-in cards, which will form the central control record of the receipt of completed questionnaires from the respondents; and (4) a master copy to be retained separately in case portions of the other copies are lost or destroyed.

115. The balance of the directory, or the entire directory if no mail canvass is planned, should be sorted by area and printed out. The supervisor of each area will be given a list of all the establishments in his area to be enumerated. Copies of this list should also be retained in the central office for check-in purposes and for the preparation of reports on the progress of the enumeration.

116. The completeness and accuracy of the directory should not be taken for granted in the field; the enumerators should be instructed to take note of non-listed establishments and to secure reports from them. Provision should be made for creating new directory listings for such added establishments. The field supervisors may also be instructed to compare the directory listings for their areas with information available from local taxation or licensing

authorities, classified directories, and other local sources. This procedure may reveal additional establishments or indicate errors in the existing directory listings.

117. The census plans may call for separate questionnaires for different industrial activities, or the same basic questionnaires with supplemental pages to cover specific activities. Such supplemental pages might consist of a printed list of products and services, with code numbers and space for the responses. The directory can be used to provide data on the number of such special questionnaires needed for each area, and in a mail canvass can provide information for the selection of the proper questionnaire to mail to each respondent. Counts developed by tabulation of the directory can form the basis for an economical and efficient printing order for questionnaires, since the order can be restricted to the number of forms actually needed for the establishments plus an allowance for incidental requirements.

Questionnaires and other data-collection forms

118. The data-collection forms are the key documents in the industrial census since they provide the means for recording all the establishment data which are to be tabulated and published. Their format, organization and content will have a significant influence on the quality of the statistical results, the cost of collection, editing and tabulation, and the promptness of publication. In all successful industrial censuses, therefore, considerable time has been devoted to the design of questionnaires and other data-collection forms. Ambiguous phrases, printing errors, and even poor spacing, have caused some countries to abandon important inquiries or to expend scarce resources in an effort to correct the effects of such mistakes. In self-enumeration, when the respondent fills out the questionnaire without the assistance of an interviewer, even a poor selection of type-faces has been known to cause widespread misinterpretation. This situation can arise because a key heading, important for understanding the captions below it, does not stand out boldly and is consequently overlooked by the respondent.

119. Pre-testing of the data-collection forms can often bring such deficiencies to light in time to correct them. Such pre-tests need not be elaborately planned or scientifically designed. The early enumeration of one or two

districts, for example, may demonstrate the adequacy or inadequacy of collection forms. If alternative types of questionnaires are proposed to cover the same group of respondents, pre-tests can help to indicate which is the preferable type to use.

120. The physical form of the questionnaire, for example, whether it is to be printed on card-stock or paper, or whether it should be assembled as a continuous sheet or booklet, should be determined after consideration of all the uses to which it will be put. If it is expected that many of the questionnaires will be filled out on typewriters, the continuous sheet or accordion-fold type is preferable to the booklet. The spacing between the lines on such a questionnaire should also correspond to typewriter spacing. If the information is to be recorded by an enumerator, as in field canvassing for an industrial directory, a sheet or large card on which data for many establishments can be recorded may be the most efficient form. The durability of the paper or card-stock used should be considered in relation to the probable amount of handling the questionnaire will receive.

121. In addition to the spaces for respondent entries, it is usual to include spaces on questionnaires for office entries, labelled "for office use only". Such entries are of three general types:

1. Control data such as establishment number, identification of the enumerator and others processing the questionnaire, and dates on which various actions take place.

2. Code numbers such as those for location, industrial activity, size of establishment and type of ownership.

3. Calculated figures such as average number of persons engaged, average earnings, and possibly value added. These figures may be intended for tabulation or merely for use in a pre-punch review of the questionnaires for reasonableness.

With access to a computer, some or all of the coding and calculation may be performed by the computer after the data have been punched and put on tape, thus eliminating the need for spaces on the questionnaires.

122. Each data item on the questionnaire should be numbered, to facilitate reference to the item in the instructions and correspondence. The data items or code numbers which are to be transferred to punch cards should also be clearly designated. If possible, the designation should include the card number and the card field number. This will facilitate card-punching and reduce errors.

Instructions accompanying questionnaires

123. The instructions for respondents should cover, concisely but completely, everything necessary to explain what is required and to avoid misunderstanding. Instructions which are too long and diffuse are less likely to be read and understood than those which are brief and to the point. If long instructions are required, however, a common practice is to include only the most essential points on the questionnaire itself and the rest in a supplementary instruction sheet. Pre-census interviews with potential respondents often provide valuable indications of the points which should be emphasized in the instructions.

124. The following is a check-list of points normally covered in instructions on questionnaires or on supplementary instruction sheets:

1. Legal authority for the census.
2. Confidentiality of returns.
3. Types of establishments required to report.
4. Acceptability of estimates if accounting records do not supply required data.
5. Date the completed report is due.
6. How to return the completed report.
7. Time period covered by inquiries spanning a full year; acceptability of periods other than the calendar year, or the year specified.
8. Definitions of the various items of information requested; rules for valuation in case of items of data reported in value terms.
9. How to report individual products or materials (on questionnaires where descriptions are not pre-printed).
10. Special instructions for particular industries where problems affecting many respondents are anticipated. For example, how to report fish-processing operations when the same establishment is engaged in catching fish.

Phrasing of questions

125. Although the instructions are important and should be carefully written, even greater care should be accorded to the wording of the questions. A major factor in securing prompt, complete and reliable replies is the phrasing of questions so that they are readily understood by respondents and field interviewers. Some respondents will fail to read separate instructions, but all,

presumably, will read the questions before entering replies. Every effort should, therefore, be exerted to keep the wording of the questions clear and concise, and to avoid ambiguities. If two interpretations of a question are equally likely, and only one is correct, half the replies will be unacceptable.

126. For some items, check-boxes are preferable to direct questions. In a check-box query, all the possible replies are printed out, and the respondent is asked to check the applicable replies. Such devices are sometimes appropriate for information about the various industrial operations performed in the establishment, or about the methods of distribution of products. They may also be employed for obtaining quantitative data if replies in terms of size classes are acceptable.

127. A question calling for verifiable facts is preferable to one which relies on the judgement of respondents or field interviewers. Rather than ask the respondent to describe his primary industrial activity, the questionnaire can request detailed data on products made and services rendered which can be used at a later stage to assign an industrial-activity code. In practice, some countries have asked for both a narrative description and data on products and services, on the assumption that if one inquiry is not answered satisfactorily, the desired information can be obtained from the other.

128. "Tailoring" the language of certain questions may be desirable or necessary to communicate with particular industries. In censuses of manufacturing, this practice is especially useful for describing such concepts as the total value of products and services or the total cost of materials and supplies. These concepts, if described briefly as in the previous sentence, may be incomprehensible to many respondents. The naming of products or materials peculiar to an industry, or a request for separate data for them, can go a long way towards ensuring that the desired answers will be forthcoming.

Training of enumerators

129. An industrial census is considerably more complex than a population census. Many individuals in the population have essentially similar characteristics, their classification is relatively simple, and the training of

enumerators in statistical concepts is less important than teaching them procedural matters such as the need to visit all households, what to do when families are absent and the value of speedy enumeration. In contrast, the statistical units of an industrial census vary greatly in such characteristics as size, organization, and industrial activity. Moreover, one of the first problems encountered by the field interviewer, that of determining which of the establishments he encounters are eligible for enumeration, hardly exists in a census of population.

130. The experience of countries in the 1963 World Programme illustrates the need for thorough training of field interviewers. In one country, 80 per cent of the initial returns were found to be so defective that the establishments had to be revisited; about a quarter of these questionnaires were regarded as having been completely redone when the enumeration was finally completed. The training of enumerators, and adequate field supervision, undoubtedly would have avoided much of this costly work. If there are restrictions on the funds available, it would be better to reduce the scope of the census (and hence the number of interviewers required) than the quality of enumeration.

131. Ideally, it would be desirable to concentrate the task of training in the hands of a few individuals in the central statistical agency. This has rarely been feasible in practice, however, and the usual first step is to train individuals who will teach others. These may be regional-office heads and other key personnel who are given an intensive background course in the central agency. Later, these persons return to their areas to train local-office supervisors and these in turn train field interviewers.

132. One country has reported an interesting alternative in which a team of enumerators, half of whom were permanent staff members of the central statistical agency, moved around the country enumerating area after area until the work was complete. Unlike a population census, the enumeration of industrial establishments can be spread out over a period of several weeks.

133. At all levels, training classes should be small so that the teacher can give attention to the progress of each individual in the class. If feasible, simulated interviews, in which an experienced person takes the part of the establishment owner or manager, should form part of the training course.

Written tests should be given at some time during the course. Some part of this may take the form of a description of a hypothetical establishment, based on which the pupil fills out a test questionnaire. Each individual approved for field work should demonstrate that he will be able to cope with at least the most frequent of the complex technical problems which arise during enumeration.

Verbatim guides

134. Some countries have been faced with training large numbers of field interviewers in a short period and have found it worthwhile to prepare detailed classroom guides. These are intended for the use of relatively inexperienced teachers, who are expected to instruct the class by following the guide practically word for word.

Home study materials

135. There are also examples of home-study reference materials which are given to prospective enumerators in advance of any classroom instruction. After several days allowed for home study, the enumerator is given a written test. If he passes the test satisfactorily, he is interviewed by the supervisor. During the interview his understanding of the principal questions is tested again, the questions on which he failed the written test are reviewed, and a supervisor takes the part of the respondent for a practice interview. Following this the supervisor accompanies him on at least one real interview. The enumerator's progress is again reviewed closely after a probationary period of several days.

Pocket manuals

136. A condensed manual of instructions for enumerators, intended to be carried on the person and consulted during the work-day, has been used successfully in many countries. Such a manual may also serve as a basic textbook for the training course. All of the procedural matters affecting enumerators, and the principal problems they may expect to encounter, should be discussed in such a manual. The subjects covered in a typical manual are outlined below:

1. Rules and regulations. Legal obligations of enumerators to provide true returns and treat information given by respondents as confidential. Actions forbidden to enumerators, such as soliciting or selling.

2. Excerpts from the law or executive order authorizing the census.

3. Procedures. Organization of field staff. How to obtain supplies and transmit completed questionnaires. What to do if respondent wants questionnaires left with him for completion or for mailing to statistical agency. What to do if respondent refuses to supply information. How to report establishments which changed owners during the census year. How to prepare progress reports.

4. The "skip-list", that is, a list of establishments the enumerator is not expected to visit. This list may include the establishments of multi-unit enterprises which are to be contacted directly by the central statistical agency.

5. Scope of the census. Definition of an establishment, and types of establishments to be included. Types of establishments commonly encountered which are not to be included.

6. List of questionnaires and other forms used in the census, and a brief description of the content, purpose and scope of each.

7. Detailed instructions for each question in the questionnaires.

8. Special instructions for particular types of establishments or industries. How to handle frequently encountered mixed activities (for example, manufacturing and trade, manufacturing and agriculture). How to enumerate household industries.

9. Multi-unit enterprises not on the "skip-list". How to identify establishments or headquarters offices of multi-unit enterprises.

10. Alphabetic index to the contents of the manual.

Special arrangements for multi-unit enterprises

137. It has been suggested that consideration should be given to the enumeration of known multi-unit enterprises directly from the central statistical agency rather than by field interviewers. The names of such enterprises may already be known to the government. If not, or if the existing lists are inadequate and need supplementation, industry associations may be a good source for such names.

138. The main advantage of this method of canvassing is that less time will be required to perform the enumeration. The field interviewer encountering an establishment of a multi-unit enterprise may spend a long time explaining his mission, only to discover that such matters are handled by the company's head office located in another interviewer's district.

139. In some countries, experienced members of the central statistical agency have undertaken to interview officials of multi-unit enterprises in advance

of the census. The purpose of such interviews is two-fold: (1) to gain the co-operation of the company officials and impress them with the importance of the census and (2) to facilitate the actual enumeration by obtaining a list of the company's establishments and information on the activities conducted at each location. This list can later serve as a basis for the selection and forwarding of the proper census questionnaires to the company. The head office of the company in many instances will prefer to supervise the filling of the questionnaires and to transmit them when completed to the statistical agency.

Coverage control

140. In a complete field enumeration, such as a canvass undertaken to establish an industrial directory, it is fundamental that information should be secured from every establishment within the scope of the census. It is also desirable to avoid duplication, which can arise when two field interviewers make a report for the same establishment because of misunderstanding.

141. An important aid to field interviewers is a set of detailed maps covering each of the enumeration districts (ED's). Such maps should clearly define the boundaries of the districts: if the boundary is a street, the allocation of establishments on each side of the street should be indicated. If accurate and detailed maps are available in the central statistical agency, ED maps can be prepared in advance. Otherwise, it may be necessary for the regional or local offices to define ED's and prepare ED maps for the interviewers.

142. ED's defined for censuses of population may serve as a starting point for the definition of industrial census ED's. Estimates of the number of industrial establishments in each of the population ED's should be obtained, and combinations made which would tend to produce ED's of approximately equal industrial importance.

ED control record

143. In addition to pocket manuals and ED maps, systematic canvassing will be aided by a control record on which is entered identifying information concerning each establishment encountered. The control record may be in the form of a set of numbered sheets, each of which provides for entries concerning

20 or more establishments, and the set covering an entire ED. A typical record of this kind might provide for the following items:

1. Serial number of entry. In some systems, this number is also entered on the corresponding questionnaire.
2. Name and address of the establishment.
3. Industrial activity or kind of business.
4. Is the establishment within the scope of the census?
5. Is the establishment part of a multi-unit establishment?
6. If a questionnaire is left with the respondent, the date it is to be picked up.
7. Date the completed questionnaire is obtained.
8. Date the questionnaire is forwarded to the central statistical agency.
9. Name of person interviewed.
10. Remarks.

144. Since the control records provide a detailed account of the progress of enumeration and the disposition of the completed questionnaires, they should be preserved for use as reference documents in later stages of processing. Separate questionnaires can be misfiled or lost, for example, and reference to the control record will indicate whether the questionnaire file for an ED or a local area is complete.

Screening of completed questionnaires by field staff

145. For practical reasons, it is desirable to detect and correct errors in the reported data as quickly as possible after the questionnaires have been completed. If there is a delay in re-interview or questioning of the respondent, the person who supplied the data originally may no longer be employed by the company, the company may have gone out of business, or the records may have been discarded. On the other hand, it is ordinarily not feasible or desirable to conduct complete editing operations, such as those discussed in the next chapter, in the field offices. The staff should, however, be able to perform a screening operation which involves a limited amount of checking of data. This might consist of seeing that the questionnaire is properly identified, that no required figures are omitted, and similar simple checks. A slightly more complicated but basic check would be to add the reported costs

(wages and salaries, materials, etc.) to see that they do not exceed the reported value of shipments and sales of services. If they do, the report should be returned to the enumerator or given to another interviewer for verification or correction.

146. Such a field review also provides some control over the performance of individual enumerators. When questionnaires are found to be faulty, the enumerator responsible should be required to correct them, if time permits. An excessive number of faulty questionnaires may mean the enumerator is incompetent and should be dismissed.

Check interviews

147. Another means of controlling and improving the quality of the reported data has been found desirable in some countries. This involves the re-interview or check interview of respondents by the local supervisor or by a specially trained enumerator. Check interviews are ordinarily conducted for only a small sample of respondents. Since the interview involves an additional demand on the time of these respondents, the interviewer should be skilled at explaining the purpose of the interview, which is basically to increase the accuracy and usefulness of the census. He should then proceed to secure new answers to the questions. If the respondent has kept a copy of his original report, the interview may involve no more than comparing the copy with the original.

148. The check interview need not cover all the items on the questionnaire, but could be limited to a few of the most important. Check interviews are of value in disclosing unacceptable practices on the part of enumerators, such as the fabrication of data.

Receipt of completed questionnaires: mail collection

149. If a directory has been established, and questionnaires mailed either to all respondents or to a sample, the mailing-list information should be transferred to a set of "check-in" cards. Space should be provided on these cards for the entry of the date the questionnaire was mailed and the date it was received. The same card, if desired, can be used for recording the dates of subsequent contacts with the respondent, such as of correspondence to obtain corrected figures or to verify the accuracy of the figures reported.

150. Since cards are sometimes lost or misfiled, a master mailing list should also be prepared. Mechanical transfer methods, such as addressograph plates, make it possible to print lists on long strips of paper which may then be stapled and bound to ensure completeness in subsequent handling. Each strip should also be numbered.

151. As the completed questionnaires are received from respondents, the corresponding control card is removed from the file. The date of receipt is entered on the questionnaire and also on the card, which is then put in a second file representing establishments for which questionnaires have been received. In large-scale receiving operations, the clerks are sometimes instructed to "check-in" obviously incomplete questionnaires, including those returned blank, and to treat letters from respondents as the equivalent of questionnaires. The correction of these problem cases, as well as others requiring further contact, is left to a subsequent screening operation.

152. At some time after the due date for filing questionnaires has passed, the original file, which now contains cards only for the establishments which have failed to return questionnaires or other required material, can be used as a mailing list for a follow-up letter reminding respondents of their obligation to report. A second or third wave of follow-up letters or field visits to the remaining delinquent respondents may be necessary.

153. For the largest establishments in the delinquent file, it may be desirable to schedule field visits at an earlier date. The preparation of reports by such establishments takes more time, and their importance in the statistical results is greater than is the case with small establishments.

154. It may be expected that not all establishments will report by mail, and, in the planning stage, some resources should be allocated to the field collection of data from large establishments which have failed to respond.

Receipt of completed questionnaires: field collection

155. Many of the procedures suggested in the preceding section are equally applicable to field collection. The principal differences arise because in field collection two sets of receipt records are usually maintained: one set in the local or area offices and another in the central agency. In some systems, the ED control record described earlier is also used as another

check on the receipt of questionnaires. In mail collection, the check-in files may be arranged alphabetically or by some numerical system. In field collection, they should normally be arranged by area to facilitate the handling of follow-up inquiries received from the central agency. A special file, duplicating the establishment cards in the central file, might be maintained in the central agency for multi-unit enterprises. The special file would facilitate correspondence with the central offices of such companies regarding missing reports. In any case, it should be made clear to all concerned where the responsibility for contacting multi-unit enterprises lies.

156. In the following flow description of a typical questionnaire handling and receipt operation, it is assumed that the central agency deals directly with the local offices, that it is desired to maintain a constant flow of completed questionnaires to the editing and coding operations, and that the central agency is responsible for the collection of data from multi-unit enterprises.

1. Interviewer collects questionnaires and prepares ED control record.

2. Interviewer turns in completed questionnaires weekly to local office. When an ED is completed, he also turns in the ED control record.

3. Local office records receipt of questionnaires in check-in file, screens questionnaires for acceptability, returns some to interviewer for correction, and forwards the rest to central agency daily. Date of forwarding is recorded on check-in cards.

4. With the last batch of questionnaires for an ED, the ED control record is forwarded to the central agency. Date of forwarding is recorded.

5. Central agency records receipt of questionnaires in check-in card file, and forwards them to editing and coding section.

6. Questionnaires rejected by editing and coding section are returned to check-in section. These questionnaires are returned to the field for verification and correction. Date of return is recorded in the check-in card file and also on the ED control record, if one has been received.

7. As ED control records are received, they are compared with the check-in card file to see that all the required questionnaires (1) have been received in the central agency and (2) have not been returned to the local office for correction. If some of the

questionnaires are still in process in the editing and coding section, this comparison is postponed until editing and coding are completed.

8. A record of incomplete and completed ED's is maintained in the central agency. As enumeration nears completion, the central agency gives local offices a list of incomplete ED's and asks for a report on the status of each.

VI. PRE-PUBLICATION PROCESSING

Editing

157. It is inevitable that many errors will be found in the inf rmation entered on the returned questionnaires, despite careful planning and collection of the data by an efficient and well-trained staff. If for no other reason, errors will occur because the concepts behind industrial census inquiries are difficult to grasp and misinterpretation of the questions can easily arise. In addition, some entries may be treated as erroneous because of processing requirements: values expressed in common fractions, for example, are not acceptable for punch card transcription but may be quite accurate by ordinary standards.

158. The basic purposes of editing statistical questionnaires are (1) to detect errors in the reported figures and (2) to prepare the questionnaires for punching and tabulation. The former is by far the more difficult process, because many entries may appear to be acceptable, but when compared with other information are seen to be questionable or obviously wrong. Sometimes there are two or more related figures in doubt, and it may be obvious that at least one of them is wrong, but impossible to decide where the error lies. Some errors can be corrected in the editing process, but in other cases it may be necessary to obtain new information from the respondents. This is true even in countries with considerable experience in industrial censuses; in many such countries the proportion of questionnaires where the respondent has had to be contacted a second time for the verification or correction of data has amounted to 25 per cent or more of the total number.

159. The preparation of questionnaires for punching and tabulation, on the other hand, usually involves simple routine changes such as the rounding of numbers, striking out excess figures, and entering control data.

160. A clerical editing operation of some kind is always necessary in industrial census processing, regardless of the method of tabulation. If adding-machine tabulation or electro-mechanical tabulation is planned, the function of the clerical section is usually to edit the questionnaires completely. The section will also obtain new information from the respondent if required

and make all changes necessary to allow the questionnaires to proceed smoothly through subsequent operations.

161. Since it is inefficient and wasteful of resources to treat all deficiencies in questionnaires alike, however trivial they may be, some standards and criteria for editing should be established. In general, the amount of effort expended on detecting and correcting errors and omissions should be in proportion to their probable effect on the published data. It follows that the reports from large establishments should normally be more carefully edited than those from small establishments.[1/]

Work units

162. Assigning work, keeping records, and making progress reports will be simplified if the questionnaires are sorted by type and bundled together in work units of convenient size.[2/] The work units should be numbered serially. A central record should be kept showing the number of questionnaires in each work unit and a running record of its history, for example, the date it was given to an editing clerk, date of completion, number of questionnaires returned to the field for review, and number transmitted to the card-punching section. Work units are usually made up of the same type of questionnaires, and it may be desirable to sort them in other ways. The questionnaires for small establishments, if these establishments are more or less fully enumerated, can be assmebled in separate work units and edited by less skilled clerks. The instructions given to these clerks may, in effect, provide for the acceptance of errors which would be unacceptable for larger establishments. Conversely, the questionnaires for the largest establishments may be sorted out and edited by supervisors, or other persons with ability and experience above the average. An intensive edit of this kind would also be desirable for questionnaires obtained by sampling the small establishments. Limits may be set for the magnitude of entries which are to be questioned: for example, re-questionning of a respondent may not be permitted if the value

---

1/ Except where the coverage of small establishments is based on a sample.
2/ Usually not more than 100 questionnaires should be assembled in one work unit.

of an item is below a certain limit. Instead, a procedure for substituting estimates for incorrect figures may be established.

163. Certain types of checks may be performed only for those questionnaires contributing most importantly to the published data. Checking of the data on output of products against related government series, for example, may be limited to establishments whose production of the specified products is above a specified figure.

Editing checks

164. Various types of editing checks have been employed. The number used and the extent of checking usually depend on the time and resources available.

Omissions. This involves finding items where an answer is required, but none appears. The procedure may not be simple for items which are not applicable to all types of respondents. It may be necessary to instruct the clerks, for example, not to question the omission of a reply to an inquiry about the cost of purchased repair and maintenance services, because not all establishments purchase such services. If it is particularly important to distinguish failure to report from inapplicability in such cases, the question may be formulated as in the following example:

> Did you purchase any repair and maintenance services during the census year? ☐ Yes ☐ No
>
> If you answer is "Yes", how much was paid for these services? $_____

If "Yes is checked and the value nevertheless omitted, it would be clear that the respondent had failed to report a figure which should have been reported.

Internal consistency checks. These consist of checks to determine whether the reported components of a total are equal to the reported total, whether the answers to the same query in different parts of the questionnaire are the same, and whether certain ratios are reasonable in the light of experience or other criteria. On a questionnaire calling for details on both the input of materials and the output of products, a "presence check" may be possible. As an example, an establishment manufacturing bread should use flour, and the clerk should be instructed to question the bread entry if no entry appears for flour. It may be noted that checks which specify the amount of raw material that should be used lead to complications and should ordinarily be avoided.

Judgement as to the reasonableness of ratios, such as the ratio between the number of operatives and total wages paid to operatives, that is, their average annual earnings, may be based on a constant range of values given in the instructions. Ratios above or below these acceptable values would be questioned by the editing clerk and unless an obvious error could be identified and corrected, the questionnaire would be returned to the respondent for correction.

The acceptability of a ratio may also be decided after comparing it with the same ratio for other establishments of a similar size, in the same industry, or in the same geographic area. In infrequent inquiries such comparisons are usually not feasible because of the difficulty, at the time of editing, of finding a sufficient number of questionnaires with similar characteristics.

External checks. The reported data and the derived ratios may also be compared with earlier figures for the same industry. Data from other countries can also be used under appropriate circumstances. Such external data, when available, are normally used as guides in setting limits of acceptability which become part of the instructions for the editing clerks.

During the course of editing, the effectiveness of the various checks employed should be examined occasionally. If application of a given editing check results in rejection of all questionnaires, for example, its tolerances are almost certainly too narrow. If, on the other hand, no questionnaires are rejected, the check should be probably discontinued or its range reduced.

Computerized editing

165. In some countries, a considerable amount of the editing formerly done clerically is now being performed on computers. As mentioned above, an abbreviated editing operation is necessary to prepare the questionnaires for punching and tabulation. In computerized editing, the detection of errors is postponed until the cards have been processed through the computer. 3/ The computer is programmed to calculate ratios, make comparisons, substitute estimates (under specified conditions) for unreported or erroneously reported figures, and "flag" or identify questionnaires which need to be reviewed manually and possibly returned to the respondent for correction. In general, the computer will perform these operations more rapidly and accurately than

3/ On some computers, the cards must first be copied on magnetic tape in a "card-to-tape" device.

clerks. There are certain important limitations, however, which must be considered:

1. Many of the questionnaires that are flagged by the computer for manual review will require correction. It is costly to correct punch cards and computer tapes. Hence if a great many questionnaires will eventually require manual handling, repunching of cards, etc., computerized editing may result in a much higher over-all cost than clerical editing.

2. The particular computer which is to be used may have too little storage capacity to perform all of the editing checks efficiently in addition to other computerized operations such as coding and tabulation. The programme should be tested as much as is necessary to determine the effect of storage limitations.

3. A computer with sufficient capacity is capable of far more checking than is usually feasible in a clerical operation. The computer may be able to perform several times as many internal-consistency checks, for example. Some census planners, enthusiastic over this extraordinary capability, have specified so many editing checks that practically every questionnaire has been flagged for manual review. As in clerical editing, the usefulness of every individual check should be considered in relation to its cost and also the over-all cumulative effect of a large number of checks.

4. Computerized editing is necessarily performed at a later stage than clerical editing. As mentioned in the previous chapter, it is desirable to question respondents as early as possible after the questionnaires are received. With a temporary field staff, computerized editing may entail so much delay in the return of questionnaires to the field that the staff may be no longer available. Thus the timing of re-questioning should be considered carefully in relation to the timing of computerized editing.

5. An important difference between computer processing and the older methods such as hand and mechanical or electro-mechanical processing is that misinterpretations giving rise to repetitive or systematic errors are often far more difficult to correct. Many costly tabulations have had to be scrapped because the computer programmer did not quite understand what the statistician wanted. Under the older processing methods, such mistakes were usually discovered and rectified before much harm had been done. The tremendous speed of computerized processing, however, means that entire projects are often completed before the systematic errors are discovered.

For this reason, sufficient time should be allowed for the thorough testing of computer programmes before large amounts of

data are processed. The usual procedure is to prepare a "test deck" of punch cards similar to those which will eventually be processed, but which contain errors and other material deliberately planned to bring all of the features of the programme into operation. The programme can be executed on the test deck at relatively little cost, and the results of the test used to correct the programme where necessary.

Coding

166. Coding is the transformation of information to numerical form with the object of simplifying tabulation. The information to be coded may be numeric or non-numeric. Assigning size codes based on numbers of persons engaged transforms one set of figures to another and simpler set for tabulation purposes. Geographic coding, on the other hand, transforms the names of provinces, counties and cities to a compact numerical form. Numeric codes are used also to denote type of ownership, industrial activity and other characteristics of this kind.

Pre-coded questionnaires

167. It is good practice is designing questionnaires to take into account first the ease of reporting and second the ease of processing. The pre-printing of code numbers on questionnaires is primarily for ease of processing, but in some systems reporting is also facilitated since the code numbers also appear in reference manuals available to enumerators and respondents.

168. A common example of the pre-printing of code numbers is found in connexion with the reporting of individual products and services. A non-pre printed inquiry on this subject would provide several blank lines and ask the respondent or interviewer to write in product descriptions and enter the numerical data. A pre-printed and pre-coded inquiry, on the other hand, would provide the code numbers and descriptions of the several products made by the industry and in effect ask the respondent to (1) select the appropriate line or lines and (2) enter the data.

169. The pre-printing of codes for industrial activity is usually not feasible, because the proper code number is frequently selected from several alternatives on the basis of the primary activity of the establishments. Codes for products, materials, type of ownership, and several other items can, however,

be entered in advance, and pre-printed code numbers are commonly used in connexion with them.

170. The advantages of the pre-printing of code numbers are (1) no subsequent clerical coding operation is necessary and (2) punching is facilitated because printed numbers are easier to read than hand-written numbers. There is a drawback in that the respondent may enter his data on the wrong line, and it may not be possible to detect the error. This is offset to some extent by the elimination of the possibility of code numbers being incorrectly entered by coding clerks.

Clerical vs. computer coding

171. The use of the computer has greatly reduced the amount of coding required, especially the amount that needs to be done clerically. In the first place, codes whose primary purpose is to facilitate the sorting of punch cards are not required. The computer can operate about as well on a data item like number of employees as on a one-digit size code based on the original figure. Moreover, if a one-digit size code is required for other reasons, the computer can readily assign one. Secondly, clerks are notoriously weak when it comes to assigning codes based on a calculation and/or comparison of relative magnitudes, while the computer is always accurate in this respect if it has been properly programmed. In general, the computer is preferred whenever a code number is to be assigned based on numerical data, particularly if complex calculations are involved. Coding systems may be simplified also: while mechanical tabulation may be facilitated by a six-digit number in which each position is significant (for example, the first two digits represent the province, the second two the country, and the third two the city), computer operations may be equally efficient using a three-digit number assigned serially to cities with the reference to county and province kept in computer storage. Many of the practices geared to hand or mechanical processes need to be re-examined if a computer is to be used.

172. Assignment of industrial-activity code numbers may be difficult to programme for the computer if industries are defined in any other way than by the simple predominance of a product or group of products. Industry codes based only on the respondent's description of his activities cannot

ordinarily be assigned by the computer. Geographic codes, likewise, normally require assignment by a clerical section, and there may be other items involving similar non-numeric information. 4/ Thus, coding clerks will be required even though a computer is u ed.

Procedural matters

173. If a substantial number of questionnaires are to be clerically edited and coded, it is usually preferable to break down the operations into short routines. Where this has been done, clerks can handle all but a small fraction of the questionnaires after a short learning period. The extra handling involved is usually compensated for by the increased speed of processing and the fact that fewer highly skilled supervisors or industry specialists will be needed.

174. Other procedural rules which have been found useful in practice are:

1. All entries in the questionnaire should be made in a distinctively coloured pencil or ball pen. The colour chosen should contrast with both respondents' entries and field enumerators' entries.

2. The colour used by coders should differ from the colour used by editors. Some countries also differentiate corrections received from respondents by entering them in a third distinctive colour.

3. When a correction is made to a respondent's entry or a field interviewer's entry, the original data should not be erased or obliterated. Instead, the editing clerk should strike a line through the entry and enter the corrected figure above it. If space is limited, the corrected figure may be entered elsewhere and its proper position indicated by a guide line.

4. Space should be provided on the questionnaire for the initials of editing or coding clerks and the date of processing. If not, this control information should be entered in the margin of the form in a prescribed position.

Tabulation

175. Modern tabulating equipment varies in complexity from the simple desk adding machine to the high-cost, high-speed, large-capacity electronic system with its central computer and specialized auxiliary devices. Theore-

4/ With large-capacity equipment, however, some success in the computerization of geographic coding has been achieved.

tically, there is an optimum set of equipment for every job. In deciding on requirements for an industrial census, the number of questionnaires, the time available for tabulation, the kinds of tabulations are among the important considerations. If computer editing as well as tabulation of questionnaires is planned, a more complex system will be required.

176. Although computerized systems are the most popular today because of their speed and adaptability to different requirements, other systems are still in use and may in fact be superior to computers in particular situations. As a practical matter, it is often possible to make use of tabulation equipment already owned by the government. If such equipment is available for the industrial census at moderate cost, the question of its complete suitability may never arise.

Adding-machine tabulation

177. If the number of questionnaires is small and the intention is to produce sub-totals for only one or two characteristics of establishments, tabulation on adding machines may be faster and more efficient than other methods. As an example, suppose all items are to be aggregated by (1) industrial activity and (2) geographic area. After editing and coding, the questionnaires are sorted by geographic area and then by industry. The data for each group of questionnaires representing an industry within an area are then summed on adding machines. 5/ The machine should be of a type which prints out the data on paper for verification. After their accuracy has been verified, the totals obtained can be posted to summary work-sheets, and these summed again to produce (1) national totals by industry and (2) area totals for all industries combined.

178. If an additional tabulation is required, for example, a table showing some or all of the items classified by size of establishment, the questionnaires must be sorted again and the operations repeated. The amount of hand-sorting usually becomes too time-consuming if the data are to be tabulated by more than two characteristics, and therefore tabulation by other means should be considered. There are devices which facilitate sorting, such as

---

5/ If there are many items, the work will be facilitated by using a machine which records several columns at once. This type of machine is sometimes called a "book-keeping" machine.

cards with marginally punched holes, on which the data are hand-posted or typed. These devices are sometimes useful in repetitive surveys but in infrequent inquiries are not likely to offer any advantages over adding-machine tabulation.

Punch cards

179. The great advantage of punch cards is the speed and accuracy with which they can be sorted and re-sorted to produce new tabulations. They are particularly useful in industrial censuses, where the basic data on activity may be classified or cross-classified by several establishment characteristics such as kind of activity, size, location and method of operation.

180. A common type of punch card is approximately 8 x 18 cm. in size, and is designed to receive 80 or 90 columns of data in the form of holes punched in each column. These entries can represent numbers from 0 through 9, alphabetic characters, punctuation and other symbols. The punching machines are of several types: some are suited only for punching numerical data while others will also handle alphabetic information. Some types of punching machines automatically print each character at the top of the card as it is being punched and this can assist in the verification of punching.

181. The initial cost of punching cards is not inconsiderable, but once the cards have been punched and their accuracy carefully verified, they may be sorted rapidly on electro-mechanical or electronic qquipment into the groups desired for tabulation. After tabulation, they may be re-sorted quickly for other tabulations. Finally, they may be retained for possible future use in special tabulations not anticipated at the time of processing the census results.

Punching errors

182. The punching of cards is an operation that cannot be performed without error, and frequently the errors are more significant than those which arise in hand transcription. By failing to skip spaces properly, for example, the punch-card operator can easily transform the number 123 into the number 123,000. If such erroneous numbers are not corrected, they can cause much difficulty when they are subsequently discovered in the review of tables prior to publication.

183. The usual method of detecting punching errors is to verify the punching, that is, to require another operator to punch the data on a verifying punch machine which stops when no punch appears on the card in the expected position. In general, however, verification is only partially effective in tests where incorrect figures have been purposely punched on cards; verifies have always failed to identify and correct some of the errors. In a verification operation of average efficiency, it may be expected that about 95 per cent of the errors will be detected.

184. In computerized tabulation systems, a recent innovation makes possible the automatic verification of certain coded information. This is the check digit, a number computed from the code number itself, which then becomes part of it. With a six-digit code number, for example, the check digit becomes the seventh digit. Check digits can be calculated in many ways. 6/ In a computerized editing operation, the check digit is compared with the rest of the code number, which is flagged as incorrect unless the check digit passes this inspection.

185. Check digits are useful as added protection in connexion with the punching or transcription of pre-printed code numbers, such as those for products and materials, and also in the case of identification numbers. Mistakes in these numbers can be especially troublesome.

Design of cards

186. Punching and verification will normally be faster and more efficient if the cards are designed so that the figures are punched in the sequence in which they appear on the questionnaire. In electro-mechanical tabulation, it is desirable to group on one card the items which are to be tabulated together.

6/ In one method, the odd-numbered digits are summed and multiplied by 3; the sum of the even-numbered digits is added to the product and, finally, the result is subtracted from the next higher number ending in 0. For instance, with the code number 123456 the check digit would be computed as follows:

$$3\ (1 + 3 + 5) + (2 + 4 + 6) = 39,\ 40 - 39 = 1$$

The complete number to be punched would be 1234561. If it were incorrectly punched as, say, 1324561, it would fail the test. Transposition of digits as in this example is a common type of punching error.

If, to facilitate punching, the items have not been punched in the proper sequence for tabulation, new cards can be created on a reproducing punch machine, but this will add to the cost. With computers, on the other hand, data fields can often be regrouped at little or no cost as they are transferred from cards to magnetic tape.

187. If there are two or more cards per questionnaire, the amount of identification and classification information required on each card is also affected by the method of tabulation. With electro-mechanical tabulation, it is customary to reproduce on each card the establishment identification number, the industry and geographic codes, and all the other codes to be used for the classification of the data, such as size codes. With computerized tabulation, the full information is needed only on the first card; on subsequent cards, the computer can associate classification codes and data without repeating the full set of codes on each card.

Control records

188. Punch cards are valuable records and precautions should be taken against loss or damage. Questionnaires should be maintained in numbered work units of a convenient size, and the number of cards derived from each work unit counted. After punching and before tabulation or transcription to computer tape, it is usually desirable to create larger work units. The cards should be counted after each stage, and the over-all counts compared frequently with the control records for the previous phase.

Production rates

189. Since card punching is a repetitive operation, production standards based on the achievement of experienced operators can be established if desired. The rate to be expected will vary considerably, depending on the number of key strokes required per card, how many alphabetic characters are present, and the extent to which key-punch operators are motivated to exert themselves. Experience indicates that production should range from 75 to 175 cards per hour in industrial census processing.

Card sorting

190. After punching and verification, the amount of mechanical sorting required depends on the tabulation method. With most types of electro-mechanical equipment, sorting by classification code number is required before each

tabulation. With some of the smaller computers, particularly those which do not receive magnetic tape, a similar amount of advance sorting is required. With larger computers, some or all of the sorting required may be performed by the computer itself, either while the records are being transferred to magnetic tape, or at a later point. Card sorting requirements, for this reason, cannot be estimated until the tabulation method has been determined.

Electro-mechanical tabulators

191. This term is used to describe the older kind of equipment for tabulating from punch cards. In some respects similar to complex adding machines, these devices have provision for handling about 120 data positions (digits) simultaneously; the adding-machine operator is in part replaced by automatic controls. The statistical results are printed out on a wide strip of paper.

192. Electro-mechanical tabulators vary in speed and capacity but a common type can process 40,000 punch cards in an average day. The speed depends in part on the number of sub-totals and totals required, because the cards stop passing through while these sums are being generated and printed. Some models are capable of calculating and printing out minor, intermediate, and major totals for the same data fields - for instance, data for cities and provinces, and for the nation as a whole. Some can produce new punch cards showing summary totals for selected categories; these summary cards may then be used in subsequent tabulations. The machines can count the cards as they pass through and if each card represents an establishment, this count may represent a statistic for publication.

Electronic computers

193. Digital computers are also like adding machines. Their unique characteristics are the speed with which calculations are made and their capacity for storing information for later use.

194. Electronic data-processing computers suited for industrial census operations are of two general types: (1) those which process punch cards directly and print out results on paper in a manner similar to the electro-mechanical devices described above, and (2) those which can receive instructions and data on magnetic tape and can produce results in the form of magnetic tape. The experience of several countries indicates that the type of

computer best suited to industrial censuses in countries at an early stage of statistical development is one of the relatively simple systems using magnetic tape. An advantage of taped records is the speed with which they can be processed. If several tabulations are to be made from the same basic records, the availability of taped records may permit important gains in processing time. In other respects, the card-compatible and tape-compatible systems are similar.

Programming

195. The tabulation of industrial census data on adding machines can conceivably be started on a few hours' notice, provided machine operators are available. In computerized tabulation, however, considerably more notice is necessary. First, a trained programmer must be found and he must take time to familiarize himself with the objectives of the census, particularly with the plans and specifications for editing and tabulation. 7/ Next, he should plan the broad outlines of the computer programme and set down its essential features in the form of a flow chart. At this stage, the programmer and the planning staff should discuss the chart in detail, taking as much time as is necessary to make certain that it is completely understood by all concerned. The objective of this review is to bring misconceptions and misinterpretations to light which, if not detected, may be extremely costly to correct later on.

196. After approval of the general plan, the programmer will begin to write the programme itself. The time required depends on the complexity of the programme and the type of computer to be used. Simple programmes may take only a few days to write, others may take several weeks. After it is written, the programme should be tested. This is usually done with a "test deck" of punch cards of the type to be tabulated, which has been designed to bring out the features of the programme and to represent the problems anticipated in dealing with the actual data. After testing and correction, the programme is ready for use in editing or tabulating the reported data.

7/ Some planners of statistical projects have taken courses in computer programming, not in order to write programmes, but to improve their ability to communicate with programmers. If a well-qualified programmer is available, this should not be necessary.

Design of machine-output tapes

197. In both electro-mechanical and electronic systems, the statistical results are usually printed out on paper tape about 40 cm. wide. The location of every digit and every blank space on this machine-output tape has to be planned in advance, and it should be laid out in such a way as to facilitate its subsequent use in the review of data, correction of errors, and posting of tables.

198. In recent years, progress has been made in designing machine-output tapes which, as they emerge from the tabulating machines, are ready to be photographed and printed by the offset process. This method is successful where the data can be presented in the form of large numbers of similar tables. In most industrial censuses, however, there are many dissimilar tables and hand-posting and typing is still the common practice. The form of the machine output tapes is less critical when this method of producing final tables is adopted.

199. The following are some practical rules to be considered in designing machine-output tapes:

1. The review of the data should be facilitated wherever feasible by the calculation of ratios and other review aids. Ratios such as average annual wages and salaries, and cost of materials as a percentage of the value of output, are particularly helpful in this respect.

2. To avoid revealing confidential information, it is usual to examine publication cells to determine how many establishments or enterprises are represented. On many tabulation machines it is possible to print out these numbers alongside the cell for review purposes. Some equipment can also print out the identification numbers and amounts reported by the largest enterprises represented in the publication cell.

3. Space should be provided on the machine tape for the entry of corrected data. If corrections are to be punched on cards and carried through mechanically, consideration might be given to punching cards directly from the machine tape. The reviewers should use a distinctively coloured pencil or pen to make corrections.

4. Strips of paper showing the column headings should be pasted on machine tapes so that reviewers can readily identify the data shown. In computerized operations, this may not be necessary if the computer can be programmed to print column headings directly on the tape.

5. If the review specifications call for one figure to be compared with another, the two figures should be close together on the tape. Where owing to conflicting requirements this cannot be done, strips of pasteboard with rectangular "windows" cut out to reveal the pertinent data may be used.

6. In tabulations based on samples of establishments, the sampling weights are usually applied to the totals for each sample group at the tabulation or review stage. Even where weights are applied in the tabulation process and aggregates ready for publication appear on the machine-output tapes, it is usually desirable to print out the unweighted data for review purposes. If tests indicate that many corrections to the basic data will be required as a result of the tabulation review, it may be best to postpone the weighting of sample data until all corrections have been made.

Internal-consistency checks

200. The basic review of tabulated totals can be very similar to the earlier editing of the questionnaires. Most if not all of the checks applied to individual establishments should again be applied to the statistical aggregates at this stage, when totals are available for geographic areas or other groupings of statistical units.

Data available from other sources

201. The statistical staff should make use of any information available which may tend to confirm or conflict with the census findings. The comparisons with independent data are usually not routine and should not be entrusted to clerks.

VII. PUBLICATION

General remarks

202. Publication is the final step in the statistical process by which the industrial census results are made available to data-users. Within the limits of quality standards and of safeguards taken to prevent the disclosure of confidential information, a statistical agency should publish, at least at the national level, all of the basic data it collects. Some selection among the many conceivable cross-classifications of basic data with establishment characteristics much be made, however, if only because of budgetary limitations. The purpose of this chapter is to offer suggestions on the choice of tables for publication, on the timing of the publication of the results, and on the content of the text which should accompany the tables. In the discussion which follows, the tabulations considered are developed from those listed in International Recommendations for Industrial Statistics. 1/

203. In chapter IV, it was pointed out that early planning of the content of statistical tables has many advantages. Errors in questionnaire design will be brought to light and corrected and the objectives of editing, coding and tabulation will be more clearly understood. If a set of table outlines is available to illustrate the discussion, the purposes of the census can be explained more simply and effectively. However, at the planning stage, it is not necessary or possible to determine the exact final form of the tables and the accompanying material such as text and footnotes. Decisions regarding these matters may be postponed until shortly before publication. The explanatory text, for example, should be based in part on the experience obtained in conducting the census and thus cannot very well be written until processing is completed or at least well advanced.

Disclosure of confidential information

204. In most countries, respondents are much more willing to supply accurate information if they are convinced it will not be used to their disadvantage, for example, by the tax collector or by their competitors. Thus, the con-

---

1/ Chapter VIII: Reference 1.

fidential treatment of individual reports facilitates the collection of data and increases the quality of the results. Confidential treatment extends not only to the questionnaires as they are being processed, but also to the published tables. The usual treatment is to suppress cells which, if published, would reveal the activities of a single statistical unit. In this situation, the unit is normally an establishment, but the rule is often extended to the activities of enterprises which control two or more establishments. The publication of a cell representing two statistical units is prohibited, because the activity of each unit might be revealed to the other. Thus publishable cells must include, as a minimum, data for three statistical units. 2/

205. In statistical tables with horizontal and/or vertical totals, a further complication occurs because if to avoid disclosure one of the components of a total is suppressed, at least one more cell must be suppressed or the "disclosure cell" can be derived by subtraction. In a given table, application of the disclosure rules may result in so many suppressions that the table is virtually useless. Such tables give a poor impression and should be omitted or combined with other tables unless there are compelling reasons for publishing them. For example, in a series of similar tables, each covering a particular geographic area, two or more areas could be combined to avoid this disclosure problem.

206. Some statistics are considered to be more confidential than others. The amount or value of specific products made and expenditures for fixed assets are examples of data generally considered to be sensitive. The number of establishments is generally regarded as non-confidential, unless the classification is so detailed as to reveal confidential information. For example, to indicate that there is one manufacturing establishment in a given city could not be regarded as confidential, but to indicate it as a cotton-textile mill with annual gross output defined within narrow limits, could well be so considered. Since the number of employees of an enterprise or an establishment is often mentioned in public by industrial executives, it is usually considered to be less confidential than other figures, particularly cost or output values.

2/ Some countries also follow a dominance rule for data considered especially sensitive. The rule involves suppressing cells where one or two units account for a very high percentage of the total.

Quality standards

207. Data should not be published when they are known to be statistically unreliable due to response errors or processing errors, or in the case of sample inquiries, to errors of estimate. Despite earlier precautions, the full effect of such errors on particular figures may not be discovered until it is time to publish the results. If it is recognized at this final stage that the quality of the statistics is very poor, it may be better to suppress them than to publish them.

Cross-classification

208. Nearly all published tables will consist of basic data items cross-classified by establishment characteristics. Among the basic data items are such figures as number of establishments; number of persons engaged; wages and salaries; cost of materials; gross output; and value added. For tabulation purposes, important establishment characteristics include location; industrial activity; size in terms of employment; and type of ownership. There are a large number of other characteristics which could conceivably be used for tabulation purposes: establishments might be classified in accordance with the size of ratios between basic data items.[3/] The number of tabulations that could be conceived is so large that in most countries the available funds could not provide for all of them. It is necessary, therefore, to try to select the few which will be of the most general interest. Examples of some of the more common types of industrial census tables are suggested by the table titles and brief descriptions which follow. This list is not intended to be comprehensive and is not necessarily in order of importance. Also, it may not account for all of the data collected in the census.[4/]

Table 1. Basic statistics for major industrial divisions, by geographic areas

This table would present basic data items such as number of establishments, number of persons engaged, number of employees,

3/ As an example, establishments might be classified in terms of the size of the ratio of wages and salaries to gross output.

4/ More detailed recommendations on the content of these tabulations will be found in tables 1 and 2 of International Recommendations for Industrial Statistics (Chapter VIII: Reference 1). The special recommendations for tabulations to be prepared as part of the 1973 World Programme of Industrial Statistics will be found in ST/STAT/44.

wages and salaries, gross output, value added, and expenditure on fixed assets. The major industrial divisions could be (1) mining, (2) manufacturing, and (3) electricity, gas and water supply. Since the industrial-division presentation is broad, it would be logical to show data here for all geographic areas, small and large, as well as national totals.

Table 2. Basic statistics by industry, for the nation as a whole

This table would present for the nation as a whole the same basic data items as in table 1, classified by kind of activity to at least the four-digit level of ISIC or the national equivalent.

Table 3. Basic statistics by industry, for selected geographic areas

This table would be similar to table 2, with separate sections for the principal geographic areas in terms of their industrial importance. The amount of detail shown by industry would differ from area to area, since it would be determined by the industrial composition of the area.

Table 4. Number of establishments, by industry and geographic area, by size of establishment

This table would present a detailed statement of the distribution of industrial establishments in terms of their location and size. Since the distribution of individual establishments is not normally regarded as confidential information, some countries have waived the disclosure rules for this kind of table.

Table 5. Cost of materials and supplies and related expenses, by industry

This table would present whatever data were collected on cost of materials and supplies, fuels, electricity, purchased industrial services, and related costs.

Table 6. Electricity generated, purchased and sold, and fuels purchased, by geographic area, for major industrial divisions

This table would present, for the electricity industry and other industries in terms of broad divisions, whatever data were collected on the production and use of electricity and fuels. Within the limits imposed by confidentiality restrictions, data should be shown for small geographic areas.

Table 7. Quantity and value of individual products and industrial services

This table would present, for the nation as a whole, whatever data were collected on products mined or manufactured and industrial services sold.

Table 8. Selected basic statistics by industry, by size of establishment

This table would present all or a part of the basic statistics shown in table 1 by industry and size class of establishment. According to the international recommendations, size classes should be determined by the average number of persons engaged during the year. The size classes recommended are 1-4, 5-9, 10-19, 20-49, 50-99, 100-199, 200-499, 500-999, 1000-1999, 2000-4999, 5000-9999, and 10,000 and over.

Table 9. Expenditure on fixed assets, by industry

This table would present, for the nation as a whole, whatever detailed data were collected on outlays for fixed assets during the census year. If only the total value were collected, it might be included in table 1 and other tables as a basic data item, making this table unnecessary.

Table 10. Value of stocks at beginning and end of year, by industry

This table would present, for the nation as a whole, whatever detailed data were collected on the value of stocks (materials, fuels and supplies; work-in-progress; and finished goods) at the beginning and end of the census year. If only the total value were collected, it might be included in table 1 and other tables as a basic data item, making this table unnecessary.

Preliminary publication

209. To be of greatest value, particularly in relation to current public or private planning, the results of the census should be published as soon as possible after the end of the census year. On the other hand, there are many features of industrial census statistics which (if of adequate quality) are still useful after years have elapsed. The census staff may be faced with urgent demands for early publication of results, but must weigh these against the need for thorough editing of the questionnaires and careful coding and tabulation. Efforts to complete these phases too rapidly may lead to earlier publication but, if the quality of the results is poor, may be self-defeating.

210. One commonly-used device for satisfying at least some of the demands for early publication is to select certain of the most important basic data items or tables for preliminary release. Tables 1 and 2 above, perhaps condensed to show less than the full detail suggested, could be considered for this purpose. Other schemes are to release individual preliminary reports on particular industries or geographic areas on a flow basis. Such reports would

contain summary data selected from the list of final tables; each report might have two or three such summary or abbreviated tables.

211. In adding-machine tabulation, and to a lesser extent in electro-mechanical tabulation, it is convenient to sort the basic records (questionnaires or punch cards) by industry or by area to correspond to the groups selected for the preliminary releases. The tabulation-review staff can then concentrate on one industry or area at a time. Electronic computers are apt to be so fast that all of the records are tabulated at one time, and the staff might be faced with a large backlog of industry and area reports to be reviewed and published. It may be better in these circumstances to plan to prepare general coverage tables for the first preliminary release, such as the combination of tables 1 and 2 suggested above.

212. Another method of compiling preliminary data which might be considered is the selection of a sample of questionnaires for processing and tabulation ahead of the rest. This procedure has seldom been successfully applied, however, because of the added cost of sample selection and control and the difficulty of assuring special treatment for sample questionnaires or punch cards when all the records are being processed by the same organization.

Technical statements

213. To assist users in understanding the published statistics, preliminary releases should include brief technical statements dealing with the scope of the census, the basic items appearing in the tables, and methods of establishment classification. A much more complete statement of this type should appear in the final publication. The following topics are suggested for inclusion:

1. Reasons for conducting the census and its legal authority. Uses of the results. Economic significance of the industrial sector.

2. Brief history of the country's previous industrial censuses, if any, and of other industrial surveys. If no historical data can be shown in the statistical tables, the reasons might be discussed.

3. Scope of the census in terms of the types of respondents and industrial activities covered. If any small size-groups or relevant activities are omitted, the probable effect of their omission on the results both in detail and over-all.

4. Statistical unit and methods of enumeration. Methods of locating and identifying respondents. Special procedures, if any, for enumerating multi-unit enterprises. Defects of enumeration and steps taken to correct them.

5. Follow-up methods. Treatment of statistical units which failed to report for one reason or another. If estimates for such cases are included in the published results, the proportion estimated and the general effect on the validity of the results.

6. Editing, coding and tabulation procedures with emphasis on the effect of the techniques on the published results.

7. Inquiries which were not completed, if any, and the reasons for not completing them.

8. Classification systems employed.

9. Definitions of the basic data items. Reproductions of the questionnaires and instructions.

10. Procedures and definitions for derived data items.

VIII. BIBLIOGRAPHY

1. International Recommendations for Industrial Statistics, Series M, No. 48. 54 pp. United Nations publication, Sales No.: E.68.XVII.10.

2. Industrial Censuses and Related Enquiries, Series F, No. 4:

   Vol. I. 384 pp. United Nations publication, Sales No.: 53.XVII.11(Vol.I).

   Vol. II. 359 pp. United Nations publication, Sales No.: 53.XVII.11(Vol.II).

3. International Standard Industrial Classification of All Economic Activities, Series M, No. 4, Rev. 2. 48 pp. United Nations publication, Sales No.: E.68.XVII.8.

4. Indexes to the International Standard Industrial Classification of All Economic Activities, Series M, No. 4, Rev. 2, Add. 1 (in preparation).

5. Standard International Trade Classification, Revised, Series M, No. 34. 135 pp. United Nations publication, Sales No.: 61.XVII.6.

6. Classification of Commodities by Industrial Origin, Series M, No. 43. 72 pp. United Nations publication, Sales No.: 66.XVII.7.

7. Bibliography of Industrial and Distributive-Trade Statistics, Series M, No. 36, Rev. 3. 139 pp. United Nations publication, Sales No.: 67.XVII.20.

United Nations publications are available from the Sales Section, United Nations, New York or Geneva.

/ In the final version of the Manual, the Bibliography will be extended to include a selection of pertinent national studies and reports. /