FACTUAL DATA BANKS IN AGRICULTURE

Proceedings of the symposium organized by the Commission of the European Communities

Luxemburg, July 12-13, 1977

Factual Data Banks in Agriculture

Factual Data Banks in Agriculture

Proceedings of the symposium organized by the Commission of the European Communities

Luxemburg, July 12-13, 1977

Centre for Agricultural Publishing and Documentation
Wageningen, the Netherlands
1978

The symposium was organized by and the publication of the proceedings arranged by
Directorate-General Scientific and Technical Information and Information Management,
Commission of the European Communities,
Kirchberg, Luxemburg

ISBN 90-220-0674-3

Cover design: Pudoc, Wageningen

Printed in the Netherlands

Contents

Preface

The exchange of information by experts is the essential feature and object of all meetings bearing the title 'symposium'. This was certainly the case with this symposium on data banks in agriculture within the European Community. It was organized by the Directorate-General for Scientific and Technical Information and Information Management (DG XIII) of the Commission of the European Communities, on the initiative of the Agricultural Working Group of the Committee for Information and Documentation in Science and Technology (CIDST).

The wide spread of subjects has been divided into three groups: animal production, including food technology; plant production, including soil science; and agricultural management information systems. Variances in description of the concept of data banks as well as a vagueness in conceptions of the word were and are typical of a first meeting on a subject which did not begin to flourish until the seventies.

The exchange of information was intended to be a basis for consideration of the possibilities for collaboration within the European Community and for the avoidance of duplication of effort in the various countries. This is of course particularly important in connection with the creation of a network for data communication in Europe (Euronet).

Discussions along these lines resulted in an ample harvest of ideas, advice, indications and recommendations on likely paths to success and on the pace of future development.

Material resulting from the symposium has been studied further by a committee consisting of:

U. Schützsack - Chairman of the Agricultural Working Group
P. Dagnelie - Vice-Chairman, in charge of the Agricultural Management Information Systems (AMIS) Group
V. Fischbach - Vice-Chairman, Animal Production Group
A.H. Strickland - Vice-Chairman, Plant Production Group
D.D. Singel - General reporter
H. Buntrock - DG XIII
R. Haas - DG XIII

and me as Chairman of the symposium and this committee.

This study has resulted in some 'conclusions and recommendations' offered to the Agricultural Working Group for further consideration. The publication of these proceedings is a further step on the road to the improvement of compatibility and collaboration between data banks within the Community.

Drs. K.E. Krolis, Chairman

Introduction

Udo Schützsack*, Institut für Dokumentationswesen, Frankfurt/Main, Federal Republic of Germany

The Member Countries of the European Communities (EC) can already look back on a long history of documentation and information (D and I) in the various fields of agricultural science. There have been, and still are, therefore, a whole series of excellent services, of regional as well as local interest, available to the various users, from the farmer to the scientist. Particularly outstanding in this context are the bibliographic services provided in the various Community languages.

Bibliographic data - Data bases

It was therefore only to be expected that the Agricultural Working Group (AWG) of the Committee for Scientific and Technical Information and Documentation (CIDST) of the European Communities (EC), which was set up in 1969, should first devote its attention to the bibliographic sector, with the emphasis on coordination of the existing services, and to the extension, in specific fields, of the range of research registers, bibliographies and abstracting journals in the form of data bases.

AGREP

The most important projects (among which I would number in particular those which have resulted in the expansion of the range of services available) include the creation of a permanent inventory of agricultural research projects in the Community Countries (AGREP). This project (which was initially organized by the Directorate-General for Scientific and Technical Information and Information Management - DG XIII) has reached the operational phase and now also receives the technical and financial support of the Directorate-General for Agriculture (DG VI).

AGRIS

The second major project is the common EC contribution to the information system of the agricultural organization of the United Nations (FAO), the World bibliography for agriculture, AGRIS. It consists in collecting and processing the literature of the Community

* U. Schützsack is chairman of the Agricultural Working Group of the Committee for Scientific and Technical Information and Documentation of the European Communities.

Countries and is backed by a wide range of supporting services in this field (EURAGRIS).

Factual and numeric data - Data banks

Early in its existence the AWG turned its attention to the collection, analysis and distribution of other data of importance to agriculture, i.e factual and numeric data. A distinction is made between collections of such data, known as *data banks*, and *data bases*.

In the EC there are a whole series of factual and numeric data banks. Their full scope can only be guessed at. Mention may be made in this context of a study financed by the Commission on the recommendation of the AWG; but this is restricted to French data banks and urgently needs extending to cover the other Member Countries.

Many of the existing data banks have been developed in a national context and are geared to the requirements of the respective countries, with the result that they are only of limited use to other countries; e.g. data collections on the structure of agriculture in a specific sector, data collections on price trends, animal husbandry, etc. But there are also national, bilateral and multilateral data banks which are used by one or all the Member Countries of the Community.

Mention should be made, in conclusion, of data banks which are worldwide in origin, coverage and use; e.g. a plant gene data bank and a feedingstuffs data bank. Information on both of these will be presented at this symposium.

From the technical viewpoint also, the available material covers a very broad spectrum, including agricultural policy and agricultural sociology, farming and meat production, including veterinary medicine, food technology and the nutritional sciences.

Aims of the symposium

The Symposium will come to grips with this collection of problems. The organizers are well aware that it can only constitute an initial exchange of views. But a certain number of indicative observations may safely be made at this stage:

- The main purpose of the Symposium is to provide a forum for the exchange of experience, not only at a technical level but also regarding methods. The following question springs to mind: "Is it possible to develop a data sheet containing a minimum number of categories which are common to all data banks, as in the case of bibliographic documentation?"
- The Symposium should provide an overall view of the data banks available in the Community; the organizers are fully aware that not all data banks can be represented here; but the basis for a complete survey does exist. The following question might be asked in this context: "Is it desirable to carry out a deeper and more comprehensive survey and, if so, what data should/could be collected? Also, what data are subject to secrecy rules?"
- The Symposium ought to examine the preconditions for possible Community action where the users' requirements are not already met by the existing services. In this context, it would be particularly

important to answer the question: "Which actions should be taken at a Community level and which should be pursued in a national context?"

I hope and trust that the speakers and other participants will have a variety of further ideas and expectations regarding this conference. It will then become clear if there are enough common interests to justify actions (whatever their form). It is with this in mind that I wish the Symposium every success!

The needs of data bank users and some data banks in food science

D.D. Singer, Ministry of Agriculture, Fisheries and Food, London, United Kingdom

Summary

The differences in the development, design, promotion and use of in-house data banks and internationally based data banks are outlined. Many of the difficulties experienced with internationally based data banks arise from separate motivation of the system organisers, those who produce the data, and those who will use it, especially when biological data are involved. Questions concerning validation, inexpert interpretation, misuse and security can lead to suspicion, distrust and misunderstanding of data bank concept. This situation must be combatted by a clear understanding of all the implications and by early and continuous contact between the sponsors, future users and system designers. Examples of data banks in the food and biological areas are discussed.

The Data Bank is nothing new. Since man first began to establish facts it soon became apparent that some form of record was essential if these facts were to be used in the future either by him who established those facts or by others. This need stems essentially from the limitations of human memory and intellect. Not all priests or rulers could be expected to remember lists of their predecessors or days when eclipses occur; few were expected to remember mathematical tables, even of the simplest kind. The data embodied in such data bases has important characteristics. First it is established by decree or by rigid experiment or proof. In the latter instance the procedure leading up to the production of the data can be verified by argument, further experiment or by the occurrence of a predicted event. There was, therefore, little doubt that the data was indeed factual. Second, the data was established for the most part by those who would be most likely to use it and was often intentionally withheld from general use as part of a social system which maintained the superiority of a ruling hierarchy. Third the cost of deriving the data and of maintaining it in organised form came out of what today we call public funds. Justification of the whole system was not questioned and the need was self-evident and unchallengeable.

This situation did not change in essentials for several thousand years. Scientific and technical data banks developed as the producer of data saw that order was necessary to employ his findings. Another need for data banks emerged - the need to find the gaps in existing knowledge. Eventually, and particularly as mechanical and engineering

sciences developed, the organisation of data became a commercial endeavour and data banks, in the form of books were published for the use of other workers within the same discipline as the authors or within an allied field. The works of Heilbron and Beilstein in organic chemistry, or Bentham and Hooker on British Flora or of Kaye and Laby on physical properties of materials were familiar to me as a student and were in sufficient demand to justify commercial production.

This situation has now changed. It seems to me that this change has been brought about not by any one factor but by a concatenation of circumstances, by a number of factors, that together have revolutionised previous concepts of not only how data can be stored and retrieved, but of how and by whom it should be used. The first of these factors is the development of the electronic computer to a point where it is possible to concentrate in a relatively small space enormous amounts of information in a form that permits retrieval in a great variety of ways. The second factor is the emergence of a new discipline, that of Information Science. In the UK alone there are now five universities affording the student an opportunity to obtain post-graduate qualifications in this subject; similar progress has been made in other countries. Third is the vast increase in the volume of work turned out by the world's technical and scientific institutions, an increase of such extent that it could in many cases not be dealt with in the absence of the computer and the information scientist.

There is, however, a fourth factor and one which poses a new challenge to those working in the biological sciences. This factor is crystallised in a statement contained in a paper written c. 1970 for the OECD Committee on Management of Environmental Quality and Information and Data. It said: "The environmental area is one in which scientific and technical information can no longer be considered the concern of scientists and technologists alone; decision makers and policy makers require information from a variety of sources, analysed and presented in ways that reflect their special requirements." This implies that the need for the data bank is generated not by those who produce the data, nor by those who will use it themselves in the course of further scientific and technical endeavour, but by others who need it to shape administrative policies, legislation and in the long run the destiny, perhaps, of our civilisation.

It is now the situation that three separately identifiable communities of workers are involved in the data bank: the data producer, working in the laboratory, who publishes in one form or another the results of experiment; the information scientist who organises the data and decides on the shape that the storage and retrieval system can take; and the user, possibly also a laboratory worker, but also very possibly a 'desk scientist' or an administrator or legislator without specialised relevant scientific knowledge or training. No longer is the format and content of the data bank under the control of the data processor and user.

This dichotomy, or to invent a new English word, trichotomy, introduces new difficulties and problems to the scientist, which in

this short paper can be discussed only briefly and not in any pre-conceived order of priority. The first which comes to mind is the question of the validity of data. In the bio-sciences, experiments are not as easily and readily repeated as they are in, for instance, the physical and engineering sciences or even the chemical and bio-chemical fields. Often the costs in relation to the overall importance to general scientific progress makes repetition unlikely. The materials and conditions involved are variable and not easily defined if they are capable of definition at all in our present state of knowledge. This is well known to the biologist, geneticist, medical investigator, agricultural or food scientist. In the course of his work he willl use his knowledge and accumulated experience to weigh up the significance of data and the use to which he will put it. If he is charged with the task of producing a data bank for use within his own organisation he will select which data he feels is reliable according to the prompting of these critical faculties and further, which requires further evidence and which might be rejected immediately. The information scientist working remotely cannot do this. Again the availability of the computer with its enormous powers of memory, tempts the information scientist - and others - to endow the data bank with a comprehensiveness that has rarely been previously considered or been possible. There is a tendency to include all data regardless of its validity and also of the likelihood of its ever being used or required with frequency. Information scientists are trained to consider how costs of data element retrieval rise as the number of data elements increases; what they cannot do is to predict the need for recall in any one case. The costs of some data banks are now rising so fast that it seems timely to debate whether the facility of rapid recall of data that is afforded is justifiable. After all scarcely any piece of data cannot be found at all by the traditional methods of library searching or by communication with a colleague. At another meeting which I attended recently the intention to put into a data bank the toxicological, biochemical, physical properties of two million environmental chemicals together with peripheral information of a commercial nature was mentioned. Such a facility would confer a tremendous advantage on those concerned with environmental problems, but would the cost involved really be justified by the saving in time, and would time be saved in the long run? If one hundred workers are available for two hundred days a year it would take one hundred years to assemble this data if each chemical occupied one man day - a gross underestimate.

Such problems are of concern to many in government and commerce who would be beneficiaries of such vast data bank enterprises. However, these are insignificant to many in industry who see the data bank as a great danger to the way in which they at present carry out their business. If, they say, we are compelled either by rigid legal mandate, or by persuasion of ethics, to give up previously privately held data to the data bank, are we not opening our business up to our competitors or to well-meaning but misguided pressure groups or to political agitators? Under the 'Freedom of Information Act' in the USA it has been said that 80% of enquiries have been made by companies anxious to determine their competitors'

activities. Even if data banks are intended for restricted use how long can it be in these days of open government in the Community and in the USA before pressure to make all data freely available becomes irresistable?

The answers to these questions are not scientific in nature. They can only be truly determined by awaiting the outcome of events. However, I believe that they should be more fully discussed and that it is the duty of the scientist to point them out to society. If we are to surmount the difficulties as they arise I suggest that it would be as well to establish even closer links and liaison between the data bank user, the information scientist and the administrator and legislator than we have hitherto enjoyed. This meeting may well be a step in this direction.

Finally, I will outline very briefly three data banks in the area of food science and technology, referring in passing to the arguments I have previously mentioned. Some thirty years ago workers in nutrition at Cambridge perceived the need to accumulate data on the nutritional properties of foodstuffs available and commonly consumed in the United Kingdom. The outcome of their work was the well-known publication of McCance and Widdowson 'The Composition of Foods'. Eventually the costs of producing the revisions of the data became too high for the funds then available, and an appeal was made to the UK Government for help. Today data on about one thousand individual foodstuffs is collected from a variety of sources including the food industry and much data is established by fresh experiment or subjected to further checking before being stored by the Ministry of Agriculture, Fisheries and Food's computers. The interpretation of this data is a joint matter for the Dunn Nutritional Laboratory at Cambridge and the Ministry's own nutrition experts before it is eventually published in the form of a book freely available. It should be noted that the need is sufficient to justify publication, but on-line computer facility is not necessary, and that the subject matter is analysed and validated as thoroughly as possible before publication. Commercial details are not published.

Another data bank with which I am currently concerned is one which will contain analytical data on contaminants in food consumed in the UK. Examination of the hundreds of thousands of analyses carried out in the UK on heavy metals in food reveal that over forty items of information can refer to one particular analytical result, but that the likelihood of each item being the focal point for data recall varies considerably. For example whereas we may want all the results for one year from one part of the country, it is unlikely that we should want to recall by date of analysis; date of production or purchase of the food is more useful. Accordingly some data items will not be stored on computer but will be retrievable by reference to the original analyst's report. The computer will be limited to 32K of core store - a mini computer - but a 5Mbyte disc will be employed. Retrieval programs will be written on demand according to the needs of the committees who will discuss and use the data. Data will be sifted by the technical secretary of the committees who will use his expertise to reject dubious material. The need for this data bank became apparent from the impossible task presented to the

committees and their secretary by the number and range of analytical results available to them.

Finally, I would refer to the work at COST on preparing the way for a data bank on the physical properties of foodstuffs. In supporting the proposal for this data bank on behalf of the UK I consulted over thirty organisations, academic and industrial. All supported the need for this data bank, albeit for a variety of reasons. Many research workers saw the data bank as an opportunity to review the present state of knowledge and to identify gaps whereas others, including those from industry, saw the need as arising from the desire to acquire more information about the nature of the material with which they work, particularly in relation to the design of sophisticated machinery now employed in modern food processing. This is an area where validation is as important as always, but where a high degree of precision or accuracy is not at present required or even possible. Although it is proposed that the Community's computer facilities be utilised, I may question whether this is justified by the complexity of the data or its volume. Certainly I feel that this is one case where a book is better than service from a bureau. However, I stress that the best way to establish this data bank with any surety of success is, as ever, to consult the users and establish their needs, and to maintain continuous and close contact between all concerned.

Animal disease data banks

G. Davies, Central Veterinary Laboratory, Ministry of Agriculture, Fisheries and Food, Weybridge, United Kingdom

Summary

At the present time there is little or no readily available information on the prevalence and distribution of disease in farm animals within the European Community. This information must be developed so that we can 1) control the spread of disease arising out of animal movements 2) measure the cost of disease to the Community and to the individual producers and 3) provide data for research and development workers.

The A04 contract report found that a great deal of information was available either in papers published by research workers or in unpublished records held by Government Departments, marketing organizations and research institutes. It proposed bringing these together in a regular publication. Because of the variety of survey methods, this data is not suitable for computer storage.

The British Veterinary Service has developed animal disease data banks that include: a) a computerized record monitoring the Brucellosis eradication scheme; b) a computerized record of all diagnoses made at 33 Government laboratories - this uses a simplified disease classification system (VIDA); c) a record (to be computerized) of all Salmonellae isolated from animals in Great Britain; d) a feasibility study of a rapid computer-based information system for use in controlling epidemics of disease such as Foot and Mouth disease.

The problems associated with the classification of disease, the identification of herds or farms, and the rapid retrieval of information from a computer file are discussed. The need to base disease information on properly collected animal population data is stressed.

Within the European Community there are 80 million cattle, 70 million pigs, 43 million sheep and 600 million poultry. These animal populations represent one of the most important of our natural resources and yet we have little immediately available factual information on the prevalence and distribution of the diseases current within them.

In this contribution I will outline the animal disease data that is available within the Community, describe the data banks that are being developed in Great Britain and speculate on future Community

developments in the field of data banks and computer based information.

But first of all, why do we want information on animal disease? In the early years of the 19th century, the Napoleonic wars spread contagious bovine pleuropneumonia throughout Europe and yet it was eradicated even before the causative organism was discovered. It was contained by controlling the movement of animals, but this was done at a time when most herds were very small and when most animal movement was to a local market or slaughterhouse. Now we are in a common agricultural market and for it to operate there must be freedom of movement for its products. Thus animals can be moved vast distances for breeding or for slaughter, and as the animals move so do the diseases. To control this trade, we must have information on the diseases to be found in the various areas of the Community and it must be readily available ant not buried away in the archives of Universities or government departments. This is the first and most pressing reason for developing data banks.

Secondly, we are working in a highly sophisticated agricultural industry. It has been said that the average farmer is on the treadmill of technological advance. Those who produce at low cost succeed, and those who do not, fail. Disease is a cost but so is its prevention and there may be quite small differences between the two. We must therefore be able to measure these differences not only on a community or national scale but also at a producer level.

Thirdly, we have developed extensive research and development facilities that, in the current economic and political climate, are having to justify their existence. Increasingly the research worker is asking for factual information on the prevalence and distribution of diseases so that he can decide if a disease is ecomomically important and worth working on, so that he can evaluate the possible predisposing factors, and so that he can estimate the effect of a new treatment.

It is with this background that there has been an increasing demand for data on the prevalence and distribution of disease. I am always surprised that although we have a wide range of scientific journals and computer-based index systems largely devoted to information on scientific techniques and laboratory experiments, we often have to recourse to brief official reports, unpublished documents and even oral opinions if we are searching for information on disease prevalence.

At the moment two international organisations publish data on the prevalence of disease. The Food and Agriculture Organization (FAO) publishes an Animal Health Year Book which records the occurrence, but not the prevalence, of disease in various countries. The Office International des Epizooties provides information on the prevalence of the most important diseases in its bulletins and circulars, but it is dependent on individual experts and government organisations.

There is no European Community organisation responsible for gathering this kind of information and even at a national level there are very few units devoted specifically to studying the epidemiology and economics of disease in animals. There are a variety of Government organisations, University clinics, marketing boards,

insurance agencies and diagnostic laboratories that collect information that has a bearing on animal disease but this has not been collated into a simple accessible body of data. All European countries have systems for counting animal populations and also for the compulsory notification of certain diseases although the list of these 'notifiable diseases' is small and varies from country to country.

The European Commission has made an initial move towards developing this kind of information by commissioning contract A04. This was designed to examine the data that was already available and that could be collated without setting up new information systems. The report of the contractor shows that over a 5 year period, no less than 250 articles have been publisehd that provide information on the prevalence of diseases in cattle in France, the United Kingdom and Ireland. This volume of published work is rapidly increasing and it clearly forms the most substantial bank of information that we have. The report goes on to examine suitable ways of compiling a record of disease prevalence based on this information. It points out that anyone searching for information of this kind requires: a) a clear definition of the population studied; b) a description of the sampling methods used; c) a definition of the diagnostic criteria. But it finds that most of the published work gives little in the way of definition of the population and that there is great variety in the sampling methods and diagnostic criteria. The report examines the prospect for creating a computer record of this considerable body of information and it comes to the conclusion that the diversity of enquiry makes such a record impossible. Clearly we have to wait until there is some standardization of survey methods.

The report does, however, describe alternative ways of condensing the information contained in the articles and it proposes that the summaries should be published as a 6 monthly booklet. This booklet would also bring together the vast store of unpublished information to be found in the archives and reports of official organisations within the Community. I firmly believe that the provision of this service would be a major step forward in disease surveillance within the Community.

If we look beyond these beginnings, we must refer to developments at the national level and I am going to confine myself to the current development in animal disease data banks in the U.K.

First of all we have the Census of Animal Population. This is based on the annual agricultural census. Mr Peter Ellis, who has recently completed a study of the epidemiology and economics of swine fever eradication in the Community, has commented on the difficulty he has had in obtaining data on the pig populations in various areas. Any worthwhile epidemiological enquiry must be based on reliable demographic data if only because it allows us to draw statistically valid samples for surveys and other enquiries.

Data from disease control schemes

In the U.K. we are entering the final stages of a brucellosis eradication scheme and for the last five years the records have been

computer-based. This has been an aid to administrators monitoring the progress and the cost of the scheme and we have been able to use the computer record to measure the efficiency of the various diagnostic tests. We have learnt two lessons from our experience with this scheme. First of all the input to the computer file must be simple so that it does not create unnecessary work in busy animal health offices. The computer record should be confined to an index that lists farms or herds and that describes only the main parameters e.g. completion of tests, breakdowns of infection and so on. Any attempt to put all the laboratory data on file will only break the system. Detailed enquiry into the efficiency of diagnostic tests can be best done by using the computer record as an index to draw samples of relevant herds that can be followed up from the files held at local offices.

The second lesson learnt is that this kind of extensive record must be based on a Data-base Management System so that the file can be searched without protracted programming.

One advantage of a compulsory scheme is that the herds or farms can be identified by number. The ownership of herds can change repeatedly and a system that records individual cows would be unwieldy. We thus have to record herds or farms and in Great Britain we use a County/Parish/Holding number (C.P.H.) which works reasonably well. An alternative is a map reference and this has advantages for computerized mapping of disease outbreaks or animal populations. This identification of herds is a problem with the next kind of record.

The diagnostic laboratory record

In Great Britain 33 government diagnostic laboratories contribute to a national record of disease in farm animals called V.I.D.A.II. it is based on a simple list of 380 diagnoses and it has been in operation since January 1975. The record is widely used by research workers and others and again it is mainly used as an index, e.g. the records provides a list of the diagnoses of muscle disease or copper poisoning made in the country during a certain period and the research worker can then refer to the individual laboratories for the detailed case records. A sample of the annual record of diagnoses is shown in Appendix II.

The original V.I.D.A. system used the SNVDO disease classification. We found this to be too complex for recording diagnoses from a large number of laboratories and it created difficulties when used as an index in that the diagnoses were not uniquely identified.

Records of Salmonella infections

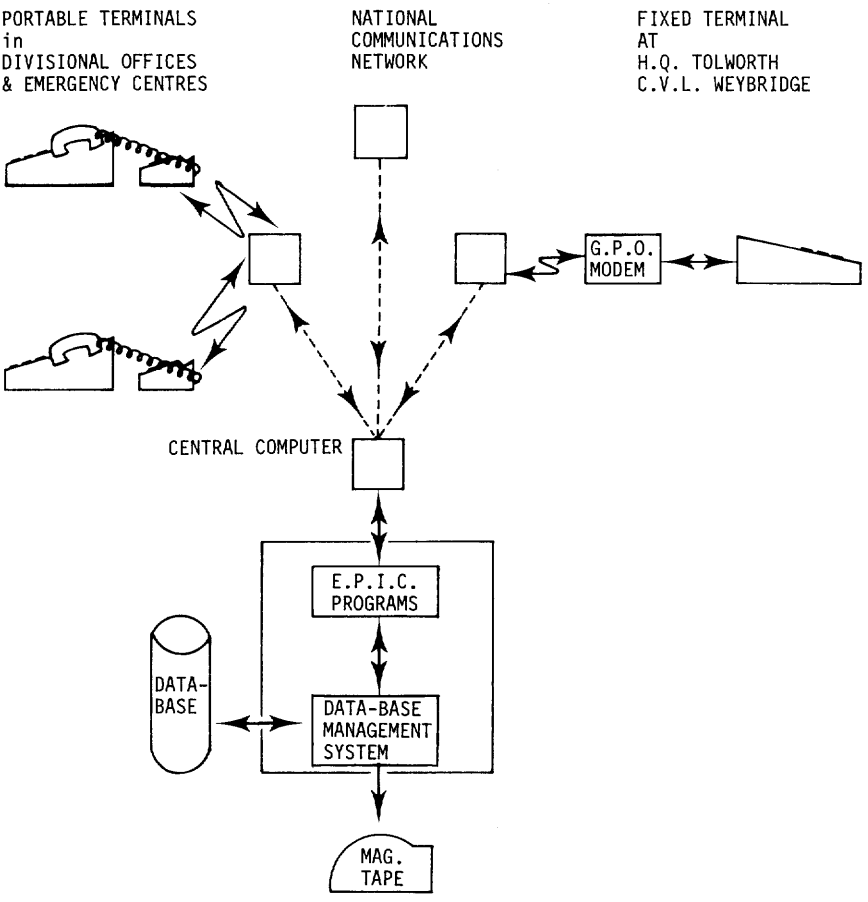
We have been operating a record of Salmonella infections in Great Britain since July 1975. It is not yet computerized because we have not finally decided on the format of the information and we are looking for a suitable data-base-management system. At the moment we include the address of the infected premises but not the herd number and this is an obstacle to any subsequent computer storage or computer mapping. We also have problems in finding a Salmonella 'incident' which is a term we use to describe a new infection on a certain farm. It differs from a Salmonella isolation because the number of isolations clearly depends on the amount of

investigation carried out by the local laboratory and they may not reflect the amount of infection or disease.

Finally, we have E.P.I.C. (Epidemic Planning and Information, Computer assisted). The United States Department of Agriculture commissioned a study of systems for dealing with sudden outbreaks of major epidemic disease such as foot and mouth disease. One of the parts of the study, carried out at the University of Minnesota by Mr W.M. Miller of the Epidemiology Unit, Weybridge, was a computer based system for providing day to day information on the progress of an outbreak. It depends on portable computer terminals connected to telephones at local offices, and these feed in to a data base information that has been gathered by veterinarians visiting the infected farms and markets. The data base can be interrogated through terminals at the national or regional headquarters, and thus it enables senior staff to direct efficiently the resources of man and materials that are available (Appendix I); This system is currently under evaluation by the British Veterinary Service.

I have described some of the present developments at national level. What is the way ahead for the Community as a whole? Or to put it another way, what are the priorities? My experience has been that we must build up a simple information before we attempt sophisticated data collection schemes. Therefore I see the information syetem suggested by the A04 report as an important first step. Then we must look at areas where political or public health needs are most pressing and we should consider a data bank that records outbreaks of notifiable diseases within the Community. In addition a number of countries have systems for recording zoonoses such as Salmonella infections and these could be brought together in a Community data bank. Finally, we have the diseases that are neither important enough to be 'notifiable' nor of serious public health concern. These diseases do, however, represent a considerable cost to agricultural production and I would suggest that the first step to recording their prevalence and distribution is to develop a European record of diagnoses derived from diagnostic laboratories in the various Community countries. The essential basis for this is agreement on the classification of disease and I belive that there is now sufficient shared interest and experience in the European diagnostic laboratories for this to be a working reality within the next decade.

Appendix I. Communications in EPIC - using commercial bureau facilities.



Group 1 Systematic diseases and those not readily classified organically

060 Hereditary and developmental anomalies
087 Congenital trembling

515 Iron deficiency anaemia
545 Malnutrition not otherwise specified

113 *Clostridium oedematiens* disease
117 *Clostridium Welchii* type C disease
110 Clostridial disease not otherwise specified
123 Colisepticaemia
122 Colibacillosis - oedema disease
120 Colibacillosis not otherwise specified (see below for enteric colibacillosis)
133 Erysipelas (other than foetopathy)
139 Leptospirosis
140 Listeriosis (other than foetopathy)
152 Pasteurellosis (to include pneumonia due to pasteurella)
161 Salmonellosis due to *S. cholerae-suis*
165 Salmonellosis due to *S. typhimurium*
160 Salmonellosis due to other serotypes
134 *Spaerophorus necrophorus* disease
171 Streptococcal disease (other than foetopathy and meningitis)
142 Tuberculosis

720 Neoplasms - not specified - any site

421 Poisoning due to arsanilic acid
420 Poisoning due to salt (to include water deprivation)
410 Poisoning due to chemical not specified
440 Poisoning due to plant not specified

Group 2 Diseases of the digestive system

121 Enteric colibacillosis
234 Epidemic diarrhoea
662 Gastric ulceration
491 Hepatosis dietetica
667 Intestinal abenomatosis

668 Jones type intestinal haemorrhage syndrome

	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec	Total 1976	Total 1977
060	3	4	4	5	1	3	1	2	6	1	1	4	35	23
087	1	1	3	-	4	1	-	1	2	-	-	-	13	11
515	-	7	5	1	2	2	4	1	1	-	1	3	27	31
545	8	19	11	12	5	10	7	5	10	9	5	19	120	115
113	1	1	7	8	3	3	2	-	-	3	2	1	31	42
117	-	2	2	1	4	1	2	-	2	2	2	-	18	30
110	1	1	1	1	2	2	2	1	-	1	1	2	15	11
123	20	15	15	18	12	8	10	16	30	5	19	10	178	166
122	11	16	12	6	9	4	6	8	8	19	11	9	119	115
120	29	24	26	26	20	10	22	22	25	21	30	11	266	240
133	3	2	10	3	2	1	2	1	10	5	4	6	49	45
139	1	-	-	-	-	-	1	-	-	-	1	1	4	3
140	-	-	-	1	-	1	-	-	-	-	-	-	2	1
152	13	11	14	16	16	6	4	9	10	13	11	12	135	129
161	1	3	4	-	1	1	1	-	-	1	1	-	13	10
165	2	2	3	-	-	-	2	1	2	2	7	2	23	18
160	4	4	7	6	3	3	6	5	5	7	7	4	61	37
134	-	1	1	-	-	1	-	2	1	-	-	1	7	3
171	10	22	14	9	13	12	6	6	10	9	7	13	131	117
142	4	3	5	-	5	4	2	3	3	6	3	3	41	18
720	-	-	-	-	-	1	1	-	1	-	-	2	5	4
421	-	1	1	1	-	-	-	1	-	-	-	-	4	9
420	1	5	5	3	2	2	4	1	7	4	2	5	41	45
410	2	2	5	-	2	1	2	1	1	-	-	-	16	15
440	-	-	-	-	-	-	-	1	-	-	-	-	1	2
121	171	184	160	131	124	123	163	168	220	209	205	125	1983	1578
234	2	-	2	-	-	1	-	-	1	-	1	1	8	4
662	5	6	7	2	3	-	1	-	5	11	10	12	62	46
491	-	4	3	2	3	2	-	1	1	1	1	1	19	13
661	6	8	6	2	2	3	2	7	7	2	10	8	63	36
668	10	12	14	7	11	7	6	4	3	3	8	10	95	102

Feed data banks and the International Network of Feed Information Centres (INFIC)

H. Haendler, Dokumentationsstelle der Universität Hohenheim, Stuttgart, Federal Republic of Germany

Summary

For the field of feed science most important presuppositions for running a data bank are given: the need for such data and the possibility of their isolation. Thus in this field the establishment of centres for systematic collection of data and data processing was started relatively early.

The today's data bank of the Documentation Centre of the Hohenheim University has been developed from an institution founded in 1949. Works started in the USA in 1963 led to the today's 'International Feedstuffs Institute' in Logan, Utah. Some institutions are dealing with collecting and making available data from - mostly own - analyses.

The combination of these Centres in the 'International Network of Feed Information Centres' (INFIC) in 1971 initiated extended works for the adaptation of the different systems. For the compatible recording and processing of data at different centres are available today: 1) an 'International Vocabulary for Describing Feeds' involving more than 5000 descriptors in three lingual versions (English, German, French), structured in six facets according to the semantic components of the denominations of feeds; 2) a coding system for 800 substances and values; 3) a system for the description of feed samples allowing to record possible influences on the contents; 4) a standardized data recording system; 5) computer software for different possibilities of data processing.

At present data processing is restricted to the Centres in Logan and Hohenheim. The Hohenheim data bank encloses more than 750 000 data units, it serves for preparing feed composition tables, different special outputs and answering questions.

Data documentation is a process in which 'informemes' are extracted from their original context in the document and put into a special form of their own, this being a data unit. These data units are the original documentation units which are processed, stored, selected and supplied. A data bank embraces all these facilities allowing storage, processing and supply of the units.

Purposeful data documentation in a specific field has two basic requirements: 1) There must be a real need for the information ('informemes') to be documented. 2) It must be possible to isolate the data to be processed. Both these requirements have been met for

many years for data in feed science.

Re 1): Since Albrecht Thaer demonstrated over 160 years ago that qualitative recording of the nutritional value of feed was a necessity for efficient agriculture, more and more detailed data on content and values of feedstuffs have been required, especially for the calculation of feed rations for high-production animals and correct utilization and make-up of existing feedstuffs.

Re 2): The possibility of isolating 'informemes' i.e. its information value out of context, is provided first and foremost by the constant structure on which the propositions to be processed are based. This always involves a (quantifiable) object/property relationship. Moreover, the standardization of laboratory methods and feed value data - thanks to the Weender analysis method, since considerably developed and improved - has gone a long way towards permitting data to be isolated from chemical and physiological studies of feedstuffs.

The fact that the two above requirements have been met has meant that data documentation in feed science has already been in operation for some time. The only other requirement would be to recognize that the need for reliable information can only be met by a systematic collection of as much relevant data as possible from specialist institutions and centres and by developing and introducing suitable methods of processing, storing and selecting these data. This recognition led to the establishment in 1949 of an 'Archive for feed value determination' in Kiel which paved the way for the current data bank for feed findings at the documentation department in the University of Hohenheim.

Certain other establishments were set up independently of this development in Germany but on the basis of the same situation in terms of needs and insight. In this context mention must be made of the particular effort made by the Utah State University in Logan (Utah, USA). Work begun in 1963 led to the creation of the International Feedstuffs Institute. In addition, efforts were made by various institutes in other countries to build up similar data collections or to put the large amount of data material available in certain research centres to systematic use for general purposes. A particular case in this respect is the Institut d'Elevage et de Médicine Vétérinaire des Pays Tropicaux (IEMVT) in Maisons-Alfort (France).

Some years ago the worldwide need for information on feedstuffs, and especially the lack of information in this field in developing countries, induced the FAO to take up contact with information centres, to run a survey on current activities and finally, in 1971, to arrange a first meeting of the heads of these centres. These initiatives led to the foundation of the International Network of Feed Information Centres (INFIC). The aim of this network is a systematic recording of all information found everywhere on foodstuffs and its availability worldwide by means of data exchange and the operation of a joint data bank with branches at various locations. Computer processing was limited first of all to the above mentioned institutions in Logan and Hohenheim which met the technical requirements (use of data processing facilities) and

acted as computer processing centres for INFIC.

The biggest problem the INFIC-partners encountered was the incompatibility of the two systems in Logan and Hohenheim. Several years' work in Hohenheim went into adapting the systems in conjunction with the institutes in the USA and France. The first task was to develop a new international system, including all the respective part systems and auxiliaries, and to adjust existing systems to this one. Although there was something to be built on, the fact that certain elements had to be used, also proved to be a handicap. Some time ago the bases for compatible data recording and processing were established. This system is already in use even though greater clarification is required in some respects and the reorganization of the extensive data stock in Hohenheim is still not complete.

The first part system created for compatible data recording and processing, the description system. Extremely high demands have to be made of the identification of informemes in data documentation. A precise description of data contents calls for adaptation to the structure in question and must take into account the respective constituents.

Accordingly, the description system itself is divided into three parts: 1) the description system for feedstuffs; 2) the coding system for contents and values; 3) the system for sample description. In addition to these, there is: 4) a system for standardized data recording; 5) the computer software required for processing as a technical system.

Re 1): To provide an accurate description of feedstuffs as the objects under examination and to identify them in terms of selection, a faceted vocabulary was devised. The six facets correspond to the description components which together represent the term for the feedstuff in question: the original material, the parts of this material, the process to which the material was subjected, the vegetation or development stage, the cutting time and the particular degree of quality. In all the vocabulary contains more than 5000 descriptions, of which more than 4300 just for the'original material' facet. Each description is found in three languages: German, English and French. A Spanish version will also be prepared. All first descriptions of a feedstuff are recorded under a reference number in a file. Automatic translation of descriptions and selective printout from the data bank in one of the system languages are possible.

Re 2): The extension of the data bank to include newer and newer properties called for corresponding codes. There are now codes for 800 different contents and values. These codes indicate quantitative values which are to be stored as the actual analysis results. In addition to the classical nutritional and active ingredients, for example notations for toxic components and contaminated substances are also provided for. There are also notations for derived values which are not determined directly but calculated from the stored values.

Re 3): The third description dimension goes beyond the original and general object/property relationship and relates to possible

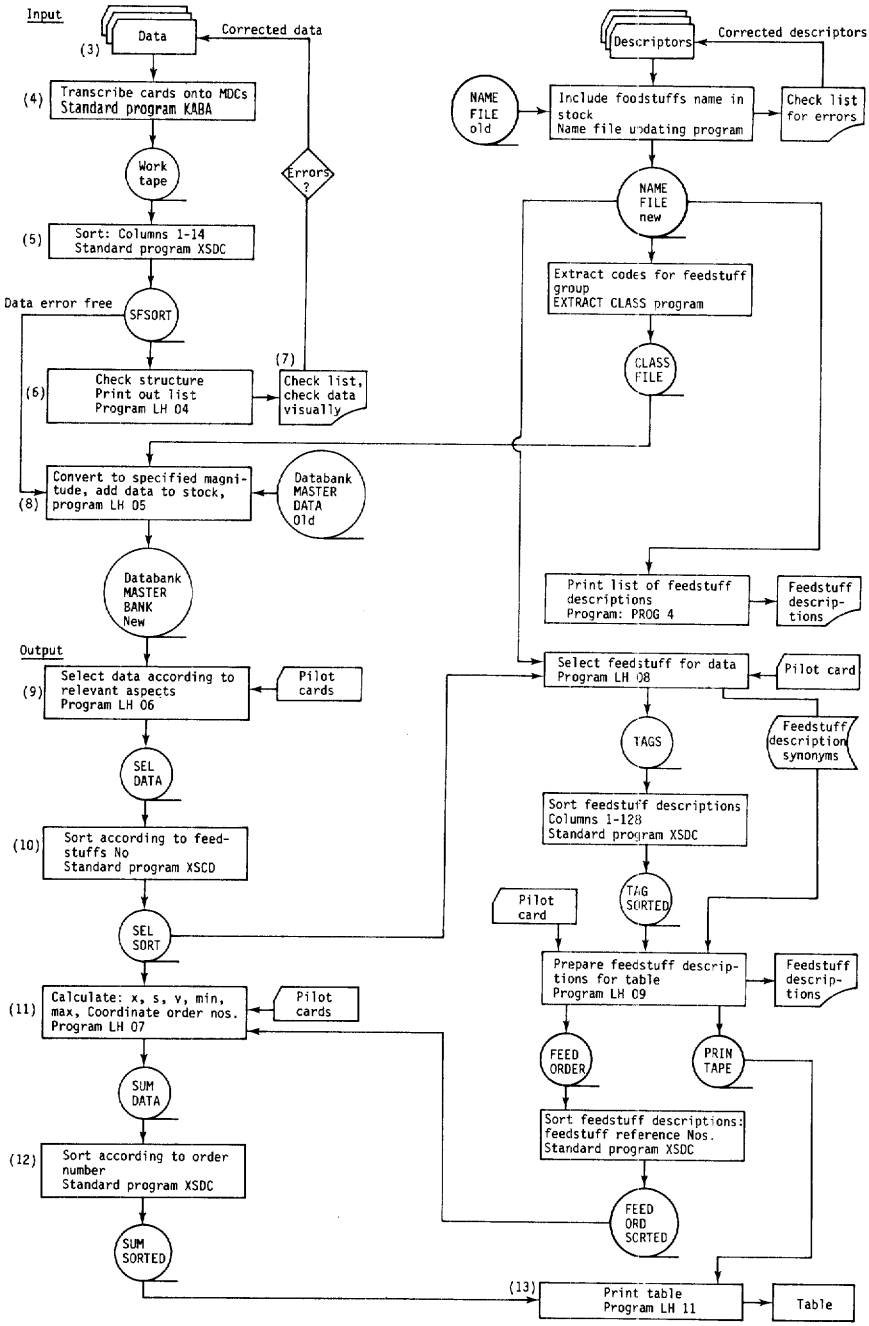
effects on these relationships by extreme factors (environmental). Consideration of all factors which have affected the particular individual feedstuff sample permits differentiation between typical and atypical samples and the assessment of such effects (e.g. crop protection measures, storage conditions, industrial or traffic emissions).

Re 4): All data must be fed into the system in accordance with a specific format. The details are laid down in a comprehensive data recording system. In this case various formal data (e.g. country of origin, laboratory, bibliographical details) are to be considered in addition to description. Data recording forms are used for practical work which facilitate correct entry in terms of format and punching onto punch cards.

Re 5): Special programmes are available for the processing of data - input, checking, collating and calculating average values and deduced values (e.g. on digestible nutrients, energy value data) - and for the printout of data.

A typical operation for data processing and the production of feedstuff tables is shown in the 'flow chart diagram on data processing' in the Annex. The data units stored at Hohenheim - currently over 750 000 - is used to compile various feed composition tables for special types of animal, for special tables or for individual enquiries. The development and improvement of description systems allows the increasing demands on the selection of individual data to be met. Exact descriptions of the set of selection descriptions and of kinds of calculation are feasible for each case, meaning not only that special information can be provided for the content and values of certain feedstuffs but also that data in accordance with origin and particular effect parameters and the like can be selected or that new values can be derived from existing ones.

Appendix. Flow chart diagram on data processing.



Notes to flow chart on data processing (processing operation)

Input

(1) Data recording on form (in accordance with recording system).
(2) Data control on form, possibly checking with the original document or with the data producer.
(3) Punching the data onto punch cards.
(4) Transfer of data from punch card to magnetic tape (work tape). Standard program KABA.
(5) Sorting (by form number and type of card), production of SFSORT tape. Standard program XSDC.
(6) Design check (after coding) and printout of a list with an error flag. Program LH 04.
(7) Visual data control of list (comparison with form) and, if necessary, production of correction punch cards and processing from stage (4).
(8) Input of SFSORT tape data (corrected) into the data bank with coordination of the codes for feedstuff groups available on the separate CLASSFILE tape (MASTERDATA tape). Conversion of values to uniform reference units and magnitudes.

Output

(9) Data selection in accordance with specific aspects (type of table, user requirements) and production of SELDATA tape. Program LH 06.
(10) Sorting (by feedstuff number), production of SELSORT tape. Standard program XSDC.
(11) Calculation of average values, deduced values (e.g. energy values), standard deviations, variation coefficients, minimum and maximum values plus coordination of order numbers for feedstuff designations (from separately produced FEEDORD-SORTED tape), production of SUMDATA tape. Program LH 07.
(12) Sorting (by order number), production of SUMSORTED tape. Standard program XSDC.
(13) Coordination of feedstuff descriptions (from separately produced PRINTAPE tape) with the SUMSORTED tape and printout of the respective table. Program LH 11.

Automation and the use of data banks in milk recording, artificial insemination (a.i.) and registration of cattle in the Netherlands

J.H. Verheijen, Common Data Processing Centre for Cattle Husbandry, Arnhem, the Netherlands

Summary

In the Netherlands data processing and storage in the area of milk recording, artificial insemination (a.i.) and registration is concentrated mainly in the computer of the GIR (foundation common data processing centre for cattle husbandry). In the foundation the central milk recording service, the a.i. associations and the Netherlands cattle herdbook participate. Milk recording is effected for nearly 1.5 million cows on 36 000 farms every three weeks, kgs milk and fat and protein content of all cows are measured. In the computer file a record of each cow is kept. The record contains complete data of the cow including descent, all test data of the current lactation, all previous lactation totals and some standardized figures. From these data the transmission of bulls is derived. Data of nearly 2 million artificial inseminations a year are also processed and stored. Monthly statistics with nonreturn rates are produced and invoicing on behalf of the associations is effected. Recently the input of officially registrated calves in the computer started. In the near future a completely integrated system for milk recording, a.i. and registration will be developed. One of the objectives will be the automatic registration of calves making use of the data of the mother and the data of a.i. or natural service.

Introduction

In the Netherlands automation of working-up milk recording data started in 1966. The 'Stichting Centrale Melkcontrole Dienst, CMD' ('Foundation Central Milk Recording Service') then centralized processing of milk recording data by means of a computer procured for that purpose. So far all work had to be done manually in the local milk recording associations provincially supervised by the 'Provinciale Melkcontrole Diensten' ('Provincial Milk Recording Services').

The computer was a 64K words tape-oriented EL X 8 of Electrologica. The input for the central computer-system housed at Arnhem was prepared by the 11 provincial services. The recordings daily received from the associations were transferred to paper tape which was daily sent to Arnhem where they were worked up by the computer. At the end of 1970 the switch-over to the mechanical administration

was completed.

The transfer of the technical and financial administration on behalf of a number of a.i. associations to the computer at Arnhem started in 1970. The input on paper tape was also prepared by the provincial services. At the end of 1973 milk recording for over 1.3 million cows, making 60% of the total number of cows, and a great deal of the administration for a.i. associations were on the computer. In that year the growing demand for quicker and better information, the experienced great possibilities to fulfil this demand and in addition the interest shown by the breeders-associations and especially 'The Royal Netherlands Cattle Herdbook Society' to make use of the computer for their registration tasks, led to the establishment of the 'Stichting Gemeenschappelijke Informatieverwerking voor de Rundveehouderij (GIR)', (Foundation Common Data Processing Centre for Cattle Husbandry). In the new foundation a.o. the CMD, a.i. associations and the Herdbook participated.

The main task of the GIR is to promote and to accomplish further concentration and integration of administrative procedures and data of the three participating dairy cattle breed improvement organisations.

In 1974 GIR took over the staff of the computer department of the CMD and the computer that shortly before had been replaced by a 128 K NCR Century 200 system with CRAM (Card Random Access Memory), discs and tapes.

In the near future some components of the system will be replaced. The CPU will be a 256 K Criterion 8570 and CRAM will be followed up by discs of 200 K bytes capacity.

Milk recording

At this moment milk recording is effected for nearly 1.5 million cows on about 36 000 farms.

The basis of milk recording is in the local breeders- and milk recording-associations. At present there are 350 of them. The farmers are visited every three weeks by an independent milk recorder. The test-day yield per cow is measured and a sample of the milk is taken. The samples are examined on fat-content and protein-content in 7 regional laboratories.

The results per cow of the test, that is to say day yield as well as the number of the sample, are entered on a form delivered by the computer ('turn around' principle). After inspection by the local administrator the forms are sent daily to the provincial services where they are transferred to magnetic tape by means of data entry systems ('key-to-disc').

This magnetic tape as well as the tape containing the results of the laboratory examinations for fat and protein is delivered to Arnhem as input for the computer. In the milk recording file on CRAM a record of each cow is kept. The record contains complete data of the cow, including descent, all historical lactation data and all test results of the current lactation. Updating occurs on a daily basis.

Every three weeks the farmer receives the results of the recent test of all his cows. Per cow the total yield in the current

lactation is given in absolute figures as well as in standardized figures. (By the standardization influences of environment, age, season of calving and stage of lactation are eliminated so that the cows can be compared one with another). Moreover totals of the test and standardized figures for the farm as a whole are produced.

On the test report the cows with a standardized production of over 10% above herd average are indicated with + and those over 10% below herd average with -.

At this moment several new indices are being developed in close cooperation with the University of Wageningen. The object of these indices is to make cows better comparable, which is of great importance from a selection point of view.

When lactation of a cow has come to an end, a complete record of the cow is delivered to the farmer, the association, the provincial service and the Herdbook. It gives all data of the cow, including descent, historical lactation data and a complete specification of the latest lactation.

Many statistics are periodically delivered to the farmer and his organisations. As of most cows the descent is officially known the results of milk recording are also used for the calculation of the production transmission of the bulls. At this moment the method of contemporary comparisons is used. In the near future a new method will be introduced. It is now being developed ('Best Linear Unbiased Prediction').

Artificial insemination (a.i.)

All inseminations by the a.i. associations are recorded for financial and technical purposes. Every month invoices are prepared and statistics are delivered, the latter giving a.o. 56 days non-return rates per bull and per technician. Insemination data are periodically delivered to the registrating authorities for the verification of the descent of the calves born.

Registration

In the course of this year the input of calves into the computer files started. The calves officially registered by the local breeders associations and those entered in the Herdbook are filed at Arnhem. Actually the control of the data of the birth declarations is still being done manually by the Herdbook and the associations.

The main purpose of the input of the calves is to create a file of officially registered calves with ascertained descent which, together with the animals already in the milk recording file, will result in a stock of living animals on which in the years to come a fully 'integrated' system can be built.

Integration

This completely integrated system for milk recording, registration and a.i. will be developed in the next few years. An important objective will be the 'automatic' registration of calves making use of the data of the mother of the calves and the a.i. data, in a

given case also the data of natural service, already in the files. It is intended to start automatic registration on a smaller scale in the course of next year.

Personal data for a cow beginning her career as a 'milk recording' cow so far had to be entered specially in the milk recording file. As a result of the integrated system these data can be taken from the registration file. This will come into operation next year.

It is being studied whether also the 'Public Health Service for Animals' with which there are already close connections will participate in the integrated system.

In the near future the integrated system will result in a complete data bank in which the most relevant information for the organized cattle breeding in the Netherlands is stored and made accessible.

Agrodata data bank

J.-L.Rastoin, Mediterranean Agronomique Institute of Montpellier, France

Summary

The purpose of AGRODATA is to make available to research workers and others employed in the agricultural food production industry information of an economic nature on the 100 leading concerns in this sector in the world. For this purpose, starting in 1971, an index has been built up at IAM comprising identification data and indicators of size for each of the 100 firms concerned. The identification comprises the name, address, structure and activities, together with details of multinational involvements. The indicators of size consist of accounting details (balance sheet and simplified trading account) recorded since 1963 and updated annually.

In data-processing terms, the bank is at present available in the form of punched cards and operates by batch processing. Apart from printout of the data in tabular form, the program (approximately 300 instructions in FORTRAN IV) provides for the calculation of a large number of coefficients and ratios and of hierarchical classifications. An on-line conversational version is currently being developed on the IBM-CALL network.

Aims: Information base on agricultural foodstuff multinationals

Agrodata, a data bank on the hundred leading companies in agricultural foodstuffs, was set up at IAM in 1971. The aim of this bank is to make available to users, especially universities, professional and government research workers and management staff concerned from different countries, basic data for the purpose of analyzing and assessing the leading firms in the world agricultural and processing industry.

Content: Approximately 20 000 technical and economic indicators

The Agrodata primary file is composed of two parts (see appendix I):
- The first part relates to indicators which permit company identification: name, address, organization, sectors of activity, branches in foreign countries. These indicators are recorded for the most recent year.
- The second part concerns financial and economic data consisting at present of 11 items: simplified balance sheet (assets: fixed assets, floating assets; liabilities: own capital, long-term debts,

short-term debts), operating data (turnover, net profit, depreciation, cash flow), number of employees. Information is stored for the years 1963 and 1974 and is updated annually on the basis of published company trading results. The number of indicators can be increased.

Management: 4 data processing sub-routines

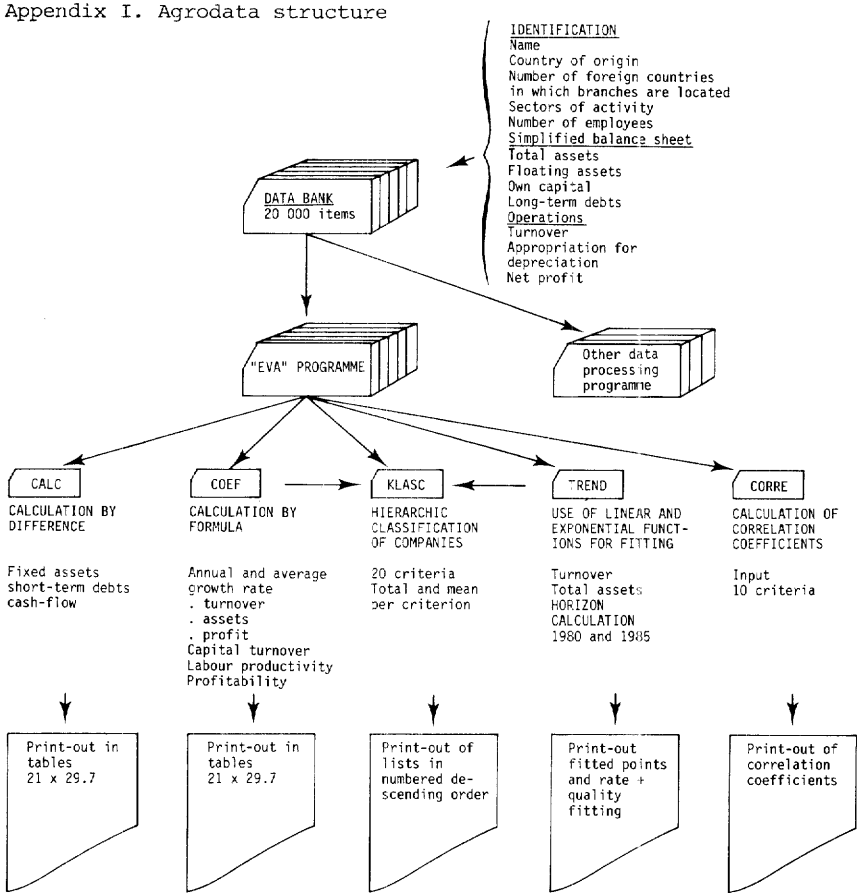
In the initial form for batch operation, the management program of the Agrodata file is written in FORTRAN IV IBM. It consists of approximately 300 instructions permitting four types of operations (see Appendix I):
- Printout of statistics on companies in 21 x 29.7 cm format (2 tables per company);
- Calculation by difference or aggregation of balance sheet and operating indicators, calculation of economic and financial ratios from basic data (productivity of labour and capital, profitability, growth rate);
- Hierarchic classification of companies in descending order on the basis of each of the 11 basic indicators;
- Trend assessment by fitting the turnover and assets on the basis of linear and exponential functions.

The updating and operation of the batch version gives rise to the publication of a yearbook.

Prospects: shared-time management and teleprocessing, extension of Agrodata

The French time-sharing IBM CALL network is currently developing a version of Agrodata. The aim is to insert the whole data base in the network so as to permit fragmentary interrogation and individual calculations from users. To date, the company 'identification' part, which was not computerized in the batch version, is already available in the IAM library of the IBM CALL network. The subfile is managed by the 'A.S.' package, which permits a large number of operations (print-out, read-out, statistical calculations, models). In addition, an extension of Agrodata is being considered. It would cover the 100 leading world cooperatives, which would permit the use of a supplementary data bank facilitating comparative analysis between capitalist and cooperative companies. This extension should take place in cooperation with European and world organizations.

Appendix I. Agrodata structure



Data banks on crop losses

F.H. Rijsdijk and J.C. Zadoks, Laboratory of Phytopathology, Agricultural University, the Netherlands

Summary

The third of a series of Agro-ecological Atlases on Cereal Growing in Europe, published under the auspices of the European Cereal Atlas Foundation, will deal with pests and diseases of cereals. Information was collected from an international enquiry and extracted from the literature.

The information was stored in a data bank before processing. During processing the data are weighed with help of certain rules of procedure expressed in simple algorithms.

Data processing leads to two kinds of output:
- Loss maps of Europe show the mean annual loss due to each cereal pest and disease by means of isoloss curves.
- Risk maps of Europe for selected pests and diseases show the relative risk incurred by the cereal crop in each climate area as determined by the climate of that area, areas with equal risk being separated by isorisk curves.

New developments in detailed disease survey combined with a computer based warning system for some cereal diseases in the Netherlands can lead to a much more detailed data base, not only for crop losses in general but specified for cereal varieties, physiologic races of diseases, interaction with fertilization and application of fungicides.

In the area of crop protection, there are a number of data banks, some functional, others in the making. A feasibility study of a data bank on pesticides is reported by Scholten (1976).

Several chemical industries have stored their data on field experiments, including yields, sequential disease assessments, and treatments, in data banks; these data banks are not, or not yet, accessible to the general public, though they contain much valuable information for studies, in epidemiology and crop losses, for example. The Pennsylvania State University in co-operation with the state authorities of the Commonwealth of Pennsylvania, USA, prepared a data bank for the purposes of diagnostics, registration, statistics, and detection. Though seemingly simple in design this undertaking is impressive in its implementation, as retrieval of data is possible by way of many different entrances. The data bank will ensure rapid dissemination of up-to-date information. The fourth type of data bank, the one to be discussed in more detail here, is

a small undertaking organized for the purpose of assembling information on crop losses of cereals in Europe. Two characteristics are conspicuous: it is a data bank designed for a single purpose, and the data are of the most divergent quality.

The data bank on crop losses in cereals is part of the operations of ECAF, the European Cereal Atlas Foundation, which has its seat in Wageningen, the Netherlands. Under the auspices of ECAF atlases on cereal growing in Europe are published. Two atlases have already been produced, one on agrometeorology (Thran & Broekhuizen, 1965), in which Europe is divided into some 80 agro-climatic sub-areas, and one on cereal growing in general (Broekhuizen, 1969). A third atlas, dealing with pests and diseases, is in preparation at the Laboratory of Phytopathology of the Agricultural University at Wageningen. Among other things, this third atlas will contain data on mean annual losses of the various cereals due to pests and diseases. It will also contain information on weather-dependent risks and diseases.

Crop loss data have been obtained in three different ways:
- Enquiry: a questionnaire was sent out and completed by some 200 colleagues all over Europe, with questions on cereal crops, crop losses per harmful agent, and other relevant data.
- Literature review: The Centre for Agricultural Publishing and Documentation (Pudoc) in Wageningen started a search of the European literature to find published information on crop losses, with special emphasis on eastern Europe.
- Personal information: Data collected by a few informants who had travelled extensively to that purpose.

All data were coded in one format and stored on punched cards and magnetic tape. Every card contained one elementary *unit of information*, characterized by crop, harmful agent, and geographic position. The total number of cards amounted to some 25 000. The maximum number of entries per card was fourteen, of which three pertained to loss data; most of the others were qualifiers. As the authors depended on the benevolence of the informants, they had to live with the unfortunate situation that most cards were incomplete. To indicate the data handling problem, it suffices to say that the loss data collected were not 'hard' crop loss data acquired in accordance to precise instructions, but rather *opinions* of informants varying from wild guesses to conclusions drawn from experimental evidence. The obvious uncertainties with respect to the *quality* of these 'soft' data led to the conclusion that any data handling must be subjective, taking into account and judging the quality of every unit of information. Objections against subjective data handling can be met by chanelling subjective judgment through certain *rules of procedure*. Rules of procedure are discussable, they can be altered, and when agreed upon they can be applied without respect of persons. The rules of procedure were transformed into simple algorithms, and these were entered into tha data handling program. The quality of the data was taken care of by means of weighing procedures:
- Age of data: the older the data the lower their weight.
- Size of area: Data from areas comparable in size to the climate

sub-areas were given a high weight; the weight decreased the more the area referred to deviated in size from that of the climate sub-areas.

Part of the information was given in the form of crop loss estimates, part as disease severity or disease prevalence data, and but few informants gave both types of data. Therefore, a rule of procedure was devised to relate losses to severities and/or prevalences. For most harmful agents enough data were available to calculate a relation of loss to severity (Fig. 4). When all the algorithms had been placed in order in the program, and all the cards had been read in by the computer, the interim result for every combination of host and harmful agent was a list of paired items, the number of the climate sub-area and the mean annual loss calculated for that sub-area (Fig. 2).

This typical data bank result, though modern in appearance, was not very useful in itself. So the next step was to produce one of the most antiquated forms of a data bank known, the geographic map. This was done again by means of a computer, which steered a plotter to position the loss data on a base map. The computer also calculated and indicated the position of the lines connecting points of equal losses, the *isoloss* curves. The final drawing of the isoloss curves is done by hand, as this is simpler, cheaper, and aesthetically more pleasing than map drawing by means of the computer. The final product of the data bank is an 'old-fashioned' map (Fig. 3). As said before, the map is based on opinions, and consequently, it represents nothing else than an opinion, the reasoned opinion of the authors. They and not the informants carry the responsibility, and the burden of proving the correctness of their opinion. They are to be blamed if their maps are incorrect. This is the consequence of what the authors call *subjective data handling*. The map so produced is primitive in the same way as the early geographic maps were primitive, these rendering the opinion of the geographer rather than objective geographic reality.

The Agro-ecological Atlas, Volume I, is a data bank in itself. The climatic data per climate sub-area have been used to simulate epidemics of yellow stripe rust (Puccinia striiformis) on wheat. As only monthly averages were available in the atlas, years with different weather types were simulated by varying around these monthly averages with the help of a random generator. The simulated yearly weather sequences were then used as inputs to a simulator of fungal epidemics, set for yellow rust (Rijsdijk, 1975). The result was a frequency distribution of terminal severities over 50 simulated years for every climate sub-area (Fig. 5). The frequency distributions give indications of the risks incurred in the various climate sub-areas, when no crop protection is practised. The risks are expressed in five classes: high, medium, low, negligible and unknown, and mapped accordingly (Fig. 6). In this way the old data bank of Atlas No. I, serves an input to the new data bank of Atlas No. III.

This data bank of crop losses in cereals is a closed bank in so far, that now no new data will be supplied since the project of producing an atlas is almost finished. Updating is not possible due to a lack of funds and organization. But new possibilities in

the same field are developing.

In the Netherlands a research project started this year to construct a computer-based warning system for cereal diseases. The purpose of this system is to give the farmers guidance in crop protection by means of advice on the use of chemicals and the choice of varieties (supervised control). The advice will be based on early detection of plant diseases in the crops, and on calculation of the rate of progress of epidemics by means of computer simulations. The early detection of plant diseases will be carried out by interested farmers with the help of the agricultural services. About 2% of the wheat crops will be taken up in the system proportionally distributed over the cereal growing areas and the main varieties. Since a check on the effectivity of this system is necessary information will not only be collected on varieties, disease, and degree of attack, but also on fertilization, crop rotation, chemicals applied, and yield, in order to calculate the profits. If the system is profitable to the farmers, every farmer can participate. It is evident that the data collected for the purpose of reducing crop losses by plant pathogens will form a data base much more detailed than the data base mentioned above.

Acknowledgements

Thanks are due to Dr J. Tammen who explained the Pennsylvania Plant Disease Information Storage and Retrieval System to the junior author.

References

Broekhuizen, S., 1969. Agro-ecological Atlas of Cereal Growing in Europe, Vol. II. Atlas of the Cereal-Growing Areas in Europe. Elsevier, Amsterdam, 157 p.

Rijsdijk, F.H., 1975. A simulator for yellow rust on wheat. C.R. Fac. Sci. Agron., Gembloux (1975), p.411-418.

Rijsdijk, F.H. & J.C. Zadoks, 1976. Assessment of risks and losses due to cereal rusts in Europe. Proc. 4th Eur. Mediterr. Cereal Rusts Conf., Interlaken, 1976, p.60-62.

Scholten, K.F., 1976. Data banks on pesticides; a feasibility study. EPPO/OILB Conference, October 1976, Paris.

Thran, P. & S. Broekhuizen, 1965. Agro-ecological Atlas of Cereal Growing in Europe. Vol I. Agro-climatic Atlas of Europe, Elsevier, Amsterdam, 150 p.

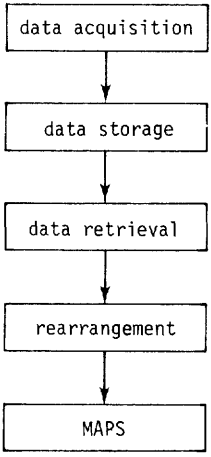


Fig. 1. Simple block diagram of data bank on crop losses

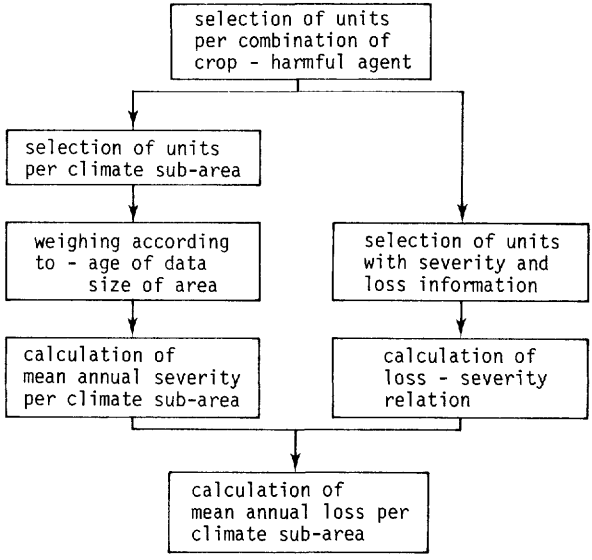


Fig. 2. Simplified block diagram of data handling

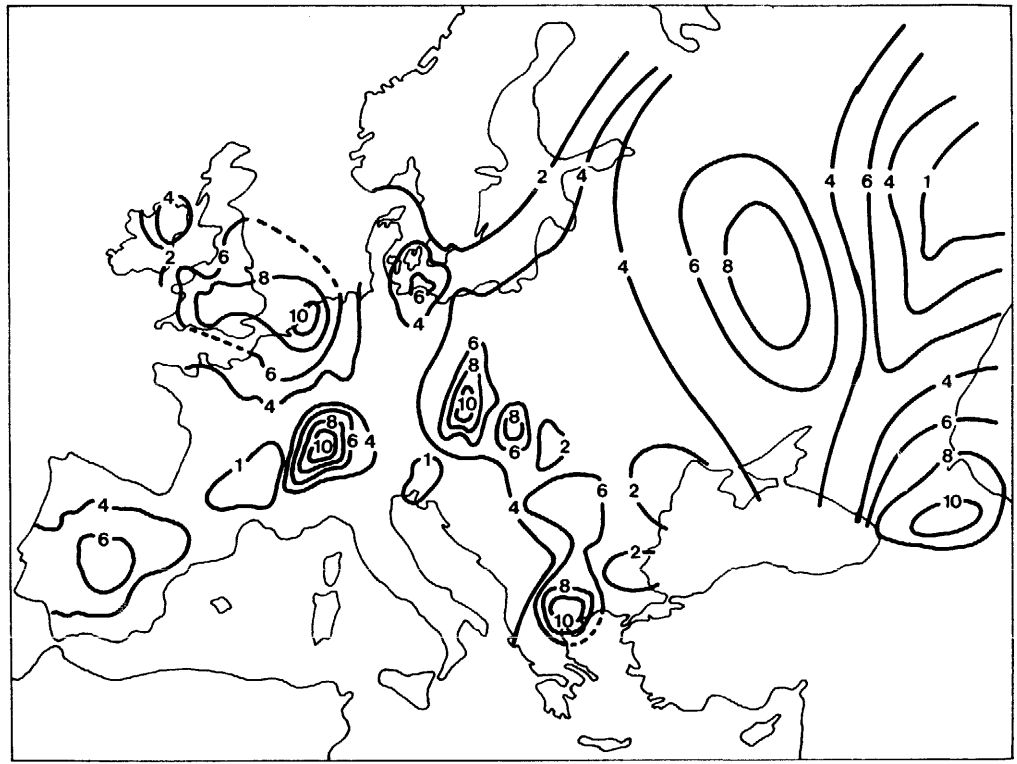


Fig. 3. Provisional map of mean annual crop losses of wheat due to yellow stripe rust (Puccinia striiformis) in Europe. Curves are *isoloss* curves connecting points of equal losses.

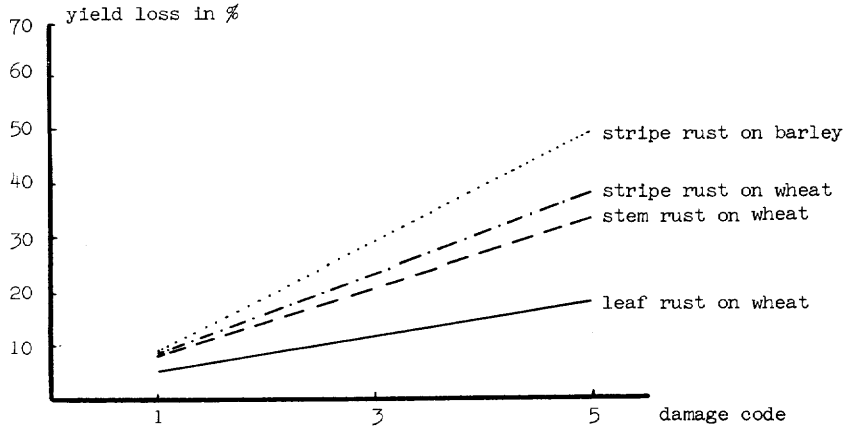


Fig. 4. Relation between damage code and yield loss in per cent, calculated from the opinion of the specialists. Abscissa - damage code: 1 = light; 3 = moderate; 5 = severe. Ordinate - yield loss in per cent.

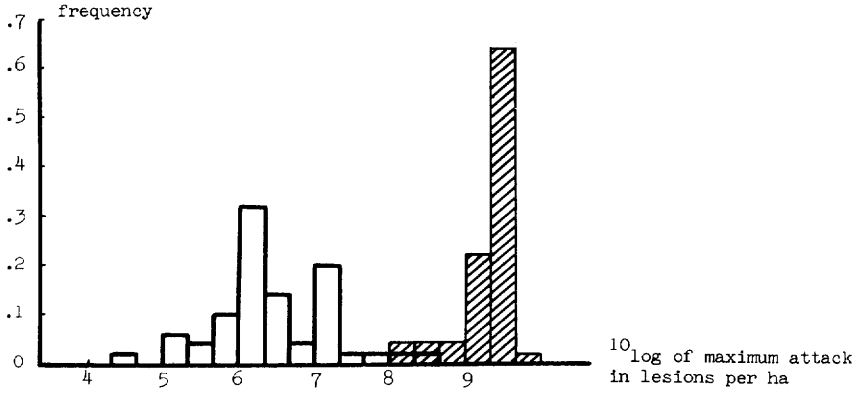


Fig. 5. Frequency distributions of the maximum attack of yellow stripe rust (Puccinia striiformis) on winter wheat during the growing season, calculated for 50 years. Abscissa - $\log_{10}$ of the maximum attack during the growing season in lesions per hectare. 100% severity equals the log value 9.7. Ordinate - relative frequency per class. Open columns: climate sub-area 37 (southern France). Hatched columns: climate sub-area 45 (east England).

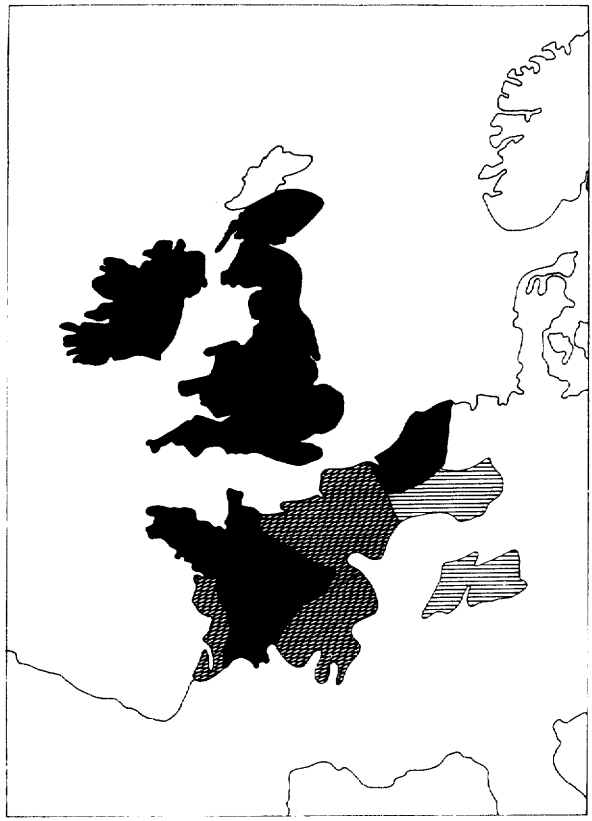


Fig. 6. Provisional map of relative risk of epidemics by yellow stripe rust (Puccinia striiformis) on winter wheat in Europe due to weather factors only. Curves delineate areas with equal relative risks, *isorisk* curves. H (risk is high) = black; M (risk is medium) = cross-hatched; L (risk is low) = hatched; N (risk is negligible) = white; 0 (no data available) (yet)) = white.

Establishment of a data base for gene bank purposes

L. Seidewitz, Institut für Pflanzenbau und Saatgutforschung, Braunsweig-Völkenrode, Federal Republic of Germany

Summary

Plant Germ Plasm Centres have been set up in order to preserve, procure and registrate plant genetic resources for future use in plant breeding. The value of plant genetic resources depends on the degree of comprehensiveness and reliability of available information. For this reason stored material is being evaluated continuously. In this connection numerous information arise which are kept in data bases after corresponding formatting. Data bases are continuously supplemented by means of data of foreign origin. By this means both the intrinsic value and the accessibility of immediately available material and of foreign resources is being increased. Our own data bank activities are supported by means of a 'Thesaurus for the International Standardisation Of Gene Bank Documentation', by means of which a terminological basis has been provided for an international information exchange.

In Braunschweig, information is stored on a worldwide basis with regard to the economic value of the material concerned. Actually, special attention has been paid to wheat and potatoes. There is a wish for our data bank to form part of a European data base concept and to solve terminological problems that we have in common with literature documentation in order to facilitate access to evaluation results that have been published elsewhere.

Data base work has been started with an elder version of the data base system SESAM (by Siemens). In continuation the so-called system FIDAS by the the Gesellschaft für Mathematik und Datenverarbeitung (Association for Mathematics and Data Processing) will be used.

As a result of the world wide change in man's natural environment, approximately 20 000 plant species are threatened with extinction. This figure is a high multiple of the plant species used in the world economy. Many potentially useful plants are to be found among the threatened species.

The natural variety of forms of plant species in use has been reduced by various processes. It is probably during the breeding of homogeneous varieties in accordance with variety-and seed-trade-legislation that the most serious process occurs, narrowing the genetic basis of cultivated species and thereby raising the spectre

of vulnerability in our economic plant species. Primitive forms with resistance properties or particular adaptability to extreme environmental conditions disappear, especially in the developing countries being identical with the natural centres of diversity, with the introduction of high yielding varieties bred for yield and are thereby lost for later cross-breeding. Owing to the use of herbicides or the cultivation of hitherto unused land, wild forms of our cultivated plants, that also constitute valuable starting material for plant breeding, are threatened with extinction.

A world network of gene banks for the preservation and use of plant genetic resources has arisen as a consequence of genic erosion caused by man. The value of this collection however, depends solely on its description according to use-oriented criteria. The genic material is being gradually evaluated in order to establish the value of its properties. To start with a systematic screening was carried out on the existence of specific properties according to the degree of specialization and capacity of available facilities, i.e. a quantitative and then a qualitative analysis was carried out on the collections.

Information for identification, morphology, physiology and information suitable for specific uses is obtained from material sorting and material examination. Worldwidely this information is increasingly being stored in data bases after corresponding formatting. The resulting opportunities for information retrieval and dissemination provide a basis for the development of plant genetic resources for selective use in plant breeding. Work is currently in progress on the problems of communication and the exchangeability of the information and data in question in several languages between various parts of the world.

The data base in Braunschweig (Brunswick) is thought to perform in future a nucleus for a concentration of information where the need is greatest, above all, for Western Europe.. As a result of the activities of a German-Dutch working party on potatoes, much attention has been paid in the last few years to potatoes in information acquisition. Likewise, much of the information currently available in the data base is on wheat. The data base was hitherto temporarily housed in a research institute about 15 km away, and was accessible via a terminal with a restriction on real-time processing. Previously, the input of about 30 000 logical records was carried out by batch processing. Owing to the conversion to 'time sharing' of the computer and operating system at the hosting research institute, a further 40 000 records are not yet directly accessible. A visual display unit is being installed in June 1977 for on-line-communication with the data base, which will be relocated on the FAL site in Braunschweig from the end of 1977.

With the installation of the visual display unit punched cards will no longer be used exclusively for data input; the on-line system will replace them widely. The SESAM data base processor, developed by Siemens, was hitherto used for the internal administration of the data base. Owing to the conversion of the data base to a time-sharing system, the FIDAS processor of the Gesellschaft für Mathematik und Datenverarbeitung (Association for Mathematics and

Data Processing), Bonn, was instituted. This system enables data bases to be segmented with a corresponding rational link-up for records and data fields. By this means it is intended to achieve redundancy-free once-only storage to a large extent. FIDAS is characterized especially by the flexability of the file editor and the report generator.

Hitherto data transcription was carried out according to one of the recommended formats for international data exchange on punched cards. In this connection two varicoloured decks of punched cards are produced each time during the punching of records compiled according to files, namely with the search-terms or descriptors which appear in files and with descriptor states. We use the search-terms in German. This card deck can be readily reproduced with foreign equivalents for data transmission. In foreign languages however, we are limited to the representable character store of our card puncher.

The aims of the Gene Bank amount to a computerized system of linking of synonyms of search-terms in various languages, so as to be able to have recourse to the data bank with the equivalent search-terms of various natural languages. The effort toward international standardization of descriptor states is to be seen in this context, so that in the future it will be possible to insert outside data on magnetic data carriers into the data bank without any problems.

Future correcting and up-dating of the data bank will be carried out as often as required, using a visual display unit, or as soon as corresponding information is ready for input.

By this means, all possibilities of on-line access should be fully exploited. While one person supervises the data bank maintenance, his colleagues who are involved in various activities in the Gene Bank will have access to the data base. The next step is to make the data base freely accessible to external users. Confidential data are not yet available for storage in the Gene Bank.

The data base is used when compiling the yearly report and to support the current Gene Bank information service of the Institute. The information store is used for the periodical listing of stored material and is also made available for information retrieval associated with specific user profiles. The FIDAS data base processor supports the statistical analysis of the data bank. The opportunity of expanding the system gradually by means of special modules for further data analysis exists.

The construction of what is now a 5-part 'Thesaurus for the International Standardization of Gene Bank Documentation' has been started in order to provide a means of communication with the data base and to define individual interest profiles. The development of international co-operation in the sphere of plant genic material preservatuon includes the extension of the present German-English and English-German Thesaurus editions to include Spanish, French and Russian.

It is considered worthwhile to make the work carried out at the data bank an integral part of European documentation in the agricultural sector, so as to be better able to respond to interests not limited to a specialized field.

Data bank of the soil science department of INRA

J.-P. Legros, INRA (Soil Science Department), Montpellier, France

Summary

The Institut National de la Recherche Agronomique Soil data bank is actually dealing with four main files. The first and the second ones are built to store, display and process soil profile data. The aim is to simplify both the writing of the soil survey reports and to open new research fields. Nearly 3000 descriptions and analysis of profiles are recorded. The third file links data related to the geographical pattern of soils. The objective is to computerize the drawing of management maps obtained from the pedologic maps associated with a lot available environment data. This file contains actually only a few examples. The last file gathers together the data necessary to identify all existing French pedologic maps. This is information for which we have often been asked.

Types of files

The bank comprises four types of files:
- descriptive file of pedological profiles (DESCRIPTION);
- analytical file of pedological profiles (ANALYSIS);
- file of pedological maps (MAPS);
- national index of pedological studies (INDEX).

Structure of the bank

The scientist in charge has the assistance of a full-time technician, a part-time programmer/analyst and trainees. The bank is installed on an IBM 360-65 computer using the OS-NVT system and owned by the University of Montpellier. This machine is equipped with 512 K of fast storage + 2000 K of LCS. Among the available compilers we use mainly the following: Cobol ANS, Fortran G, PL1 F, NLT. In the department we have a MITRA 15 terminal.

Description and analysis files

Data acquisition is by means of a field proforma drawn up on the multiple-choice principle. These fieldsheets refer to the terminology of specially prepared glossaries. What is actually recorded are the main descriptive and analytical characteristics of the soil types, together with certain details of their geological, geomorphological and vegetation environment. This is naturally supplemented

by such information as soil classification, name of the horizons, reliability and representiveness of the whole profile.

The information is coded and recorded in a standard format on cards. Between 20 and 25 punched cards are needed for each soil profile. The data are then checked and transferred to magnetic tape.

Five specific working programs have so far been developed for the print-out and statistical processing of data (contingency tables, averages, standard deviations, modal value etc.). Nearly 3000 profiles have so far been stored.

Maps file

This file contains a record of all data appearing on the geological, pedological and phytosociological maps which together concern a particular region. For this purpose the area is divided using a regular square grid.

The aim is to use these basic data for the automatic generation of land utilization maps concerning the area under consideration (suitability for various agricultural uses, drainage and irrigation needs, soil potentiality etc.).

At the processing level the system (termed INRA 6) comprises three essential functions:

- statistical function: extraction of minimum, average and maximum values, summary of information on areas in question, calculations on degree of accuracy obtained, etc.;
- factor combination function: data can be combined in order to retrieve, for example, soils considered suitable for the cultivation of peach trees, which are defined as 'deep, not chalky, on slightly sloping surfaces'. Various systems of logic can be used for retrieval: the Boolean retrieval is implemented, another is being tested;
- graphic function: the results of the sorting and calculating operations are represented in the form of histograms, maps produced by the printer or block diagrams obtained on a plotting board by connecting the system to the CARTOLAB library (J.L. Mallet, 1976)

At the programming level, INRA 6 is made up of a series of PL1 subprograms comprising a total of about 2000 statements. The system is activated by means of a number of very brief command programs which are kept on cards.

Index file

Following a considerable number of requests, we have created a file to index every pedological study carried out in France with a soil map: département, scale, subject, year, title, area concerned, authors' names, name of organization responsible for the study, reference number of this organization, name of requesting body. A print-out program has been used to print the index, which also contains all necessary references on the persons and organizations involved in this work.

At present 1400 titles have been collected. A special program has been developed to allow various statistics to be obtained: surface covered each year,surface covered per département, per subject etc.

Development prospects

At present efforts are being devoted essentially to the INRA 6 computer-assisted cartography system, since the data bank is functioning normally under its own steam. In the future a certain number of systems are to be put into operation with a very practical aim in view: to promote greater knowledge about and improved protection and use of the agricultural environment.

Additional documentation

Agence de Cooperation Culturelle et Technique, 1977 - forthcoming. Glossaire de description des horizons, glossaire de description des profils, glossaire de description de l'environnement, glossaire des techniques analytiques. 9 Avenue de Messine, Paris.

Favrot, J.C., J. Argeles, P. Falipon, E Hensel & J.P. Legros, 1977. Répertoire National des Etudes Pédologiques. INRA, Montpellier, No. 301, 305 p.

Legros, J.P. 1974. Fiche de description du profil pédologique d'après les glossaires de pédologie du Groupe Informatique et Biosphère. SES No 158, 2nd edition, ESNA, Montpellier.

Legros, J.P., 1974. Le stockage informatique des données pédologiques - avantages et contraintes. Colloque Informatique et Environnement, Arlon, Belgium.

Legros, J.P. & E. Hensel, 1978 - forthcoming. INRA 6 Computer-assisted cartographic system. 2nd meeting on Soil Information Systems, Sofia, Bulgaria.

ORSTOM, 1969. Glossaire de description des horizons en vue du traitement informatique. 70 route d'Aulnay, Bondy.

Mallet, J.L., 1976. Programmes de cartographie automatique, présentation de la bibliothèque CARTOLAB. Science de la Terre, No. 7.

An integrated information system for soil survey

S.W. Bie and J. Schelling, Netherlands Soil Survey Institute, Wageningen, the Netherlands

Summary

Introduction
A pilot project to establish a computer-based information system for environmental data was executed in the Netherlands (1972-1976) using earth science information of the Soil Survey Institute and Geological Survey.

Outline of design
The system has been designed to allow the input, storage and retrieval of point, line, area and time information in on-line and off-line modes. It aims at capturing archive material with minimum human transliteration. The system is of modular design. The system is written in FORTRAN (ANSI IV) and is implemented on a CDC Cyber 72 (131 K Words), with a Nova 1200 controller of the graphics module.

Input
Input modules include fixed format, variable and free language input. The current version of the system accepts punch cards, paper tape, magnetic tape and terminal input. Much use is made of optical character readers for capturing information.

File organization
The system supports two central systems for file organization. The heart of the system is the *relational* data base management system in G-EXEC of the Natural Environment Research Council in Britain. G-EXEC includes a large number of application programs. G-EXEC contains a module for interactive input and vetting of data, but the system itself is batch oriented.

The WIA system supports GRASP of the US Department of the Interior. This is a very user-friendly system requiring minimal training.

Interfaces
Data from all input modules and transfer of data between G-EXEC and GRASP takes place via the standard interfaces. The cartographic modules will be connected in a similar way.

The G-EXEC module includes programs to produce and accept generalized FILEMATCH files. Any external supplier or user of WIA data only has to write one interface: from his/her own system to FILEMATCH.

Automated cartography
The WIA system includes two sets of cartographic packages. On the main CDC Cyber 72 computer are currently operational 3 software packages developed by Harvard University (USA): the line printer mapping system SYMAP, the vector system CALFORM and the 3D block diagram system SYMVU. A complete stand-alone automated cartography unit is incorporated as part of the WIA system. It is a purchased 'turn key' system from the US firm Computervision. The WIA Computervision system includes a large combined digitizer/plotting table, a 19" storage CRT unit for interactive editing, and disk, magnetic tape and teletype peripherals. Also included is extensive software for cartographic preparations, with special emphasis on polygon manupulation. The system is controlled by a Nova 1200-type minicomputer, currently with 32 K Words. The output magnetic tapes of the Computervision system may be accepted by G-EXEC for further processing if required, and also linked to SYMAP whereby transformation from raster to vector maps become possible.

Application packages
G-EXEC and GRASP contain extensive statistical packages for general use. G-EXEC allows the user of the system to freely add his own application program.

Conclusion
The WIA system is an attempt to utilize existing knowledge in Europe and North America to build an up-to-date user-friendly generalized information system for environmental sciences. It is a system with modular design and generalized interfaces to allow adaptation to local needs and communication of information between system. It was adopted from 1977 by Netherlands Soil Survey Institute and Netherlands Geological Survey.

More powerful data handling allows new methodological research to progress. International soil data transfer is currently limited due to lack of normalization in classification and analytical methods. A joint Dutch-German soil mapping project exemplifies a possible future approach. Much remains to be done to arrive at a European land classification. A number of basic soil problems must be solved before international exchange of soil data becomes practical.

In the Netherlands, the Ministry of Economic Affairs has financed from 1972 a pilot project to establish a computer-based information system for environmental data. By the end of 1976 the basic system was nearly complete. In 1977 systematic use of this WIA system (Werkgemeenschap Informatiesysteem Aardwetenschappen) began by the Netherlands Soil Survey Institute and the Netherlands Geological Survey.

Organizationally it is worth noting that the WIA project was an interdepartemental one, linking ministries responsible for agri-

culture, mineral and petroleum exploration, meteorology, rural and urban planning, remote sensing and cadastral mapping. This sharply focussed project, with its own staff and direct financing, eliminated many of the normal stumbling blocks associated with interdepartmental cooperation. It also strengthened the ties between the Soil Survey and the Geological Survey linking the data of the two closely related organizations. The funds made available for the project was *f* 1.1 million (approx. US $ 400 000).

Outline of design

The system has been designed to allow the input, storage and retrieval of point, line area, and time information in on-line and off-line modes. In particular, it aims at allowing existing archive material to be captured for computer use, with minimum human transliteration. We have thought it especially important to provide a number of user-friendly entries into the system, allowing the 'computer-naive' user to operate a sophisticated system. The philosophy of this approach was outlined by De Heer and Bie (1975). The system designers have strived to utilize recent advances in hardware and software in a modular (building-block) design.

We have attempted to use already existing modules wherever possible. Specifically we have aimed at creating a generalized system that may be applicable to both small and large problems. The modular design allows large parts of the system to remain non-functional

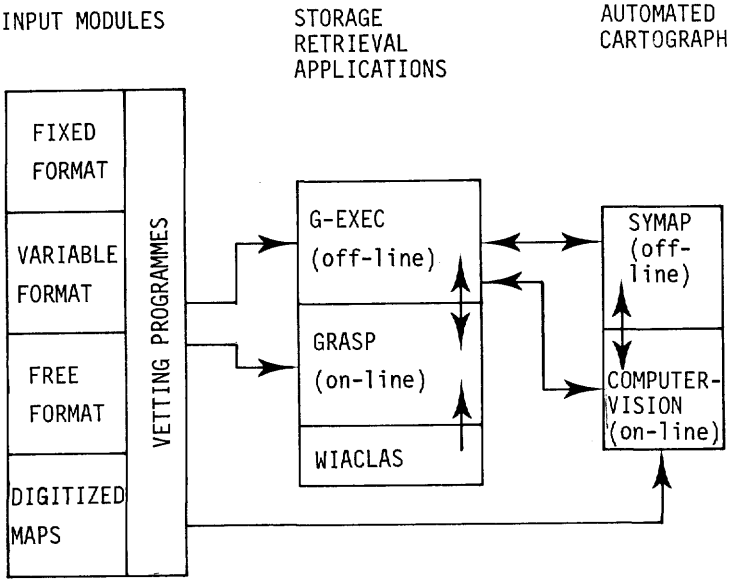


Fig. 1. Information system earth sciences.

when only minor parts are needed, thus reducing processing costs.

The system is almost completely written in ANSI FORTRAN IV, and is currently implemented on a control Data Corporation Cyber 72, with 131 K Words under NOS/BE operating system.

A NOVA 1200 minicomputer acts as a controller of the graphics modules. As many of the software modules originate from other machines, the system is largely 'portable'.

We can consider the system as having five basic compartments: input, file organization, automated cartography, application programs and interfaces (Fig. 1). The latter consists of modules scattered among the other compartments and although frequently invisible to the user, they constitute a most essential element of a modular soil information system. The success of such a design is totally dependent on the performance of the interfaces.

Input modules

Input modules include normal fixed format input, variable input where the attribute is identified by its name (or tag), and free language input where word lists are employed to identify attribute-value pairs from normal or near-normal text.

Fixed format input

An example of fixed format is the input of soil analytical data via mark sense forms. For each attribute (soil property) has been reserved a fixed field of fixed length. With this input form (read optically by a commercial service bureau) short lines are drawn with a soft pencil across the values that apply.

Variable format input

The landscape forms attempt to describe the attributes of the visible landscape per space or mass (e.g. of trees). As only few of the total number of attributes are present in any one space, the identification by tag ($ plus three letters) forms a useful alternative. Only those attributes are punched (or otherwise entered into the computer) that happen to have values (see Bie, 1974).

Free format input

Borelogs or soil profile descriptions constitute an extreme form of variable input. Here the attribute name is frequently implicit rather than explicit. The value *blue* defines implicitly the attribute-value pair *colour: blue*.

Our archives consist of large numbers of such records (more than 0.5 million). There is no personnel available to change this manually to computer-acceptable input. We have constructed a system (Bie et al., 1975) that works thus:

1. A (geologically, pedologically) unskilled typist types the information on sheets according to specific typing instructions.
2. The typed manuscripts are read by an optical reader, processed by a formatting program and dumped on magnetic tape.

3. The tape block is read into a vetting program. The program automatically identifies attribute-value pairs by identifying the words (values) in word lists. Alternative spellings and terms, together with commonly occurring errors, are automatically normalized. Errors, to which the program cannot find a solution, are printed on a terminal for the scientist to solve.

Conclusion

Reflected in the input modules described above is the following philosophy of the designers:
- make the input preparation user-friendly, preferably accept present type input;
- do not use scientific staff for mundane tasks, let the computer do the donkey-work;
- leave the last word with the scientist.

File organization

The system supports two central systems for file organization. The heart of the system is the *relational* data base management system G-EXEC of the Natural Environment Research Council in Britain (Jeffery et al., 1975). This is a powerful data base management system particularly developed for environmental data. It utilizes the concept of relations between data (rather than hierarchical structures) as originally formulated by Codd. These structural relations enable geographical relations in depth and space to be expressed efficiently, and to link data that are logically linked also when the linkage is not hierarchical. G-EXEC contains a large number of application programs, including numerical analysis of data and plotting routines. In the first phase of operation of the WIA system work with G-EXEC is confined to our computer support staff. We expect especially motivated soil scientists to operate G-EXEC independently in the second phase over a year or so.

To enable less computer-minded soil scientists to utilize the computer to interrogate their files, to perform simple statistical analysis and prepare tables for further study, we have also implemented the real-time interactive GRASP retrieval system of the US Geological Survey (Bowen & Botbol, 1975). Here light terminals connected by telephone lines to the Cyber (e.g. teletypes or video displays) enable easy access to the computer. Any reasonably motivated soil scientist can, with minimal instruction, get down to practical work.

A beginners manual (Bie et al., 1977) of which an English version is also available (Bie et al., 1977a), attempts to lower the threshold further, and pays attention to error messages that may occur as result of illegal commands, typing errors etc. GRASP is essentially also a relational file system, employing the simplest case - the single relation. We feel that GRASP is a powerful educational tool though limited in its ability to perform complex retrievals, it wets the appetite of the earth scientists and give them incentive to develop further data handling interests.

GRASP and G-EXEC are linked through internal two-way interfaces

allowing e.g. a preliminary retrieval to be performed in batch mode under G-EXEC, and be passed on for further interactive work under GRASP. Externally G-EXEC can produce the FILEMATCH format, acceptable for input in the French SIGMI system, the West-German DASCH system, the Canadian SAFRAS and the generally available SYSTEM 2000. G-EXEC can also accept FILEMATCH files produced by other systems. This concept of file transfer is pioneered by COGEODATA of the International Union of Geological Sciences, and deserves close attention from the soil science world.

Automated cartography

The WIA system includes two sets of cartographic packages. The Cyber supports three software packages developed by Harvard University (USA): the lineprinter mapping system SYMAP (current version implemented 5. 19), the vector system CALFORM and the three-dimensional block diagram system SYMVU.

SYMAP, with its ability to handle data in raster rather than vector(line) format, allows maps to be constructed from point data with continuous values (contour or isoline map), with class value (proximal map) or just a line map translated into raster format (conformant map). We have extended the SYMAP system by adding modules that allow several maps to be overlaid within the computer, with user-selectable algorithms between the overlays, so that a synthesis map can be constructed. This extension to the multivariate case offers much prospect in soil classification and in environmental mapping. Examples of the technique have been published by De Gruijter and Bie (1975), Bie (1976) and De Gruijter (1977). A recent addition to this module allows the normal lineprinter map to be changed into vector format. Smoothing algorithms will (cosmetically) create a normal map image, if desired. The image can be edited further in the Computervision system. This is a complete 'stand-alone' cartographic unit, incorporated as part of the WIA system. It is a purchased 'turn-key' system from the US firm Computervision.

At the moment the WIA Computervision system includes a large combined digitizer/plotter, a 19" storage CRT display for interactive editing and a 14 M words disk, a 9 track 800 B.P.I. magnetic tape unit and teletype peripherals. Also included is extensive software for cartographic use, with special emphasis on polygon manipulation so important for soil maps. The system is controlled by a NOVA 1200 minicomputer, currently with 32 K Words. At the moment only one user can operate the system at any one time, a multiuser capability will shortly be added. The system can support 4 users, and also be equipped with an automatic scanning option, which *automatically* digitizes line patterns. This offers much hope for more efficient automation in the future than is possible with current hand digitizing methods.

Two-way interfaces exist to link the Computervision system with the Cyber, for use with G-EXEC and SYMAP. This allows information on the digitized borehole/profile description locations to be passed back to the main computer for inclusion with the descriptions themselves.

Application programs

All programs described above contain application software. G-EXEC and GRASP have extensive statistical packages for general use. SYMAP and the Computervision system also include various possibilities for user manipulation and statistical summaries.

Added to this is the WIACLAS module operating on the Cyber system. This is an interactive identification program that allows an unknown individual to be allocated to a class on the basis of its similarity to other, already classified, individuals. WIACLAS is self-learning, in that new identifications may be added to existing ones, thus increasing the 'experience' of the system. Interesting applications of WIACLAS are in geological stratigraphy (Bie, 1976) and in soil suitability classification (Bie et al., 1976b).

Potential impact of a soil information system

The WIA system is an attempt to utilize existing knowledge in Europe and North America to build a user-friendly generalized information system for environmental sciences. It is a system attempting to offer facilities for the many activities of a soil survey organization, with modular design and generalized internal and external interfaces.

Whilst we are pleased with the progress made in the machanization of tedious and unattractive tasks in manual soil data handling (and make our large archives potentially usable) the last two years of developments have strengthened our feeling (Schelling, 1975; Bie, 1975) that with soil information systems will come a new and more powerful approach to our branch of soil science. A new and deeper insight into soil science problems may be obtained from direct access to the 'atoms' of the soil system: the raw data. This realization is not confined to soil science; other earth scientists draw the same conclusion (Hubaux, 1973). A phase of building soil information systems will be superseded by increased research in methodology (e.g. Webster, 1975). We hope that in the Netherlands our WIA system will provide a tool also towards this purpose.

The main tasks now to be faced in the earth sciences are these:
1. International and interdisciplinary data exchange. With respect to soil taxonomy soil science is still in the pre-Linnaeus stage. There is no internationally accepted general classification, and no normalized set of soil characteristics or laboratory methods. These are probably the worst drawbacks for international cooperation.

The International Society of Soil Science has since 1975 had a Working Group on Soil Information Systems in its Commission V. The purpose of this group is mainly to exchange knowledge and experience. The group has met twice in Europe and a subgroup once in Australia. The proceedings, a newsletter and a project catalogue are results of its activities.

FAO has made a first sketch for an International Soil Data Bank. This ambitious idea will require much money and energy, commodities that are not easily available at present.

Our geological colleagues are in general further advanced. They

have COGEODATA, in the International Union of Geological Sciences. This powerful group does not only exchange information, but makes overviews of the state of the art, and has developed the exchange format between some of the main existing systems, FILEMATCH.

The International Council of Scientific Unions Committee on Data for Science and Technology has set up a task Group on Space and Time Dependent Data. The object of this group is to set up and execute a project that would improve and accelerate the interdisciplinary transfer of knowledge of space and time dependent data handling methods.

2. Cooperation between soil survey institutes in Europe. The publication of a soil map of the world on scale 1 : 5 million is in progress, the sheet of Europe will probably be the last one to appear. A soil map 1 : 1 million of Europe has been under construction for many years. Several parts are present in manuscript form. A discussion on land classification has started on European level, but it will probably take several years to come to a common method. (Technical Consultation on Land Evaluation for Europe, Nitra, Czechoslovakia, 1975).

In a joint project between the German and Dutch soil surveys one 1 : 50 000 soil map covering border areas has been published, both with the German and the Dutch map legends. A second map sheet will be surveyed in the next years.

3. The need for exchange of soil data in Europe. There is no doubt, that in the near future for environmental and agricultural purposes exchange of soil information will be needed. As long as several basic problems are not solved, the total exercise cannot be initiated. Among these basic problems the normalisation of classification, parameters and methods of analysis is of prime importance.

References

Bie, S.W., 1974. The representation of geological and soil information within an automated data handling system. Bull. Int. Ass. Engineering Geology, 10, p.27-29.

Bie, S.W., 1975. Towards integrated integrated soil information systems. In: Bie, S.W. (ed.) Soil Information Systems. Pudoc, Wageningen, the Netherlands. p. 73-78.

Bie S.W., T. de Heer, C.J. van Staalduinen, J.R.E. Lieftinck & W.M. Dam, 1975. Automated data capture from poorly structured geological borelogs. Geol. Jb. Hannover A 28: p.15-25.

Bie, S.W., 1976. New methodologies for geological surveys. Computers and geosciences 2: 341-344.

Bie, S.W., C.A. Benschop, J.R.E. Lieftinck & N. Hofstede, 1976a. GRASP for beginners. Version 1. Stichting voor Bodemkartering, Wageningen, the Netherlands.

Bie, S.W., J.R.E. Lieftinck, K.R. van Lynden & A.W.Waenink, 1976b. Computer-aided interactive soil suitability classification - a simple Bayesian approach. Neth, J. Agric. Sci. 24: 179-186.

Bie, S.W., C.A. Benschop, J.R.E. Lieftinck, N. Hofstede & W.G.J.J. Hamer,1977. GRASP for beginners. Version 2. Stichting voor Bodemkartering, Wageningen, the Netherlands.

Bowen, R.W. & J.M. Botbol, 1975. The geological retrieval and

synopsis program (GRASP). U.S. Geol. Survey professional paper 966.
Gruijter, J.J. de & S.W. Bie, 1975. A discrete approach to automated mapping of multivariate systems. Trans. Comm. III, Internat, Cartographic Assoc., Enschede. p. 17-28.
Gruijter, J.J. de, 1977. Numerical classification of soils and its application in survey. Pudoc, Wageningen, the Netherlands.
Heer, T. de & S.W. Bie, 1975. Particular requirements for the Dutch WIA Earth Science Information System. Geol. Survey Canada Paper 74-63, p.78-81.
Hubaux, A., 1973. A new geological tool - the data. Earth Science Rev. 9: 159-196.
Jeffery, K.G., E.M. Gill, S. Henley & J.M. Cubitt, 1975. G-EXEC system user's manual. Atlas Computing Division, Chilton, England.
Schelling, J., 1975. The role of soil information systems. In: Bie, S.W. (ed.) Soil Information Systems. Pudoc, Wageningen, the Netherlands, p. 13-16.
Webster, R., 1975. Sampling, classification and quality control. In: Bie, S.W. (ed.) Soil Information Systems. Pudoc, Wageningen, the Netherlands, p. 65-72.

An automated information system for soil survey

R. Webster, Soil Survey of England and Wales, Rothamsted Experimental Station, Harpenden, United Kingdom

Summary

The Soil Survey of England and Wales has some fifty field scientists, mainly engaged on mapping. Surveyors record about 120 000 items of information per year in a computer compatible form, and these records are used to prepare reports on the soil of each site in a conventional layout automatically. This information, however, represents only about one tenth of all observations. The remainder are generalized into classifications or maps and thereby lost.

The Survey aims to record and handle all observations, and is studying both the requirements of its clients and the methods needed to furnish them. It has therefore begun a pilot project for one region. In collaboration with Oxford University it has established a regional store for Oxfordshire, containing soil information from a variety of sources. The data will be used to demonstrate the feasibility by automation to serve the Survey's clients and to explore real needs. A systems analysis has been made, and the data management program, G-EXEC, is being implemented for the purpose of the project.

The Soil Survey of England and Wales has some 50 field scientists operating from 19 offices throughout the country. The Survey's main task at present is to prepare maps of the soil of selected 100 km^2 areas at a scale of 1 : 25 000. The maps are made from observations made at some 30 to 40 points per km^2. The soil is examined using a spade or auger at each point, and its colour (including mottling), texture, stoniness, presence of carbonate and other properties observed at each of several depths. A few surveyors record all they observe, some record all they observe at a proportion of the inspection points only, some record only a few soil properties at most inspection points, and some record only a soil class name that subsumes all they see at each point. Between 1000 and 1000 km^2 are surveyed each year, and the results presented as maps showing soil types separated by boundary lines.

For each soil type shown on a map one or more detailed descriptions are made of the soil from pits dug to about 1.5 m. About thirty characters are recorded at each of several levels or horizons in the field. Samples are taken from pits for particle size and chemical analysis, for determining the moisture characteristic, and for microscopic examination. Some 10 to 30 properties are measured in the laboratory on samples from each level or horizon. Data derive

from several sources; namely field centres, the chemistry laboratory at Rothamsted Experimental Station, the soil water laboratory at Derby, and the micromorphological laboratory at Rothamsted.

Several hundred pits are dug and described each year. All the descriptions and results of analysis are being recorded in computer-compatible form. At a rough estimate the Survey now accumulates 60 items of information on each of four levels or horizons in the soil at each of 500 sites, together with ancillary information, each year. This is about 120 000 items per year.

All other observations are recorded only if surveyors wish. In the past there has been no satisfactory means of handling records of such observations and this has led to the present practice of generalizing and thereby losing specific site information at a very early stage in survey. Nevertheless in the West Midland region staff do record all observations in full, and accumulate about 200 000 items of information annually. If staff elsewhere record all their observations there will be a further 1 M or so items per year to be added.

Computing

The Survey has had access to a large general purpose computer (a System 4-70) at Rothamsted for several years. The machine has been used mainly for statistical analysis. But programs have also been written to screen files of coded field descriptions and translate them into text and to process the records of chemical analysis. Raw laboratory measurements of soil water are entered into the computer and converted to values of moisture retention, soil density, stoniness and particle size. In the latter form they are stored in 80-byte records. The aim now is to bring together the various data handling activities into a single management system.

Study of a prototype data store

We wish to retain as much information as possible in an easily accessible form so that it can be exploited, and to use the computer to handle all or almost all data that the Survey can collect. We are not yet sure how best to do this, partly because we lack sufficient computing expertise, but mainly because we do not know fully what either surveyors or their clients wish to know. Equally neither surveyors nor clients can visualise, let alone specify, the range of their requirements without some practical experience of using an automated data store.

A research project has been started jointly by Oxford University and the Survey to resolve this impasse by establishing an experimental prototype data store for one region, well primed with information to serve a range of users. The intention is to use the store and data handling facilities to demonstrate the feasibility of such a system and to explore real requirements. The region chosen is that of Oxfordshire, for which data are available from the laboratories mentioned above and from several others, including those of the Ministry of Agriculture, Soil Science Laboratory at Oxford, the Institute of Geological Sciences, and the Imperial College of Science

and Technology in London. A data handling system is therefore needed to do the following:
- Accept all kinds of information about the soil at sites, deriving from several field surveyors and several laboratories, which may use different recording conventions and different formats.
- Provide simple facilities for validation, editing and up-dating.
- Output to line printer, and to temporary workfiles or more permanent files which may then be used for deriving new varieties, statistical computation, translation of coded input into text, mapping, and input to existing standard programs such as SYMAP, the mapping program of the Laboratory for Computer Graphics at Harvard, and GENSTAT, the Rothamsted general statistical program.
- Most important of all, information must be held in such a way that it can be retrieved or collated for at least geographic area, soil type, local government division and the more important individual soil characters. Such output is what finally interests the user of such a system.

It is desirable that the system is easy to use from the simplest terminal, namely a teletype linked to the Post Office telephone network, by staff with little computing expertise and who are unlikely to use the computer sufficiently frequently to justify their becoming expert.

We have completed the systems analysis, and are considering existing software for our purpose. The program G-EXEC, written by the Natural Environmental Research Council computer unit mainly for geological data seems the most suited, and is being implemented on the Rothamsted computer. We expect it to be working during the summer of this year.

For the purpose of the experiment the data handling need not be computionally efficient. The important factors are ease of use and early availability. Efficiency can be considered and improved when we know more clearly users' requirements.

POSEIDON and the ORSTOM pedology data bank

R. van den Driessche, A. Garcia Gomez, A. Aubry and A. Giey, Office de la Recherche Scientifique et Technique Outre-Mer (ORSTOM), Bondy, France

Summary

Over the last 8 years, a user-friendly soil information system has been implemented at ORSTOM on an 1108 computer, under the acronym POSEIDON. Plan language is used as input, and statistics (non-parametric) is the main output. Poseidon has a wide range of applicability, and has already served into 10 data files. Any vocabulary derived from binary, ordinal, nominal and interval variables (maximum 500), plus spatial and temporal referencing, may be used. The files are updated (maximum 10 000 units and 10 subunits) and retrieved (maximum 100 conditions).

Cluster analysis and tests are performed on the retrieved records. Surfaces are obtained on the lineprinter by polynominal interpolation techniques. The most important programs have been published in 1974, other last year. Training a limited number of soil scientists is offered once a year.

Purpose

Since its establishment in 1968, the ORSTOM Pedology Data Bank has been providing assistance in computerized management and making statistical comparisons of soil profiles which have been described overseas.

Manual coding of the descriptions was out of the question due to the lack of personnel. Automatic coding meant that a unique terminology would have to be used by all ORSTOM pedologists, and that a DBMS with all finctions (storage, updating, retrieval) would have to be created, accepting the plain language of the new terminology. The terminology was drawn up in cooperation with all the interested French organizations on the initiative of the DGRST and published in 1969 and 1971. Management requirements were met by creating, in early 1970, a DBMS whose functions can be used in plain language. This management software was published in 1974 on an example not related to pedology, as complete independence from a given terminology was a requirement of the system.

Research into the statistical comparison of soil profiles is still going on, and several multivariate nonparametric methods have already been of use in this work.

The system is better known under the acronym POSEIDON for 'Procédures Opérationnelles en Statistique Et Informatique pour DOnnées en langage Naturel' released in 1975.

Hardware

A remote batch SFENA TMF400 card reader and lineprinter multidiscipline terminal is linked by a 4800 bauds line to a SPERRY-UNIVAC 1108 computer with 192 K words of STAD in Paris and to an IBM 37'/168 computer of CNRS-CIRCE in Orsay.

This year, the Pedology Data Bank accounts for 20% of the use of the terminal.

Main files

Profile descriptions are stored and updated on magnetic tape. A user-friendly retrieval procedure is available.
- Referral file: Geographical coordinates (latitude longitude in degrees and minutes), donor, date, French soil group etc., make up a 14 variates record. This is spatial and temporal referencing for all the soil profiles fed into the data bank.
- Horizons file: This file includes records containing 112 binary, nominal, ordinal, and interval variates used to describe profile horizons (an average of 4 horizons per profile with a maximum of 10).
- Environments file: This recent file contains descriptions relating to geomorphology, geology, hydrology, vegetation, human influence. There is room for 70 variates as SOGREAH. Homologous retrieved subfiles are amalgamated, e.g. for contouring purposes.

Two files on central Africa alone were also set up using the INEAC terminology to process profiles described by this institute. The first contains one record for each of the ten fixed depths in each profile, and this record contains 42 variates. The second file describes and locates the sites with records of 24 variates. Both have served as a reference file.

In addition to the above, other files have been used for specific tasks. POSEIDON has a wide range of applicability. An experiment is under way on the proposed FAO terminology.

POSEIDON as a DBMS

Following management programs are all in the implemented stage for the 1108 computer:
- Retrieval. Subfiles are retrieved by reading one of the tapes supporting the file in question, specifying the appropriate vocabulary (e.g. that of the retriever language), and supplying a complete Boolean expression. Logical AND OR operators and relational LE, LT, EQ, NE, GE, GT operators are used. Operands (variates and values) are taken from the vocabulary, in plain language or in code. There are no parentheses: AND has priority over OR. As missing values are coded automatically, it is possible to retrieve only those records which contain specified variates. The records which answer the Boolean conditions are: a) coded on tape or cards in an I format and identified, record length being determined by the total number of variates, and, b) printed in plain languege, the values being arranged in the order of the vocabulary. In the case of files comprising two levels, the printout covers the entire unit (profile), the coded subunits (horizons) being marked.

Fortran program SGYT, 906 lines, 80 K words.
- Geosearch. Questioning of the referral file may be accompanied by a request for a mapped nondistorted representation of the sites which satisfy certain conditions and which are in the geographic area determined by two parallels and two meridians. Linked Fortran programs COOR, 240 lines, 9 K words, SGYT, CARTO, 507 lines, 17 K words.
- Storage. Source documents are handwritten or typed descriptions, as well as checked lists of terms. Descriptions are fed into the files in plain language on cards, with a period as separator between values. Within a subunit, values may be in any order, and the order may vary from one subunit to another. Two compulsory keywords control the unit/subunit hierarchy. The program checks input with the wording of the vocabulary. The printout is annotated in the event of discrepancy or redundancy, while only validated values are coded magnetically. Up to 500 words of comments may be added to a unit; 10 subunits are allowed per unit, 500 values and 100 chapters per subunit, 100 000 subunits per file. The vocabularies and files of descriptions are loaded on tape: 3 tapes are used in circular permutation for reasons of safety. Fortran program DGY, 943 lines, 69 K words.
- Updating. Correction, addition, or deletion of values are made by rereading the cards of the subunits concerned. Addition of subunits to existing units is by reading additional cards. Unit or subunit deletion is by control card. Same DGY program.
- Vocabularies. Variates and values are included in a vocabulary. Values may consist of 1 to 150 alphanumerical characters (with only /-,*$) and must each be different since values may be in any order when storage, updating, or retrieval takes place. The wording of the values in the first three files can be found in the two published glossaries. Included in the vocabulary are all the equivalencies of the type 'value in plain = same value in numerical code'. A vocabulary may accept 60 000 words. Translated vocabularies permit automated translation of the records. Fortran program RGY, 763 lines, 68 K words.

The POSEIDON multivariate nonparametric statistical software

In the presentation of following programs, m stands for the total number of units or subunits, v for the number of variates choosen by format cards, missing values being accepted in all records.
- Clusters. Any dissimilarity matrix or any distance matrix may be used to build clusters defined by a smaller within-cluster average distance than the between-clusters average distance. Printout of the clusters content and average distances ($m \leq 100$). Fortran program AGGLOM, 427 lines, 89 K words.
- Stepwise clustering. Based on the smallest rank dissimilarities, this stepwise clustering program outputs the clusters content at each step. Binary, ordinal and/or interval variates are accepted ($m \leq 127$, $v \leq 2000$, $mv \leq 38\ 100$). Fortran program CAVT, 282 lines, 87 K words.
- Dissimilarities based on probability. Adapted from Goodall's program (1966), this version outputs a dissimilarity matrix on

binary, nominal, ordinal, and/or interval variates ($m \leq 100$). 852 lines of Fortran, 38 K words.
- Multi-level dissimilarities. Input covers records of values from ordinal and/or interval variates ($m \leq 250$, $v \leq 400$) and their maximums. Printout of the dissimilarities takes several forms, Fortran program DISSIM, 201 lines, 47 K words.
- Allocation. A profile is allocated to a soil type from a reference file by the smallest computed multi-level dissimilarity index ($m \leq 750$, $v \leq 400$). Linked Fortran programs SGYT, MERGE, 38 lines, 8 K, and IDENT, 99 lines, 12 K words, after the DGY program.

In addition to the above programs for the 1108 computer, 18 Fortran programs run either on the 1108 or on the 370. However, some of the latter do not accept missing values.

Users

The ORSTOM Pedology Data Bank has a permanent staff of two who deal with work requested by certain pedologists. Interested parties can gain independent access to the software and files after familiarizing themselves with the models, algorithms, and control cards of the programs of their choice. Presently there are on magnetic tape 8000 retrievable records plus 6676 retrieved records. This does not include a few parasite files lacking horizon descriptions, e.g. the one with 1210 profiles.

Training

A maximum of 10 trainees may be taken on for 3 weeks once a year. Training is free. Trainees are allowed access to the terminal, to the software and files for a limited overall CPU time.

References (truncated list)

Goodall,D.W., 1966. A new similarity index based on probability. Biometrics 22: 882-907.

Girard, M.C., 1968. Approche statistique de la notion de série sur un exemple pris dans la bordure septentrionale de la Beauce. Doctoral thesis, University of Paris, 440 p.

Glossaire de pédologie, description des horizons en vue du traitement informatique, 1969. Rapport du groupe informatique pédologique (DGRST). Init. Doc.Tech.ORSTOM 13, 82 p.

Glossaire de pédologie, description de l'environnement en vue du traitement informatique, 1971. Report of DGRST working group. Informatique et Biosphère, Paris, 173 p.

Van den Driessche, R., A. Garcia Gomez & A. Giey, 1974. Un système informatique en langage naturel, application didactique aux coloris ISCC-NBS. Report presented under DRME CONTRACT No. 73/332 (Direction des Recherches et Moyens d'Essais, Ministère de la Défense). Init. Doc. Tech. ORSTOM 25, 118 p.

Van den Driessche, R., A. Garcia Gomez & A. Giey, 1974. SIDA satellite imagery descriptors analysis. Report presented under DRME CONTRACT No. 73/332 (Direction des Recherches et Moyens d'Essais, Ministère de la Défense). ORSTOM, 165 p.

Cointepas, J.P., 1974. Application d'une analyse multivariable à quelques problèmes de pédogénèse en Tunisie. Cah.ORSTOM Sér.Pédol. 12: 145-70.

Van den Driessche, R., A. Garcia Gomez, A. Giey & A. Aubry, 1975. POSEIDON procédures opérationelles en statistique et informatique pour données en langage naturel. Report presented under DRME CONTRACT No. 73/332 (Direction des Recherches et Moyens d'Essais, Ministère de la Défense). Cah. ORSTOM Sér. Pédol. 13, 31-4. Duplicated 400 p.

Van den Driessche, R.A. Garcia Gomez & A.M. Aubry, 1975. A multi-level dissimilarity index between soil profiles. Cah. ORSTOM Sér. Pédol. 13, 173 p.

Martin, D. &. A.M. Aubry, 1975. Comparaison de profils du Congo par des distances. Cah. ORSTOM Sér. Pédol. 13: 175-90.

Aubry, A.M., 1976. Programmes Fortran pour distances de rangs, constellations et corrélations. Init. Doc. Tech.ORSTOM 30, 56 p.

Martin, D., 1976. Analyse multivariable de données analytiques de sols ferrallitiques du Centre Caméroun. Cah. ORSTOM Sér. Pédol. 14: 3-11.

Girard, M.C., 1976. Recherche d'une méthodologie pédologique en matière de traitement statistique des données de sol, applications à la taxonomie et à la cartographie. Science du Sol,4: 177-204.

Some comments on agricultural data banks: the user's viewpoint

P. Dagnelie, Faculté des Sciences Agronomiques de l'Etat, Gembloux, Belgium

Summary

This report describes the experience gained with the creation of data banks for various branches of the agronomique sciences (vegetable and animal products, soil science, scientific documentation etc.)

It illustrates on the one hand the necessity always to take very careful account of users' requirements and on the other hand the advisability in many cases to start from certain small systems already in existence.

Introduction

Data processing is now a basic tool for scientific research and business management in several fields. This is particularly true of agriculture in a general sense (crop and livestock farming, forestry, soil and environmental sciences etc.).

The rapid development of computer science in recent years has produced numerous files which, although not initially designed with data banks in mind, have nevertheless become more and more akin to them. This aspect should not be forgotten when it comes to integrating documentation systems on an international basis. We shall illustrate this by giving some examples of small-scale data banks and some relevant comments.

Examples of small-scale data banks

The concept of data banks

The expression 'data bank' almost always conjures up a picture of a large amount of information, together with a substantial equipment for automatic processing. It seems to us, however, that a larger meaning can be given to the term and that it can be applied just as well to relatively small information systems.

This interpretation would appear to be confirmed by such definitions as that proposed by Ginguay and Lauret (1973): "Data bank: data assembled in a large file and constituting the sum of accumulated information on a subject of interest to a group of users; techniques which allow the users access to the assembled data, by means of terminals for example, in order to find answers for their particular questions and to update the information." From this definition, we can deduce that the basic requirements for a data bank are the existence of one or more files which are jointly

available to several users and the possibility of consulting and updating them.

Context of the work

The context of the work referred to below is a teaching and research institute for agriculture. The institute has its own department of statistics and computer science and its own equipment. More precisely, the work was carried out within a teaching institute (State Faculty of Agriculture at Gembloux, Ministry of Education) and a research institute (State Centre for Agricultural Research at Gembloux, Ministry of Agriculture), together with special research centres (sponsored principally by the Institute for the Promotion of Scientific Research in Industry and Agriculture) with a pooled service of statistics and data processing.

This department has a staff of about twenty (lecturers, research workers, post-graduate students and technicians) serving about 700 people in all. The equipment is a relatively small computer (ICL 2903 with a main memory of 144 K bytes and an auxiliary memory of 15 M bytes). The numerous problems dealt with cover a very wide range, such as statistics, biometry, physics, economics etc. (Dagnelie,, 1975).

Examples

Files have been compiled for use by several people and programmes have been devised for the updating and consulting in the following fields:

- plant breeding: data on the main properties of different varieties of widely-grown plants and on the effect of different types of manure and plant protection;
- agricultural experiments: data on variations in field experiments;
- animal husbandry: data assembled to aid the management of large herds of cattle;
- zoology: data assembled principally for the purpose of drawing maps showing the distribution of invertebrates;
- meteorology: data which can be consulted for different research projects (yield forecasts for widely-grown crops, studies of the flow rates of rivers etc.);
- air pollution: data on pollution caused by heavy metals, especially in urban areas and close to motorways;
- soil science: data on soil profiles, environment and results of analyses;
- rural economics: data assembled during agricultural accounting operations;
- forestry economics: data for the management of forest areas.

Some of these fields come from a predominantly individual interest or are linked to the work of a single research team. Others, however, are the result of collaboration within the institute, between two or more research teams, or of collaboration at national or international level, between two or more institutes.

Comments

The various files or data banks referred to above were compiled at the request of the 'customers' of the Department of Statistics and Computer Science, in accordance with their needs. We feel that this is a vital factor, since we have found from experience that users very often tend to ignore the system which is set up when it does not answer their specific needs or when it becomes too complex or cumbersome.

The various examples listed will give some idea of the range of problems which can be encountered in a single agricultural teaching and research institute. It is reasonable to assume that this situation is not unusual and that many similar data banks exist throughout the EEC member states.

The potential value of these banks should be remembered when international systems are being set up. Alongside an 'on-line' data network for the EEC it would probably be useful to draw up a list of existing small-scale data banks and to coordinate as far as possible the various systems which relate to the same subject. Exchanges of information should also be encouraged, even on a fairly unsophisticated scale (exchanges of punched cards, magnetic tapes etc.).

Action along these lines would be closer to the practical possibilities and would provide a significant promotion to relations between research institutes. It would encourage the use of simple data processing facilities and, further away, of more elaborate systems, such as Euronet.

A final comment concerns the cost of information transmission systems. Teaching and research institutes in many countries presently suffer from financial difficulties, and it is obvious that any system that aims to be useful must not only satisfy the needs of the users and be easy to operate, but must also be as inexpensive as possible.

References

Dagnelie, P., 1975. Dix années d'informatique en agronomie. Ann. Gembloux 81: 253-260.

Ginguay, M & A. Lauret, 1973. Lexique d'informatique, Masson, Paris.

MEDISTAT: Mediterranean countries socio-economic data bank

M. Allaya, Mediterranean Agronomic Institute of Montpellier, France

Summary

Field: Macro-economics, national accounts, international trade, agricultural food production systems.

The geographical area adopted for MEDISTAT is based on the Club of Rome's ' Regionalized hierarchical world model', adapted for the analysis of the Euro-Mediterranean region. Although all countries of the world are included in the model, there are only detailed data for the countries of the Euro-Mediterranean region, other countries being grouped together in broad regions.

The model makes it possible to carry out analyses by country or group of countries for the Mediterranean, Europe and Africa, and by region for the rest of the world.

The chronological series entered in the bank cover the following fields from 1965, including the most up-to-date data for recent years:

- total population and active population by branch or sector;
- overall data on national accounts;
- agricultural production of food by crop;
- external trade in agricultural food products by product and country (imports and exports, by quantity and by value);
- balance of payments;
- production structures.

Nature of service:

The purpose of the MEDISTAT data bank is to make available to specialists (research workers, teachers, interested parties in firms or public authorities) a certain amount of statistical data in the socio-economic field on the Euro-Mediterranean region, together with a set of programmes for processing these data for analysis purposes.

The MEDISTAT bank forms part of the researrch on the Mediterranean region carried out by the Institut Agronomique Méditerranéen.

The user can extract from the bank the desired sub-sets of data and compose tables as required; he can then request a certain number of processing functions (calculation of averages, sub-totals, percentages, adjustments, statistical calculations etc.).

Identification data

Name: MEDISTAT (Mediterranean countries socio-economic data bank)
Address: Mediterranean Agronomic Institute (I.A.M.)
B.P. 1239
34011 Montpellier Cédex. Tel. 63 28 82
Management: Head of the Mediterranean Agronomic Institute and
Mr M. Allaya.

Description

Field

Macro-economics, national accounts, international trade, agricultural foor production systems.

The geographical area adopted for MEDISTAT is based on the Club of Rome 'Regionalized and hierarchical world model', adopted for the analysis of the Euro-Mediterranean region. Although all countries of the world are included in the model, there are only detailed data for the countries of the Euro-Mediterranean region, other countries being grouped together in broad regions.

The model makes it possible to carry out analyses by country or group of countries for the Mediterranean, Europe and Africa, and by region for the rest of the world.

The chronological series entered in the bank cover the following fields from 1965, including the most up-to-date data for recent years:

- total population and active population by branch or sector;
- overall data on national accounts;
- agricultural production of food by crop;
- external trade in agricultural food products by product and country (imports and exports, by quantity and by value);
- balance of payments;
- production structures.

Nature of service

The purpose of the MEDISTAT data bank is to make available to specialists (research workers, teachers, interested parties in firms or public authorities) a certain amount of statistical data in the socio-economic field on the Euro-Mediterranean region, together with a set of programmes for processing these data for analysis purposes.

The MEDISTAT bank forms part of the research on the Mediterranean region carried out by the Institut Agronomique Méditeranéen.

The user can extract from the bank the desired sub-sets of data and compose tables as required; he can then request a certain number of processing functions (calculation of averages, sub-totals, percentages, fittings, statistical calculations etc.).

Data sources

1) National statistical yearbooks; 2) international statistical

yearbooks - OECD, Eurostat, FAO, UNO; 3) specialized reviews on food technology; 4) mediterranean reviews; 5) research carried out by the IAM.

Acquisition and compilation of data

at IAM punched cards are used to input data from the original tables. The data are first stored on disc and then recorded on tape after they have been checked, sorted and validated.

Data processing

Programmes written in PL1 and in FORTRAN IV are used for tabulation, statistical calculation and classification etc.

The data are processed by means of a simple language generating the structure of the tables to be edited by extraction from the archive files; This language allows the five functions (+,-, * ,** , /) to be used and provides almost unlimited levels of parenthesis. The results can be retrieved and processed using the library programmes (programmes for statistics, classification etc.).

At present the processing is done on an IBM 360 of 512 K memory and 2000 K LCS (Large Core Storage). On average, the programme at present uses 250 K. It should be noted that the larger the edition, the greater the size of memory required. A new model now undergoing tests will allow the memory size required for average output to be reduced to approximately 100 K.

Interrogation of the bank

This data bank is similar to a data base in that it only contains basic data. It differs from the data base because the data are not inter-connected but regrouped according to need by the programme. This makes the processing more flexible for it is possible to produce all the re-groupings required and, for example, to work on the Europe of the six, the nine, the twelve etc. The MEDISTAT files are divided into unconnected sub-sets in accordance with certain criteria which facilitate processing (periods, products,country, flow of products) etc.

At present only batch processing is used; conversational processing on a time sharing network is planned.

Output

Output is simple or structured in the form of tables via the printer, (see table 1, Exemple de sortie), curve plotter, magnetic tape and in some cases punched cards.

Method of operation and cost

Two methods of operation are possible:
- single or repeated interrogation
- participation.

For more information, please contact the IAM.

Table 1. Example of output (MEDISTAT). 6165: average 1961-1965, % 6165: average 1961-1965 as percentage of world total (=100%), A: average 1969-1971, 7175: average 1971-1975, % 7175: average as percentage of world total (=100%), T6373: averaged annual growth during the period 1961/'65-1971/'75.

PRODUCTION
PISB 9 PRIX MARCHE CSTANT <1975> , TX DE CHANGE: 1975
UNITE: DOLLARS US 10**6

	6165	%6165	7175	%7175	T6373
FRANCE	181798.00	5.27	309141.60	5.42	5.44
ITALIE	104698.60	3.03	163450.00	2.86	4.54
PMEM CEE	286496.60	8.31	472591.60	8.28	5.12
ESPAGNE	41874.00	1.21	78390.40	1.37	6.46
GRECE	9281.20	0.26	18762.00	0.32	7.28
MALTE	164.80	0.00	312.20	0.00	6.59
PORTUGAL	7513.60	0.21	13638.00	0.23	6.13
TURQUIE	15279.60	0.44	28592.80	0.50	6.46
YOUGOSLAV	17408.40	0.50	30746.80	0.53	5.84
CHYPRE	558.00	0.01	1049.00	0.01	6.50
PMEN MEMBR	92079.60	2.67	171491.20	3.00	6.40
MEDIT NORD	378576.20	10.98	644082.80	11.29	5.45
ISRAEL	4748.00	0.13	11238.60	0.19	8.98
LIBAN	1964.00	0.05	3572.80	0.06	6.15
EGYPTE	7588.00	0.22	11090.40	0.19	3.86
JORDANIE	829.60	0.02	1156.20	0.02	3.37
SYRIE	1894.20	0.05	3452.20	0.06	6.17
KOWEIT			3370.00A	0.00	*
YEMEN	470.40	0.01	709.00	0.01	4.18
YEMEN SUD	267.00	0.00	211.20	0.00	-2.32
PAYS DU GO	0.00	0.00	0.00	0.00	*
ARAB. SEDU	4520.40	0.13	15836.20	0.27	13.34
IRAN	14020.20	0.40	40360.40	0.70	11.14
IRAK	4107.60	0.11	7585.80	0.13	6.31
MOY ORIENT	40409.40	1.17	95212.80	1.67	8.93
ALGERIE	4820.40	0.13	9111.80	0.15	6.56
MAROC	4543.60	0.13	6890.80	0.12	4.24
TUNISIE	1628.20	0.04	3154.40	0.05	6.82
LIBYE	2166.00	0.06	8358.60	0.14	14.44
MAGHREB	13158.20	0.38	27515.60	0.48	7.64
MEDIT SUD	53567.60	1.55	122728.40	2.15	8.63
MEDITERRAN	432143.80	12.54	766811.20	13.45	5.89
SOUDAN	3189.20	0.09	3836.80	0.06	1.86
MAURITANI	245.20	0.00	448.40	0.00	6.21
SOMALIE	360.60	0.01	514.60	0.00	3.61
PAYS ARABE	38594.40	1.11	75929.20	1.33	6.99
AMERIQ.N.	1108697.00	32.17	1628194.00	28.56	3.91
ETATS-UNI	1023461.00	29.69	1487376.00	26.09	3.80
CANADA	85239.80	2.47	140823.40	2.47	5.14
AMERIQ.L.	170049.80	4.93	301399.60	5.28	5.88
AFRIQUE R	55697.00	1.61	95732.40	1.67	5.55
EUROPE O.E	1036224.40	30.06	1600792.00	28.08	4.44
EUROPE E.R	430655.20	12.49	833070.80	14.61	6.81
U.R.S.S.	430655.20	12.49	543770.80	9.53	2.35
OCEANIE R	59882.40	1.73	96568.40	1.69	4.88
JAPON	186204.40	5.40	456539.00	8.00	9.37
ASIE PLANI	179302.40	5.20	292769.40	5.13	5.01
CHINE R.P.	171682.00	4.98	280597.60	4.92	5.03
INDE	61113.40	1.77	82381.00	1.44	3.02
MONDE TOTA	3446074.00	100.00	5700790.00	100.00	5.15

Products

MEDISTAT has been in operation since January 1977. It has permitted the production of the yearbook of the Mediterranean countries No. 2 which has just been published by the IAM. The production of other yearbooks by group of countries or products is planned.

CRONOS and agricultural statistics

D. Byk, Statistical Office, Commission of the European Communities, Luxemburg

Summary

General description of the base
CRONOS is a management system for a statistics data base entered as time series.

The time series are grouped in subject fields corresponding to major statistical categories (e.g. national accounts, energy, foreign trade). Each field is allotted a storage area known as a file subset. In addition to the various specialized subject fields, CRONOS also has a general priority field intended to supply general and short-term economic information (ICG). This field will be accessible at all times during periods of conversational access to the data base; the other fields, on the other hand, may only be accessible during 'working sessions'.

Certain series of specialized fields (master series) are likewise included in the ICG, where they are automatically updated (subordinate series).

In the conversational mode, it is possible under certain conditions to store series emanating several fields in a temporary storage area and to link up these series in a computation or a tabular or graphic display.

Agricultural statistics in CRONOS
- ZPA1: Animal products. Management of this area is the responsibility of Division D2 of Eurostat 'Agricultural balance sheets and products'.

It comprises monthly and annual series relating to total animal stocks and production of milk, eggs and meat, with effect from 1964.

The material covered by this field is used in Eurostat publications: Monthly statistics for:Meat, Monthly statistics for: Milk, Monthly statistics for: Eggs.
- PACØ: Agricultural prices and accounts. The management of this field is the responsibility of Division D1 of Eurostat 'Agricultural accounts and agrarian structure'.

This field is at present being set up and will comprise monthly and annual series of agricultural purchase and selling prices, indices of purchase and selling prices as well as all the agricultural accounts series.

Presentation of the CRONOS system

Aims

CRONOS is a management system for a data base consisting of statistical data recorded in the form of time series with fixed intervals. This system, which was devised for the departments of the European Communities, aims to facilitate more efficient and more intensive use of the main information available on economic and social trends.

CRONOS makes it possible to create a large-scale data base and provides a wide range of processing facilities which are directly accessible by statisticians, without recourse to professional programming.

The system is designed to provide a rapid response to questions concerning information in the base, whether in the form of raw data, processed data stored and automatically updated in the base, or data computed to meet a specific request.

The system can be used simultaneously by a number of users, who communicate with it either by conventional 'batch processing' or directly via an individual terminal - this is known as 'on-line' access or conversational mode operation.

Data based management system

A data base management system consists of a store of data on a computer medium (magnetic tape or disc) and a library of general programs for the creation, updating or interrogation of the base.

A data base contains information managed (i.e. created, updated and consulted) by a number of independent users. Centralizing the data avoids redundancy and facilitates the exchange of information. A data base accepts information in specific logical structures which the system maps onto physical storage structures with which the user does not need to be concerned.

A management system consists of a set of general programs for the creation, updating and interrogation of the base. These programs, grouped in logical functions corresponding to the usual operations to be carried out on the base, can be run directly by users without the need for programmers. In particular, users may define their data and the processing to be done by means of simple languages.

General description of the base

The CRONOS data base essentially contains time series divided into domains corresponding to major categories of statistics (e.g. economic statistics, national accounts, energy). A storage space, known as an SEF (sub-set of files), is allocated to each domain. An SEF may contain up to 150 000 monthly series or 300 000 annual series.

A time series is identified by a numerical code. It consists of a set of terms or values representing the changes in a phenomenon recorded at fixed intervals (monthly, quarterly, half-yearly or annually). Labels and parameters may also be associated with a series.

There are two types of series:

- primary series, whose terms are supplied directly by the user;
- derived series, which are calculated and updated by the system using computation formulae that are themselves defined by the user in a computing language.

The total capacity of the base is limited only by the resources of the available computer; it is planned to include in the EEC base some 600 000 time series, about 400 000 of which were available at the end of 1976.

General description of the management system

The system contains two main catergories of functions accessible to users:
- management functions modifying the content of the base: these are used to create and update the data base;
- interrogation functions for using the data in the base without modifying its content.

For security reasons, use of the management functions is restricted to authorized users. Use of the interrogation functions may also be restricted in the case of confidential data.

The management functions are used in batch mode. The interrogation functions are available in batch and conversational modes, allowance being made for the characteristics of each access method.

Agricultural statistics in CRONOS

ZPA1 domain

This domain is managed by division D2 of Eurostat, 'Agricultural balance sheets and products'.

General features

The domain designated by the abbreviation ZPA1 includes 5 sets of series:
- monthly hatchery statistics (eggs put in incubation - chicks hatched - external sales of chicks);
- monthly milk statistics (yield of cow's milk and products obtained);
- monthly statistics for slaughter animals and breeding animals (animals slaughtered - gross domestic production - external trade);
- results of surveys on cattle and pig stocks;
- results of structural surveys (cattle rearing farms - pig farms - hatcheries).

Structure of the classification plan

The classification plan of the ZPA1 SEF is divided into 5 main sub-codes (2, 3, 1, 1, 2).
Sub-code 1: identificatuon of various series sets (2 digits);
sub-code 2: product identification (the form of this identification varies from set to set and can be further broken down into 2 or 3 sub-codes) (3 digits);

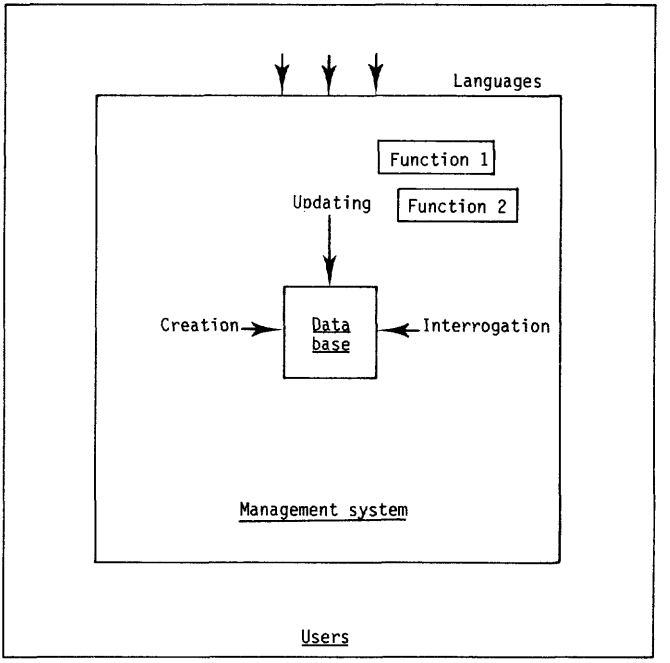


Fig. 1. Data base management system

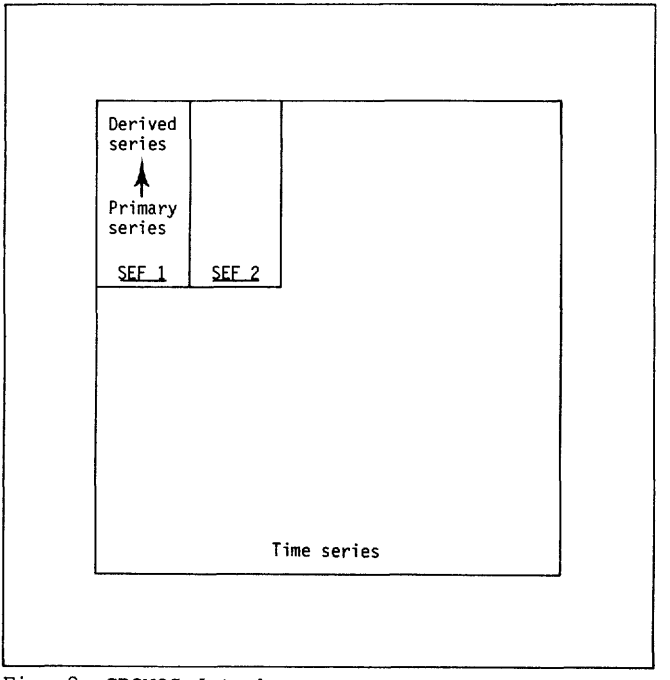


Fig. 2. CRONOS data base

sub-code 3: specification of unit of time (1 digit);
sub-code 4: distinction between absolute figures, cumulative figures and variations (%) from previous year (1 digit);
sub-code 5: country identification (2 digits).

Each series in the domain is identified by a nine-digit code consisting of a combination of these 5 sub-codes.

PACO domain - Agricultural prices and accounts

This domain is managed by Division D1 of Eurostat 'Agricultural Accounts and Agrarian Structure'.

General information on agricultural prices

Statistics on agricultural prices gathered by the SOEC are subdivided into four categories:
- Producer prices and purchase prices: the series in this set are prices in national currency and their equivalents in EUA. The annual series refer either to calendar years or to farming years.
- Index of producer prices: the primary series used for the indices are either series of national indices or price series; methods of compiling aggregates differ considerably from product to product and from country to country. Coverage is different for the monthly series (33 products) and for the annual series (48 products).
- Indices of purchase prices: the situation for calculating aggregates is comparable to that for indices of producer prices. This set is currently being developed.

The codification of agricultural prices

The codification of agricultural prices is as follows:
- sub-code 1: identifies the type of price data: price or index, sales or purchase prices;
- sub-code 2: identifies the product;
- sub-code 3: identifies the trade phase, the degree of elaboration and the periodicity;
- sub-code 4: country sub-code.

The structure of this codification is identical to that of ZPA1.

General features of agricultural accounts

The constitution of this set is at present being investigated. Data on agricultural accounts relate to the compilation of four major aggregates:
- production;
- intermediate consumption;
- value added;
- fixed capital formation.

There is a specific type of questionnaire for each of these aggregates. However, for production there are two types: one for quantities and one for values. Data on production are characterized firstly by a product (e.g. cereals, barley) and secondly by an aggregate (e.g. gross production, final stocks).

Data on intermediate consumption are also classified under 12 products with 9 aggregates. The structure of the other two aggregates is similar. The structure of forestry accounts is of the same type as that of agricultural accounts. This set of series, however, is much smaller.

The following code structure is to be adopted for agricultural accounts:

- sub-code 1: identification of data type;
- sub-code 2: identification of products;
- sub-code 3: identification of aggregate;
- sub-code 4: identification of country.

Conclusion

The agricultural data bases form an important part of the CRONOS system. Current developments mean that it will be easier to make useful comparisons between data from different sources and of different types, thus opening the way to improved coherence in the information system.

System for supplying calculation data in the Federal Republic of Germany

H. Kühner, Kuratorium für Technik und Bauwesen in der Landwirtschaft, Darmstadt, Federal Republic of Germany

Summary

Due to the rapidly progressing development of agriculture the demand for a range of data available in sufficient quantity and quality has increased over the last 20 years. In order to meet these requirements, various establishments in the Federal Republic of Germany engaged in the early sixties in compiling and grouping data and issuing them in terms of catalogues and data collections. The way, however, those data were established would not follow a uniform principle; frequently they were inaccurately or not at all described and thus hardly resisted comparison.

Since the decentralized and mostly sporadic way of proceeding in 'collecting data' could no longer meet the demands, the Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL) (= Farming and Farm Construction Technology Board) was charged in 1971 to set up a working party within which not only a new system of establishing data should be worked out but also revised the one of supplying calculation references standardized on federal level for the overall range of agriculture, fruit and vine-growing, horticulture as well as rural domestic economy.

These problems can only be solved with the use of electronic data processing. Therefore an information system determined and prepared for computer use (dIS) was developed in which the data are stored in such a way that, if desired, any data aggregates whatever may be produced by computation and recalled in the required combination. The innovation consists in the agencies commissioned not providing the KTBL with finished results, hence final values, as previously but is describing the way data are to be calculated in terms of functions. Final products of the data establishment are original data which allow time, natural and cost data to be calculated under any combinable conditions whatever.

The efficiency of the dIS relies on facts such as follows:
- The systematic way of preparing and storing data allows the individual data to be modified at any time with minimum expenditure. Updating is consequently only dependent on the supply of the original data.
- The combination or the origin of the computed information

is reducable because maximum possible details are recorded.
- In principle the same programs allow the production of both standard data with normalized influencing variables and farm and model-specific data. Advisory services and training centres mainly require standard data as currently published for example in the 'KTBL-Taschenbuch für Arbeits- und Betriebswirtschaft' (= KTBL Pocket-book for Labour and Farm Management) or in the 'Datensammlung für die Betriebsplanung in der Landwirtschaft' (= Data Collection for Farm Planning).
The decentralized system of establishing data with the use of standard survey sheets and their central storage ensure the conformity of the results, with the dIS being used as decentralized points.
Data storage and processing are organized in such a way that accuracy and type of output desired in each case require the user a minimum input.

The importance of having up-to-date, well compiled data in a suitable form for situation analysis and decision-making for individual use and for regional and national economic planning is widely recognized. The mounting requirements experienced in recent years as regards the quality and exhaustiveness of data can be met only by a coordinated approach backed up by the required technical and financial resources.

Setting up of the KTBL working party on Calculation references

For the above reasons, and for the general lack of uniformity in data acquisition and supply, the Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL - Farming and Farm Construction Technology Board) was asked in the early 1970s to set up a working party to prepare and systematically update calculation data valid throughout the Federal Republik (i.e. supraregional) and covering all sectors of agriculture, fruit and vine-growing, horticulture and agricultural accounts. Those responsible agreed at the outset that apart from the task of coordinating the work, the methods of compiling and updating the data needed to be revised and rationalised, i.e. it was necessary to use electronic data processing (EDP) and to set up an information system making use of a data bank.

Definition:Data bank based information systems use the latest EDP techniques to store information in the smallest form required (elements) so that any data aggregates may be produced by computation and recalled in the required form.

The working methods of the KTBL working party may be summarized as follows:
- Data are obtained on a decentralized basis from suitable specialist institutions in accordance with agreed rules. Original data suitable for the data bank are sent to the KTBL.
- The KTBL prepares and stores the original data and draws up programs for computing the calculation data.
- The KTBL issues publications containing calculation references e.g. KTBL-Taschenbuch für Arbeits- und Betriebswirtschaft (KTBL Notebook

on Labour and Farm Management); Datensammlung für die Betriebsplanung in der Landwirtschaft (Data on Operational Planning in Agriculture).

New ways of establishing data

In preparing original data for the data bank, data compilers no longer supply the KTBL with finished results, but indicate how to compute the results (the final data) using functional equations. The factors governing each individual target value are therefore examined, together with the mathematical relationships between these factors. An essential feature of this method is that the most complex combinations can be reduced to small, easily identifiable elements or segments.

Thus, for example, the process 'Pflügen' (Ploughing) contains about 50 segments which describe accurately and in a logical sequence the process and the variables and express them as a function of working time. The values (e.g. driving speed, working width, turning and preparation time) which are measured and statistically verified by data compilers represent the content of these functional equations. The variables relating to individual farms (e.g. plot size and shape, distance from field to farmhouse) are determined by means of a questionnaire and included in the program. Thus it is possible to calculate the time rquired for each part of a particular job and also for the entire process.

The simple example given shows the thoroughness with which data must be compiled for the data bank. However, it can also be concluded that with the functional approach a great deal of duplication can be avoided by exchanging elements, and the possibilities for combining various processes are almost limitless. Data must however, be compiled in each sector according to a standard procedure.

Standard processing system

The original data for preparation and storage are coded centrally in accordance with a standard procedure and then fed into the data bank and tested. For this purpose it is necessary to develop systems, i.e. universal service programs dealing with storage, documentation and the data bank contents and computaion of calculation data (see Fig. 1). Explanation of Fig. 1: the functions and variables supplied to the KTBL are stored and updated in data bank stores (dbs) following processing and coding. The 'documentary' programs on the left-hand side of the diagram are used mainly to check the contents of the data bank and as information for data compilers, so that they can apply existing functions in their work. In the case of the 'computational' programs (each function stored must be solved by computation in accordance with the values of the variables) there is a 'universal' and 'subject-orientated' process.

- Universal: e.g. calculation of time required for certain specific processes, taking into account plot size and shape, distance from farmhouse to field, productivity level, amount of manure for

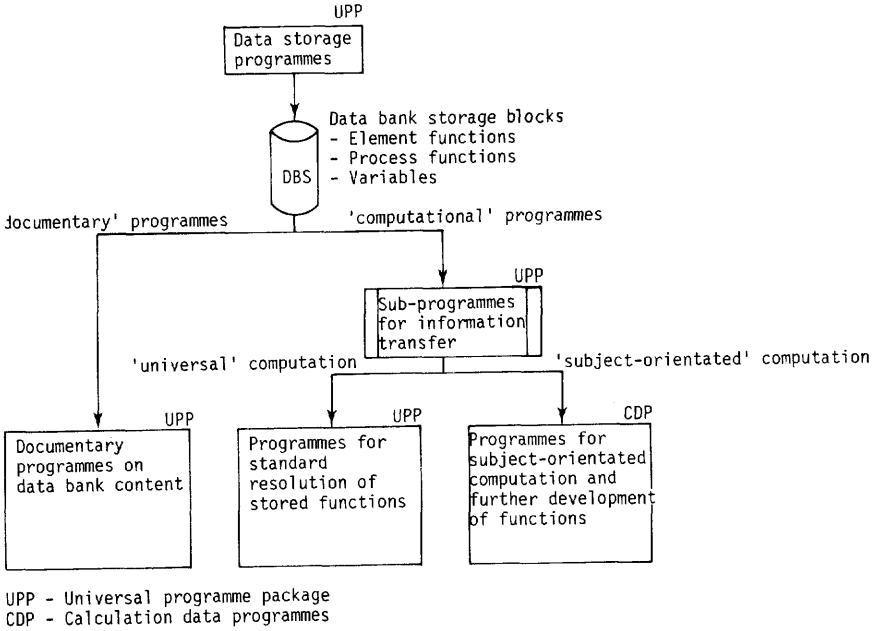


Fig. 1. Uniform processing system.

spreading. These programs are referred to as 'universal' because they can be used to analyse not only data on time requirements but also, for example, nutrient requirements or effectiveness.
- Subject-orientated: e.g. calculation of time requirements of entire sectors of agriculture over certain periods; nutrient requirements of the animal husbandry sector and estimation of effectivenes of farm fodder.

The main difference between the 'universal' and 'subject-orientated' processes is that with the latter process, the output data are more comprehensive, which means that more complex questions can be answered.

The processing system referred to Fig. 1 has the following main advantages:
- the same program can be used to compile standard data, and data relating to specific models, farms or regions. *Standardized data* are published for example, in data collections in which all variable parameters such as plot size or shape, road length, number of animals and feedingstuffs are fixed or standardized. *Date relating to specific farms or models* are used for reckoning calculation data which deviate from the standard parameters in as much as in the same program the value of the standardized or preset parameters are replaced by specific values applying to the farm or model in question.
- The system may be used for all sectors of agriculture and forestry, it may be extended at any time and is easy to keep up-to-date. Changes in the data stored can be carried out very quickly and at

very low cost, since for each change only one piece of data in an element function entered only once in the data bank required to be altered (because there is extremely little redundancy, i.e. data of similar content and value and affected by the same set of variables are found only once in storage). This makes data compilation much easier and helps to avoid duplication.

Possibilities of supplying data

Data output/supply is organized as shown in Fig. 2. Explanation of Fig. 2: the diagram shows the flow of data via the central processing system to the user. Broadly speaking, data are chanelled according to one of the two following methods:
- computation and publication of standard data and rapid information (EDP printouts as a supplement to data collections), via the KTBL;
- computation and transfer of data for regional calculation references and for specific information, via regional computer centres.

The underlying principle of this information flow is that data development and exploitation can be treated separately. The KTBL is responsible for the development and maintenance of the data bank system, for the issuing of standard information and for the transfer of data, user programs and instructions (copies) to regional computer centres. Advisory bodies and research and administration estab-

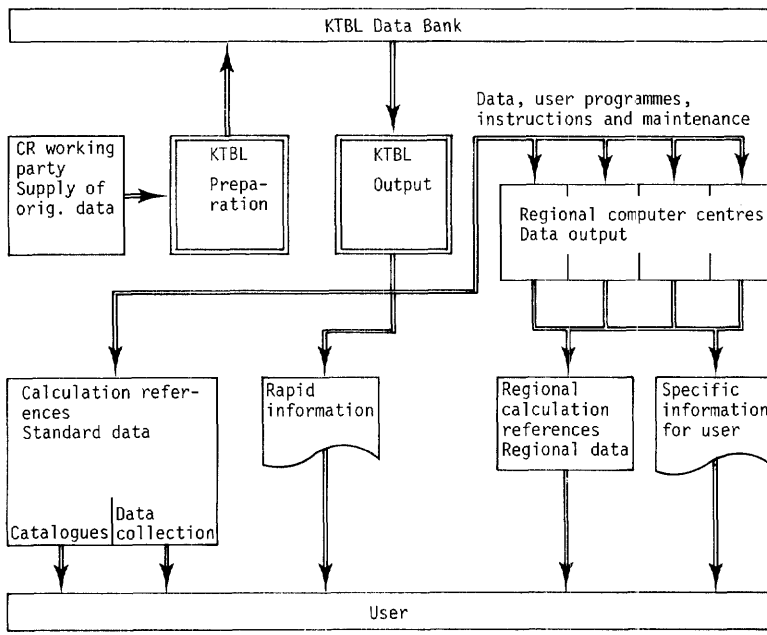


Fig. 2. Data supply via KTBL Data Bank.

lishments in the Federal German Länder use these copies and their regional computer centres to process questions and issue regional data as required. This method ensures that the final information output always refers to the same *uniform data base*. The data bank system can therefore be used to full advantage, while there is still room for individual responsibility.

Summary

The KTBL working party on 'calculation references' was formed to meet the urgent need to improve the supply of data for all purposes of agricultural training, guidance and research. Despite considerable difficulties in the early stages it was possible within a few years to build up a functional information system based on a data bank. Data supply is carried out using a completely new and substantially more efficient method. The system has been kept variable, it may be freely extended and may be interrogated for a wide variety of purposes. It is to be hoped that the work thus put in hand will never have to be abandoned for financial reasons, because the need for data will increase rather than diminish.

References

Auernhammer, H., 1976. Eine integrierte Methode zur Arbeitszeitanalyse. Planzeiterstellung und Modellkalkulation landwirtschaftlicher Arbeiten. KTBL-Schrift 203, Hiltrup.

Brundke, M., P. Jäger & H. Staude, 1974. Arbeitszeitbedarf in der Schweinehaltung. Benutzeranleitung für den Einsatz in der elektronischen Datenverarbeitung. KTBL-Arbeitspapier, Darmstadt.

Egloff, K., 1976. Daten unde Kalkulationsunterlagen. Ausbildung und Beratung 29 (2): 32-36.

Egloff, K. & E. Reisch, 1971. KTBL-Arbeitsgemeinschaft 'Kalkulationsunterlagen': Die Entwicklung und Konzeption der Arbeitsweise. Ausbildung und Beratung 24 (11): 186-187.

Gindele, E.H., 1972. Die Bedeutung agrarstruktureller Elemente für eine rationelle Arbeitserledigung in der Feldwirtschaft. KTBL-Schrift 156, Hultrup.

Hammer, W., 1975. Arbeitszeit- und Beanspruchungsfunktionen. Grundlagen für landwirtschaftliche Planungsdaten. KTBL-Schrift 202, Hiltrup.

Hesselbach, J., 1971. KTBL-Arbeitsgemeinschaft 'Kalkulationsunterlagen':Aufbau und Probleme einer Datenbank für die Betriebsplanung in der Landwirtschaft. Ausbildung und Beratung 24 (11): 187-189

KTBL, 1975. Datensammlung für die Betriebsplanung in der Landwirtschaft. 4th edition.

KTBL, 1975. Datensammlung für die Kalkulation der Kosten und des Arbeitszeitbedarfs im Haushalt.

KTBL, 1976. KTBL-Taschenbuch für Arbeits- und Betriebswirtschaft. Daten für die Betriebskalkulation in der Landwirtschaft. 8th edition.

Rosegger, S. & F.P. Sörgel, 1976. Ermittlung von technischen und arbeitswirtschaftlichen Planungsdaten für die pflanzliche Pro-

duktion. Landbauforschung Völkenrode, Sonderheft 32.
Thamling, C.H., 1975. Produktionsabläufe der Rindermast. Systemanalyse unter betriebsspezifischen Datenannahmen. KTBL-Schrift 199, Hiltrup.

Establishing a technical information centre for forestry

S. Leinert, Kuratorium für Waldarbeit und Forsttechnik, Federal Republic of Germany

Summary

Forestry faces very great difficulties in developing efficient and economic systems of mechanization which do not damage soil and remaining trees. This is the case especially under middle-European conditions characterized by highly intensified and differentiated sylviculture. To clarify this very special situation some typical forest problems are shortly dealt with.

It is planned to establish an information data bank system - called technical information centre for forestry - during the period from 1976 to 1979. This data bank is planned to be at the disposal of forestry in Middle and-Western, partly also in Northern Europe. This system represents technical and economic data and is directed towards the needs of practice. The target groups are forestry and forest science, producer of forest machinery and the timber industry. Structure and function of this system are shortly described.

The project has already been launched second half of 1976. Thus only the actual stand and the so long made experiences can be referred. Finally the project will be judged considering its potential importance and an outlook shall be given on further developments.

Background

The economic situation in the German forestry industry has been deteriorating rapidly since the end of the 1950's. Spiralling labour costs were accompanied by a stagnation, even a decline, in timber prices. Whilst earnings in timber cutting have increased six-fold in the last 20 years, it was only last year that timber prices reached the 1955 level again. It was thought that this extremely difficult situation could be solved by increasing the mechanization of the operations, but this process has encountered much greater difficulties than in agriculture.

For the most part the German forest is composed of a mixed stand in which the individual trees are of very different size and quality. The characteristics of the wood, which influence the processing, also vary very widely for each species. Moreover, since the various trees have different cutting cycles ranging from 80 to 240 years, approximately half the total increment is cut for intermediate yield - so-called 'thinning'. Finally, if one takes into

account that over the centuries the forest has been driven back by agriculture to the worst locations which, by their very nature pose particular difficulties for mechanization, it is immediately apparent that there would be major problems in any mechanization of these operations concerned with the care of the soil and the tree crop. A further disadvantage was that in contrast to agriculture, there was no compact, efficient market, served effectively by strong companies. Instead, the late advent of mechanization in the Central European forest industry led to scattered developments, which were often only of local importance and - with the exception of power saws - in no case involved large numbers. The impressive developments in North America and Scandinavia could not be adopted because of the completely different silvicultural objectives involved.

For these reasons the history of the mechanization of the forest industry during the last 15 years is marred by a complex and often unsuitable range of machinery, high investment costs per unit, high repair costs and a generally high level of risk.

Since the survival of the forest industry and low-cost supplies of raw wood to the timber industry can only be assured by effective mechanization, a three-year project was set up, 'Creation of a Forestry Information Centre', financed jointly by the Federal Government and the Länder, with the task of improving conditions in this field for the forest industry of central Europe.

Description of the system

The Forestry Information Centre is primarily engaged in the compilation, analysis and conversion of the technical and silvicultural information needed for the mechanization of forest-work. The following steps are planned for the development of this information system, which is at present still based on a manual card filing system:

- Firstly, information will be compiled on the technical aids which are of special importance for certain forestry operations. The categories of machines covered are systematically classified according to type, stating the most important operating sequences on which a technical assessment and an assessment of the suitability for forest work can be based.
- At the same time realistic models for the main operating conditions in the forest will be developed to provide information on the most important factors involved in the design and use of machines. The optimum working procedures at different degrees of mechanization will be worked out for these models, which are based on practical experience.
- In the light of practical experience - especially on that of the machine yards - a picture of the operational limits as well as of the performance and cost of the individual machines and systems is to be built up.
- Finally, it is intended to translate the silvicultural parameters obtained from these models into technical criteria for the design of machines. This will enable the engineer to design suitable forest machines, dispensing with the expensive 'trial and error' method used until now.

In principle, it is planned to base the project on manual card index systems. However, since a transfer to a computer-based system is probably essential in the longer term, the indexes will be transferred to a suitable computer system during the second half of the project period.

The appended diagram illustrates the sequence and interrelationship of the various tasks during the project period.

State of work

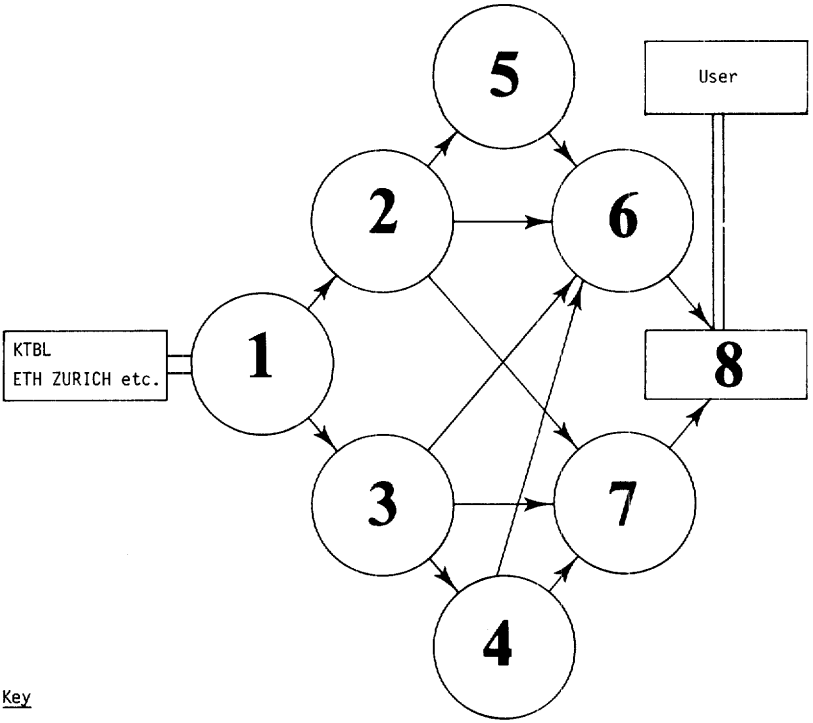
As far as the compilation of technical information is concerned, the following categories of machines have already been fully covered with regard to technical data and method of operation: skidder, forwarder, agricultural tractor, timber bob, winches, tongs, and mulching machines.

On the silvicultural side of the project, the experience gained in practice with these machines was recorded on the spot in 30 forest machine offices or machine yards in the Federal Republic of Germany, In addition, the necessary data were obtained for the construction of realistic stand models for spruce, pine and beech (trees).

To supplement this project, a model of such an information centre is at present being constructed in cooperation with the Institutes for Biometrics and Information Science of the Universities of Freiburg, Göttingen and Munich for the 'INTERFORST 78' trade fair to be held in Munich in May of next year. In addition to the data mentioned above this project, which uses the IBM system STAIRS-DL/I (Storage and Information Retrieval System-Data Language/I) and will work on-line, will provide information on the main timber harvesting processes and how they can suitably be combined and mechanized.

Since, as has already been mentioned, the project is still in its initial phase, we can only report on first experience. Even at this early stage extreme interest has been shown by the forest industry in this project. The help and cooperation of the numerous scientific and professional groups involved in the project are equally good. There have already been many cases in which the investigations enabled expert advice to be given to the industry during investment programs, so it can be expected that if the work continues as planned this project will fulfil its objectives and contribute to the efficient mechanization of the forest industry of Central Europe.

Appendix. Diagram showing the structure of the project.



Key

1	Preparation	
2	Compilation of technical information	(machines)
3	Compilation of silvicultural information	(models)
4	Compilation of silvicultural information	(operating sequences)
5	Processing of technical information	(types)
6	Processing of technical information	(criteria)
7	Processing of silvicultural information	(models)
8	Formalization	(storage, output)

KTBL: Kuratorium für Technik und Bauwesen in der Landwirtschaft (Farming and Farm Construction Technology Board)
ETH ZURICH: Eidgenössische Technische Hochschule (Federal Institute of Technology)

The IDAIC data bank project

E. Romagnoli, Istituto di Diritto Agrario Internazionale e Comparato, Firenze, Italy

Summary

The State intervention in the field of the economical reports and in the agriculture has considerably increased both the number of laws and administrative acts, whose good knowledgment is absolutely necessary for all agricultural businessmen who want to make good and rewarding choices. Consequently a system which will quickly provide precise and concise information about it, will be highly recommended.

For this purpose the Istituto di Diritto Agrario Internazionale e Comparato (IDAIC), which is a public body, has included in its programs the project for a 'Data Bank' as regard to the agricultural legislation.

In particular the IDAIC is planning a data formulation exploiting the following sources:

- agricultural jurisprudence (both of the High Court of the European Community, of the Courts of the countries which are members of the EC and of the Italian Courts but without consideration of decisions of the Supreme Court of Italy);
- agricultural legislation of the EC and of the countries-members of the EC;
- legislation projects of EC and bills actually issued in countries-members of EC;
- researches, both Italian and of other countries-members of EC, dealing in the agricultural doctrine, without consideration, as regard to Italy, of that being published in the various specialized magazines or reviews;
- agricultural customary rights;
- administrative practice.

These data, being added to those actually elaborated in Italy by the Ufficio del Maasimario presso la Suprema Corte di Cassazione (decisions of Supreme Court) and by the Istituto per la documentazione giuridica del Consiglio Nazionale delle Ricerche (studies of agricultural law published in the specialized magazines) will be condensed according to an 'abstracts' criteria, will be ordered and memorized by a computer, and will supply the 'intermediate users' with all the concise and precise information they may need.

The cooperation with the Ufficio del Massimario della Corte di Cassazione, the Ufficio Studi of the Parliament,

the Corte Costituzionale, the European Community, the FAO, together with the Italian and foreign members of the scientific committee of IDAIC, will highly facilitate the search of the sources of information.

The complexity

The complexity of modern methods and the abandonment by the State of the former laissez-faire attitudes, both as regards relations between owners, farmers and workers and as regards production trends, have unavoidably led to an increase in the number of regulations affecting relations between private citizens and government in the administrative control of economic issues.

This growth in the volume of legislation and administrative regulations is more marked in agriculture than in other areas of the economy. Since a good entrepreneur cannot take effective decisions unless he knows what the legal consequences of his action will be (incentives, mere supervision, no effect, active discouragement or prohibition) it seems desirable to offer agricultural entrepreneurs or public organizations fast, accurate and concise information on regulations affecting their work and on the administrative procedures for implementing them.

With this aim in view the Institute of International and Comparative Agricultural Law (Istituto di diritto agrario internazionale e comparato (IDAIC), a public body established in Florence, Piazza d'Azeglio 39, of which I have the honour to be Chairman, has included plans for a data bank on agricultural law in its program. The project is based on existing models which cover all aspects of law and not just one area, but which on the other hand only collect the data produced by one source. I am referring in particular to the records office of the Supreme Court of Cassation (Ufficio del Massimario della Suprema Corte di Cassazione) which electronically processes, compiles and makes available to users all the abstracts of the judgements of that Court, and to legal data bank of the Institute for Legal Documentation of the National Research Council (Consiglio Nazionale delle Ricerche) which electronically processes compiles and makes available to users abstracts of papers appearing in legal journals.

The data bank to be set up by the Institute of International and Comparative Agricultural Law will perform a similar task, and in addition to the data processed by the records office of the Court of Cassation and those processed by the data bank of the Institute of Legal Documentation will include data from the following sources:
- judgements by the Court of Justice of the European Communities on matters concerning agriculture; important decisions in the Member States on matters concerning agriculture; important decisions by Italian courts on matters concerning agriculture;
- Italian legislation on agriculture; regulations and circulars issued by public authorities in Italy on matters concerning agriculture; important measures taken by the European Economic Community with regard to agriculture; important laws, regulations and circulars on matters concerning agriculture in the Member States;
- important draft legislation in Italy and abroad; important

proposals for regulations and directives in the European Community;
- scientific works published independently and not as contribution to journals; foreign papers (from the Member States);
- agricultural customary law;
- administrative practice, new forms of liaison between practical operators not expressly included in existing regulations.

Abstracts based on data from the above sources and classified according to an index carefully drawn up by the leading experts in the field, will be entered in a computer which can then provide the intermediate user with accurate and comprehensive information on any problem of agricultural law, with an indication of the source so that, if he wishes, the user can immediately study the question in greater depth.

For some time the IDAIC has been developing indexes for material relating to agricultural legislation which, since they are intended to provide the framework for a comprehensive information service in this sector,are more detailed than the indexes used for similar, non-sectorial services such as those referred to above.

The 'replies' provided by a computer containing stored data processed and classified on the basis of the criteria mentioned above must satisfy the requirements not only of the intermediate user who wants to have the main elements of the legal data and problems presented in a form which is comprehensible to the layman, but also of the specialist for whom the task of searching for sources could be made considerably easier.

The preparatory work for the management of a data bank on agricultural law could be facilitated by cooperation agreements with the records office of the Court of Cassation and the data bank of the Institute for Legal Documentation, both of which are normally willing to cooperate in the dissemination of legal data.

As regards the other sources of information, the IDAIC would establish contact with the research departments of the Lower and Upper Houses of the Italian Parliament, the Constitutional Court, the European Economic Community, the FAO, with foreign associates and members of the Institute's scientific committee and with the Ministry of Agriculture and Forestry.

The project outlined above would be of interest to the 'intermediate users' of information, such as the provincial agricultural inspectorates, the development boards, the first and second degree cooperatives, the land reclamation syndicate, the land improvement syndicates, trade union and professional organisations in agriculture, the producers' associations, and farmers. Associations and individuals approached so far have shown considerable interest in the project, which they consider will be very useful. As was said before, the project itself will benefit not only the intermediate users but also those whose work is concerned with practical aspects of law, such as magistrates, officials, barristers, notaries and academic lawyers.

IDAIC Data Bank project

This list only concerns Italian agrarian law. Lists will be compiled

from other countries, taking into account relevant differences.
1. Agricultural law: sources, historical sources, agricultural codes, community regulations, national legislation, regional legislation, administrative regulations.
2. Agriculture: agricultural structures, agricultural development, agricultural guidance (and guarantee), forestry, stockbreeding, fish farming, activities connected with agriculture.
3. Land ownership: social role of land ownership, public ownership, collective ownership, joint use, public concerns, private property, latifundi, uncultivated land, land reform, reorganization of the land, land consolidation, land reclamation, land improvement, land cooperatives, forest ownership.
4. Agricultural entrepreneur: entrepreneurs farming as a main occupation, owner-occupier/tenant farmer, joint undertakings, staff working for entrepreneur, cessation annuities.
5. Farmers' cooperatives: productive cooperatives, servicing cooperatives, processing cooperatives.
6. Agricultural contracts: tenancy, leasing to tenant farmers, share farming on a family basis, share farming on an individual basis, profit sharing, livestock leasing, land improvement tenancy, heritable lease.
7. Farms: farm unit, succession on farm, stocks.
8. Agricultural credit: operating loans (production loans, equipment acquisition loans), land improvement loans, agricultural credit guarantees, credit notes, agricultural privileges, arrangement for agricultural credit.
9. Labour in agriculture: farm managers, farm employees, agricultural wage earners, farm hands, pensions, cessation annuities.
10. Agricultural trial: judges specializing in agriculture, judges competent in matters of agriculture, arbitrators, experts.
11. Agricultural administration: Ministry of Agriculture, national-level administration, regions, government agencies.
12. Common agricultural market: agricultural products, interventions, guarantees, target prices, intervention prices, levies, refunds, producers' associations, EAGGF, intervention agencies, subconstractors.
13. Agricultural taxes: land registration, rates, tax on income from agriculture, VAT.
14. Rural welfare.
15. Agricultural associations.

IMAG data service

G. van de Werken, Institute of Agricultural Engineering, Wageningen, the Netherlands

Summary

The IMAG-DATASERVICE consists of a number of data bases and a number of calculation programs about management, in this case labour and economics concerning arable farm, cattle breeding and horticulture. The data bases are built up as matrices with activities in the rows and specifications in the columns. These data bases have been put as data files into the memory of a computer to be available for electronical working up. Computer programs are developed to help in decision making. Several of these programs have access to the data files thus using them as information sources.

The data of the files have been obtained by special studies by collecting data of former investigations and by using Predetermined Motion Time Systems. Till now some files are in use as complete filled up files, others are incomplete or are in preparation. Examples of files are: data of field work activities, transport and loading/unloading activities, activities in glasshouses.

About fifteen programs are available in the IMAG-DATASERVICE. Most of these programs use data from the data files. These programs give quantitative information on a wide scale of problems. This scale runs from calculating a tasktime of a single operation to a whole Management Information System in horticulture.

The data bases are in use as:
- archives;
- background to provide computerprograms of information.

The data bases and much more the computerprograms are used to become quantitative information at planning, comparison of work methods and machines. Users are:
- farmers and extension service;
- manufacturers and dealers of machines;

research workers;
schools.

The computer which is used at the moment is a DEC 10 installed in the Computer Centre of the Agricultural University at Wageningen. The installation has a time sharing and batch system.

Introduction

During 1972 the Institute of Agricultural Engineering at Wageningen started a data service consisting of two data bases and four computer programs, which used data from the bases. This system is called IMAG-DATASERVICE.

In the meantime the data service grew to a data bank (data bases) and a calculation service (programs). Programs and data bases were developed on farm management concerning arable farming, cattle breeding and horticulture. Several research workers in the institute developed programs in their specialities.

Data files

After the Second World War, many time studies were made on all kinds of activities in agriculture. Besides these time studies Predetermined Motion Time Systems were developed and are available now. From these sources, completed with special studies, data bases have been made. The data bases have been stored in files in the memory of a computer so that the data can be used by programs.

The data files are built up as matrices with activities in the rows and specifications in the columns. All activities are coded in the first element of the rows. The meaning of the codes is to be found in the coding lists. A number of data files has been completed, other files are so far incomplete and some files will be set up in the near future.

The files are updated from time to time when new activities have to be added or when some elements change their value, e.g. due to the progress of technology in machines. Files are accessible as direct access files or as sequential files.

In Appendix 1 a list of files, operational and in preparation, is given, and the size when complete is mentioned or estimated.

The contents of the files are very different, dependent on the subject. As an example, the file of fieldwork, consists at this moment of 194 rows of fieldwork activities and 31 columns of specification.

In the columns are to be found a number of element times, factors for rest and disturbance, a complex of 8 characters for the working route on the plot and the type of the machine, preparation time, factors concerning the annual costs of depreciation and upkeep and interest.

Programs

At the moment 17 programs are available. Most use one or more data files as additional input besides actual input provided by the user. The intention of the programs is to calculate quantitative information to help the farmer in decision making.

The working field of the program is very wide. A number of programs calculates data for partial problems like task times for milking and all kinds of field work, matching of harvest systems, etc. Furthermore programs for more complex problems are available e.g. Farm Labour Planning system, calculating with classes of workability

and the Management Information System in Horticulture.

These programs use parts of other programs. This is possible while a number of programs has been organized as a main program with subroutines. In the main program input/output is organized and in the subroutines the calculation procedure is handled.

A part of the programs is written as a conversational program, other programs at batch versions, and some have both possibilities. Input is given according to a real situation on a farm, alternatives or fictive situation. In each program the number of input data given by the user is held as low as possible in relation to the accuracy of description of the situation and the exactness of calculation. The data to be provided by the user are mostly technical data. To help using the programs user-guides (in Dutch) are available in which input and output are explained. Appendix 2 contains a list of programs, operational and in preparation. The programs are written in FORTRAN IV.

Application

The IMAG-DATASERVICE has been set up to help farm managers in decision making on different levels of complexity: a single activity, harvest of a crop, or the whole farm organization.

Specific farm situations are approximated and problems formulated. A solution can be obtained stepwise or by optimization. Besides background information for the programs the data files are also used as archives. They are the source of information about elements of all kinds of activities and can be given by the computer. Complete data bases will be published too. The IMAG-DATASERVICE is firstly set up for Dutch agriculture, so that the language used is Dutch. But the system is also applicable in other parts of the world; of course other activities would be added.

The users consist of the following groups:
- farmers and extension service;
- manufacturers and dealers of machines;
- research workers;
- schools.

Farmers can consult their regional extension service. Specialists can help formulating questions and choosing suitable solution methods i.c. programs. Other people can contact the IMAG-DATASERVICE directly. Schools with a terminal connection to the same Computer Centre can run the programs directly.

To test and introduce the possibilities of the IMAG-DATASERVICE seminars and workshops have been organized and are still going on with farmers, members of the extension service and lecturers of schools.

Costs

Per program costs are fixed. They consist of initial costs and an amount per run.

Computer

The IMAG-DATASERVICE is operating in the computer of the Agricultural University at Wageningen. The computer is of the type DEC system 10 with time sharing and batch systems. Some technical data of the computer are:
- central memory 256 K words;
- 4 disc drives;
- magnetic tape units;
- card reader;
- paper tape puncher and reader;
- plotter and cassette reader;
- line printer and graphic terminal
- 40 asynchrone telephone lines for terminals.

Input/Output

The IMAG-DATASERVICE is using this computer via terminals and by means of cards and papertape input. The way of program input is related to the type of the program. Conversational input comes via the terminal directly from the user. Another way is via the punched cards, papertape making and input file. This is also possible by typing a file via the terminal. Short output is printed out by the terminal, more extensive output comes via the lineprinter of the Computer Centre. Output width variates from 72 characters to 132 characters per line.

Conclusions

The IMAG-DATASERVICE is running five years. Calculations were made to solve special farm problems, to do investigations and to make catalogues. The number of data bases and programs grew. This number is still growing; several data bases and programs are developed and tested.

In a time of increasing difficulties in farm management it is necessary to have suitable tools to come to a good decision.

References

Kroeze, G.H., 1975. Arbeidsbegroten per computer. IMAG Publ. 32, IMAG, Wageningen.

Kroeze, G.H. & R.K. Oving, 1973. Chains of operations, their organization and costs. Research Report 1, ILR, Wageningen.

Oving, R.K., 1971. Farming task and machinery. Research Report 3. ILR, Wageningen.

Taaktijden voor de Landbouw deel I (veldwerkzaamheden), 1970. Verantwoording van de toegepaste rekenmethoden. ILR, Wageningen.

Taaktijden voor de Landbouw deel II (stalwerkzaamheden), 1973. Uitgangspunten en verantwoording van de toegepaste rekenmethoden. ILR, Wageningen.

Werken, G. van de & M.M. de Lint. Taaktijden voor bewerkingsketens bij veldwerkzaamheden. IMAG, Wageningen (in preparation).

Appendix 1. List of data files

	Name of the file	Dimensions (rows x variables)
1	Field work activities	194 x 31
2	Transporting and (un)loading related to fieldwork	146 x 20
3	Cattle breeding activities	150 x 29
4	Transporting and (un)loading related to cattle breeding	140 x 15
5	Data about activities: - period of performance) - sensibility to the weather) 100 x 45 - type of distribution)	
6	Distribution tables over the year of activities	50 x 27
7	Workability data of fieldwork	20 x 24
8	Time standards of activities in amenity horticulture	?
9	Horticultural activities	5000 x 27
10	Transporting in horticulture I	750 x 17
11	Transporting in horticulture II	750 x 18
12	Production data in horticulture	1000 x 27
13	Data about prices etc.	500 x 27
14	Elemental data of constructing farm buildings	2000 x 7

Remark: The dimensions of the file concern the situation when the files have been completed.

Appendix 2. List of programs

IMAG 20: Task times for cattle breeding
IMAG 55: Task times for field work activities
IMAG 57: Task times for chains of operations of fieldwork
IMAG 83: Matching chains of operations of fieldwork
IMAG 04: Calculation of costs of chains of operations of fieldwork
IMAG 30: Task times for milking
IMAG 31: Labour demand of milking during the year
IMAG 32: Optimizing storage and spreading of animal manure
IMAG 35: Determination of optimal replacing moment of machines and calculation of costs
IMAG 33: Farming task and machinery
IMAG 34: Calculation of labour income on an arable farm
IMAG 40: Farm Labour Planning System
IMAG 80: Comparison of hand picking and mechanical cutting of mushrooms
IMAG 82: Determining optimal picking moment of mushrooms
IMAG 81: Economical comparison of working methods in horticulture
IMAG 83: Task times in horticulture
IMAG 88: Management Informations System in horticulture
---- Information System for costs of constructing agricultural buildings

Bavarian agricultural information system (BALIS)

J. Haimerl, Bayerisches Staatsministerium für Ernährung, Landwirtschaft und Forsten, Bonn-Duisdorf, Federal Republic of Germany

Summary

BALIS is a logical concept for organization and evaluation of agricultural data. It is realized as a system of databases and a software system to store and manipulate those data. By the time it consists of 40 physical IMS- databases. Various databases are concatenated by pointers or have secondary indices.

BALIS is a database system, covering the area of the agricultural administration of Bavaria. There are data concerning the 'primary-statistic' and data concerning 'statistic of agricultural administration'. The data are stored in different aggregated form at BALIS-databases.

The 'Agrarleitplan', as a special planning system of the 'Bayerisches Landesplanungsgesetz', is structured in two concatenated databases. The different subsidies and grants of agriculture does yield a subsystem of BALIS. The data of economic situation of farmers together with their bookkeeping data also results in a subsystem, either areas of existing databases or separate databases. The architecture and planning of agricultural buildings is another subsystem (ISBAU) of 16 physical databases. For the plant production (soil analysing , etc.) are structured databases or parts of existing databases. The animal production makes up a further subsystem of BALIS. A separate subsystem is being built for the monthly market reports.

The documentation of the scientific agricultural literature is organized and structured in STAIRS/DLI.

For the processing and evaluation of data, stored under BALIS, a software system was developed which does enable the enduser to define, to modify and to start his data processing at the terminal. This will be realized by a communicationmonitor and a data- and programdocumentation system.

Introduction

BALIS is a logical concept for the organization and evaluation of agricultural data using data processing methods. BALIS consists of two components which are responsible, 1) for data-organization and data-storage (DORIS), 2) for information retrieval and data evaluation (LINGUA).

DORIS currently covers approximately 40 physical data banks plus

further sequentially structured datasets. The DORIS-data banks use IMS as basic software, the other datasets, which are short or medium termed, are VSAM (virtual sequential access method) organized.

LINGUA is a teleprocessing oriented command and control-system, which enables the user to define his problems and execute them using a screen as input/output device. Once defined the problems (Jobs) may be saved and thereafter changed at any time. The user take advantage of the following capabilities: special command-language and a complete programming- and data documentation system. Existing programs are stored as functions in a modul library. Security-measures provide in each case the necessary degree of security of access.

DORIS is divided into several logical subsystems (see Scheme 1), which may be independent data banks or parts of data banks.

This contribution describes those subsystems which have a formated structure data structure; it does not consider agricultural documentation, which was the subject of an earlier symposium.

Agricultural structure/agricultural planning subsystem

The main aspects are:
- the field of joint primary and specialized statistics;
- the agricultural guideline planning as a specialized planning feature based on the Bavarian Agricultural Planning and Development Act of 8 August 1974. This is a specialized plan within the context of the Bavarian law on regional planning and development.

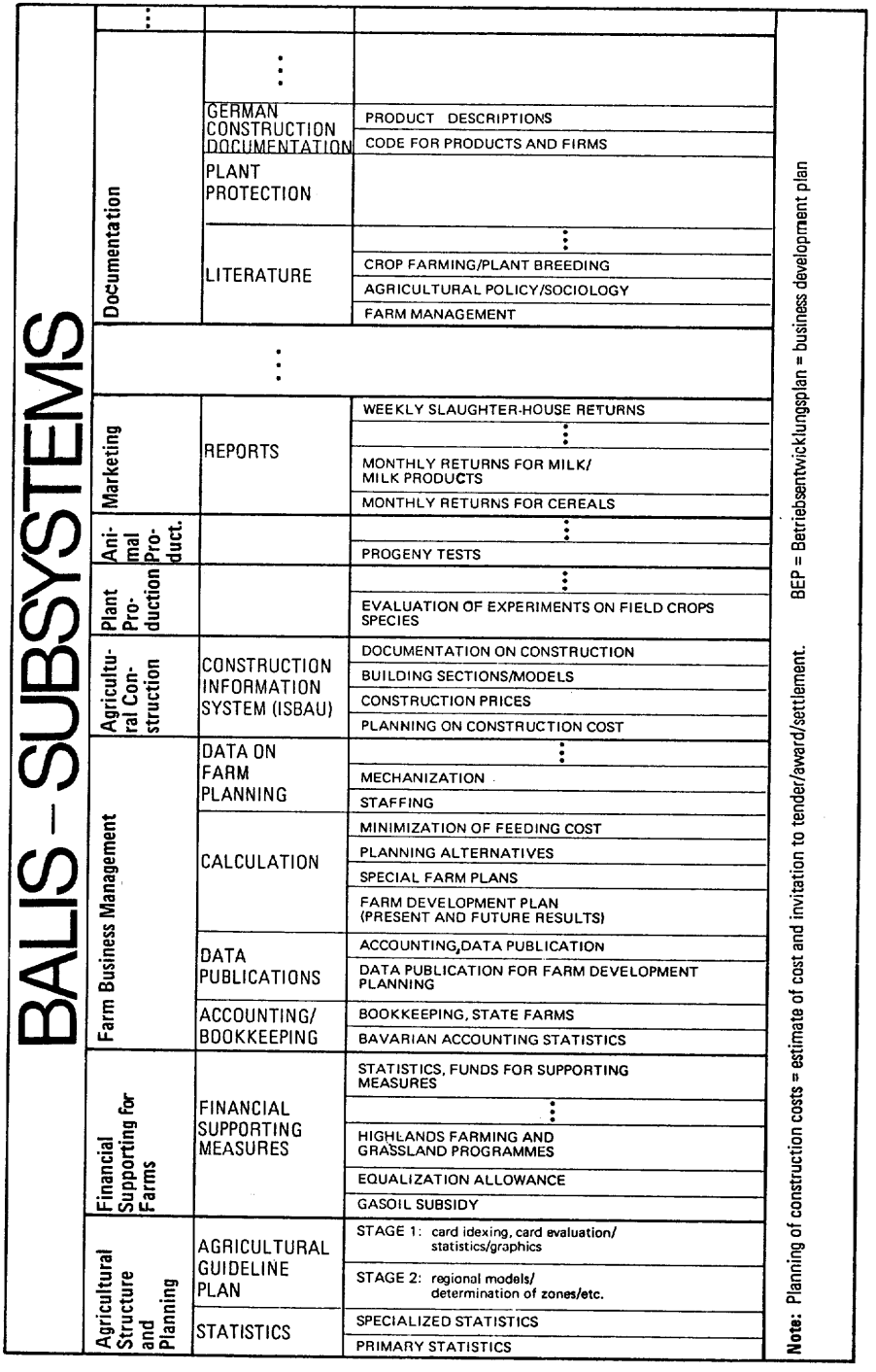
The statistical data are collected in a special statistical data bank. A statistical evaluation system makes it possible to:
- determine areas to be evaluated;
- build up tables;
- print tables;
- output data;
- define graphics;
- plot graphics.

Data for the agricultural guideline plan are stored in two logically connected data banks. The agricultural guideline plan consists of two steps:
- recording and evaluating available areas of land suitable for agricultural purposes;
- working out of objectives-conception and statements, in particular with regard to the use of these areas.

The aims of the agricultural guideline plan are:
- to guarantee food supplies;
- to maintain, cultivate and organize arable land;
- to safeguard agricultural interests in the context of planning;
- to evaluate areas under different production conditions;
- to secure suitable land for agricultural purposes;
- to work out objectives for areas not suitable to agriculture;
- to establish foundations for purpose-oriented regional developments.

In short, to elaborate guidelines for a well thcught-out and logical development of rural areas.

Scheme 1. BALIS subsystems.



'Financial supporting' subsystem

Data concerning the financial support of individual farms are stored as part of the farm data bank. Data for supporting market structure form part of another data bank. A systematically standardized data section was worked out for all supporting measures in the allowance sheet.

Every support measure has a unique code number for the type of support. The financial resources for supporting purposes are provided with a code, the first position denoting the support group - for example subsidy, public loan etc. Each decision to award a grant also has a mandatory basic data section. Data specific to support measures may be recorded if necessary.

'Farm management/auditing' subsystem

A basic accounting entry for every farm provided with accounting statistics is stored in the farm data bank. Data on transactions and characteristic values calculated from these data are stored in a segment type which is hierarchically dependent of the principal entry. The Datasheets 'Farm development plan' and 'Accounting' are stored as segments of the farm data bank. Calculations for the organization of agricultural farms, the minimizing of costs for individual production processes (for example feeding) and the optimization of agricultural models in the context of regional planning and development are done using linear optimization methods. Standard data (for example standard matrices) and data relating to specific problems (input and output data) are stored in a special data bank system. Up to five alternative organizational patterns may be calculated for each individual farm (up to 10 combinations for the minimization of food costs). A table generating system makes it possible to present the calculated results in the desired form.

The results of the actual organization patterns and the selected target organization patterns of optimization for supporting individual investment in agriculture and forestry are stored in the farm data bank. Farm planning data concerning staff, mechanization etc. are not yet stored.

'Agricultural construction' subsystem

The agricultural construction information system 'ISBAU" consists of 17 physical data banks. The basis is the standard service manual of the GAEB (Gemeinsamer Ausschuss für Elektronik in Bauwesen) - Joint Committee on Electronics in Building Engineering). The catalogue of achievements which are required in this context are issued accordingly. These catalogues produced by data processing methods are sent in the course of a write out to tender to the tenderers. Received offers containing price quotations are stored and checked. Thus the most advantageous offer is obtained. A special Item-Catalogue is compiled for each user. The priced items are applied in accordance with a pre-determined schedule for the determination of regionalized and average prices per item broken down into time categories. From the individual prices per item structural units/

models may be quoted in conjunction with priced items for planning of building costs (cost estimate).

'Plant production' subsystem (ISPFLANZ)

Data in this subsystem come from:
- The land survey (part of the farm data bank): In addition to special basic data on the farmer making the request a detailed field report and laboratory analyses of soil samples (approximately 50 000 per year) are collected;
- quality control (special data bank): surveys of harvested products and laboratory analyses are carried out for this purpose;
- The Bavarian survey on the different varieties of field crops (special data bank). Results are collected and evaluated according to various mathodological considerations.
- Crops file (part of the farm data bank): For selected cultivated areas all activities from preparation of the seed bed to the harvest including all expenditure and yield recordings are collected and evaluated.
- Meteorological and climatic reports from meteorological stations (special data bank) - still at the planning stage.

'Animal production' subsystem

Several data banks produce data for the animal production subsystem. These are data recorded from various yields, for example milk yield, records of the number of animals slaughtered and progeny tests. The farmers mutual aid organizations such as producer groups (for example milk producers, beef and pork producers and piglet breeders etc.) record and manage these data, which are stored at the Bavarian Ministry of State for Food, Agriculture and Forestry. It is therefore possible to use the software available at the Ministry.

'Market structure- reports' subsystem

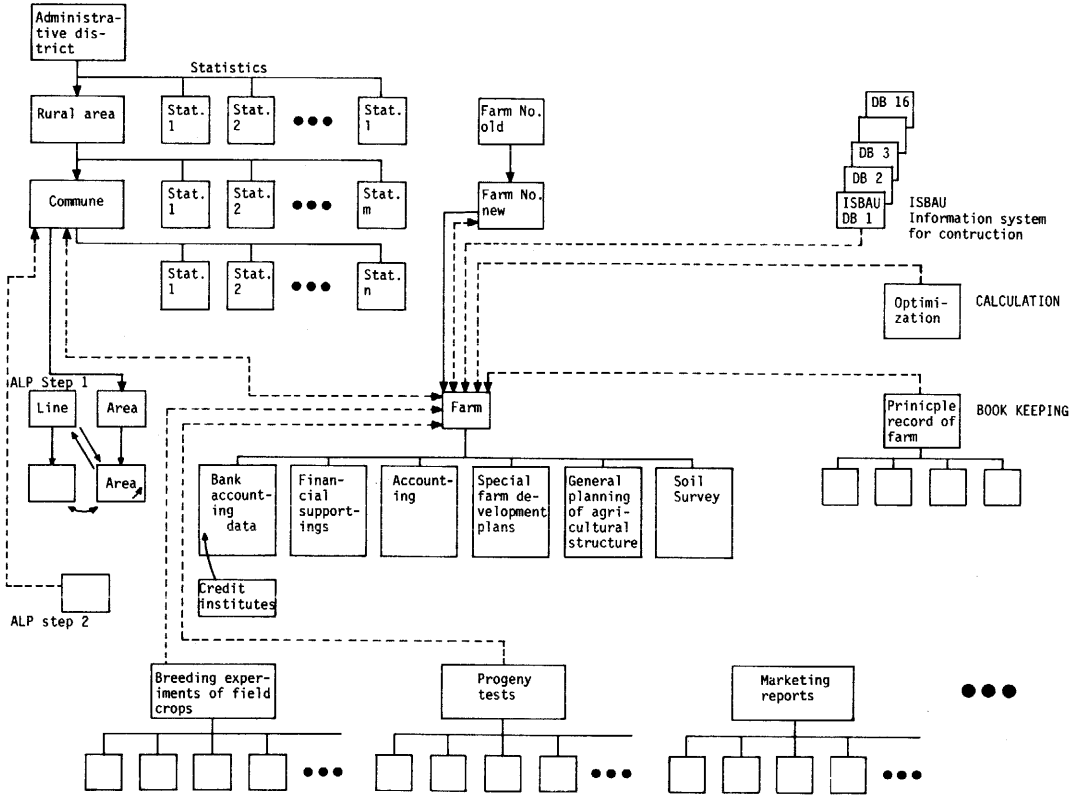
Administrative marketing data are collected in a market structure data bank. These include:
- data to support the market structure;
- the monthly informations for the food and agriculture report (in accordance with the notifications to be made by warehouses and diaries)
. on cereals and cereal products (trends in level of stocks);
. for milk and milk products (trends in level of stocks);
- weekly deliveries of livestock to Bavarian slaughter houses.

The monthly information for the food and agriculture report and the weekly trends in livestock supplies and price structure at Bavarian slaughter houses are drawn up.

Final observations

This report can give only an overview of BALIS by setting out the individual subsystems. It was not possible to consider in detail many methodological interesting problems. Considerations on the

Fig. 1. Simplified structure of the BALIS Data Banks - excluding VSAM data sets (January 1976).



technical aspects of data banks involved in the retrieval and evaluation of the stored data were by-passed. There will, however, bu further opportunities to consider some aspects of these problems in the discussion.

Discussions

Discussions of the sessions

The Symposium dealt with three areas in agriculture: animal production and food technology, plant production and agricultural management information systems. The discussions that took place in the separate sessions involved principles common to all three areas and for this reason they are summarized together. As a preliminary to the discussions, it was agreed that the definition of the term 'Data Bank' might be taken for granted but at the same time it should be understood that a Data Bank should be a permanent system, deriving its data from several sources and managed in such a way as to be put to many different applications according to the wishes of the user.

The following factors featured prominently in the discussion:

- need for Data Banks;
- cost of setting up and maintenance;
- rights of Data Bank owners and users;
- validity and updating of contents of Data Banks;
- interpretation of contents of Data Banks;
- links and exchanges between Data Banks;
- role of government and industry in use and support of Data Banks;
- training of experts and users of Data Banks.

These are dealt with below but are not allocated any priority.

Need for Data Banks

Many factors prompt the need to set up a Data Bank. Examples were given where the need for organized data systems arose from a clear administrative or scientific problem. In other cases the Data Bank was established as an academic exercise which subsequently expanded and proved to be of wide interest. Sometimes data accumulated within an organisation until it reached an extent and complexity which handicapped its use without classification, organisation and computerisation. Ultimately, the need is justified by the extent to which the Data Bank is used and it was stressed by most delegations that the purpose of the Data Bank is to meet the need of users of data and not of the data processors. It also became apparent that in many if not most cases, Data Banks will fulfil needs that were not possible to identify at the outset and that the justification for continued existence may change with the passing of time.

Cost of setting up and maintenance

In most of the systems described hardware and programming expertise was available within the organisation. Discussion of costs therefore centred about system and software development and the staff required to manage the Data Bank. It appeared that the number of staff directly involved was small, possibly only one and rarely more than three or four. However back-up requirements might be large depending upon circumstances, and in some cases working groups or advisory committees have been established. Development times of between one year and six years were cited. It was generally agreed that most parent organisations failed to recognise the importance of Data Banks thereby making development a slow process and future existence unsettled. Generally information on the costs of the Data Banks described was sparse. In most organisations those engaged in Data Bank work are also working in other areas and it is difficult to make appropriate financial assessments. Sometimes (e.g. Agrodata described by M. Allaya) the initial costs are borne by the parent institute (in this case the Institut Mediterannéen-CHEAM) but subsequently some income is derived by sales of publications based upon the Data Bank contents. However it may be said that no instances of a truly commercial approach to meeting costs were cited.

Rights of Data Bank owners and users

The problem of the rights of access to data was raised but no clear conclusions were reached. Commonly the providers of data have no control over the use to which their data is put once it is put into the system, and this may be a factor that contributes to the apparent reluctance to provide data that some workers have encountered. For instance only 50 out of 500 requests for information from M.Allaya were answered. Some questions inevitably raise problems of confidentiality and may be of strategic importance to private industry. Nevertheless there is a considerable volume of information within both governments and industry which could be profitably shared on a national and even a world basis.

In some areas, for instance Data Banks on milk production, soil science, animal disease, where the system is set up primarily for the benefit of the agricultural community, this problem is only of minor concern. However the other areas were cited (e.g. food contamination) where the Data Bank owner would be reluctant to allow access by outside persons or organisations because of the danger of misuse of the data by those without background knowledge or necessary expertise. There was however no intention to withhold the information but to provide it only together with expert advice on its interpretation (See: Interpretation of contents of Data Banks).

Validity and updating of contents of Data Banks

It was pointed out that there is no standard of the value of data, its authenticity or validity. In one instance (Dr Haendler, INFIC) it was practice at one time to eliminate doubtful data. Currently all data is put into the system but some is tagged for certain purposes e.g. because its validity is in doubt or because it is atypical.

In another case (D.D. Singer - Food Contaminants) expert selection of data will be made before insertion into the Data Bank. The conclusion that must be reached is that it must be left to the system management to decide what approach is best in individual cases. Nevertheless it is felt that the system user must be made aware of what the practice is in particular cases.

A similar variety of approaches exist with regard to updating of the data, although in some respects it is clear that much depends upon the sources of data. In some Data Banks (e.g. AGRODATA) a great part of the data is available on an annual basis although previous data remains valid since it refers to previous years. Similar considerations apply to MEDISTAT and CHRONOS. In others, data does not arise at specific intervals in time and therefore is available more or less continuously. Such data may make previous records invalid or out of date and here again different methods are applied. The practice employed by BALIS is to remove old data to an archive from whence it can be retrieved if required; users are expected to indicate the period of their interest.

Interpretation of contents of Data Banks

The problem of interpretation of data is associated closely with most of the other factors discussed. It is clear that where access is direct and unrestricted there can be no influence exerted by the data bank owners on the interpretation of data and the use to which it can be put. This is of less concern if the data has been previously published. An opposite extreme is data derived from work specially commissioned e.g. by governments, but of undoubted public interest. This type of data involves contentious subjects and data bank-owners may feel that no data should be imparted without expert comment. Many of the data banks discussed were however concerned with specialist areas where such considerations do not apply.

Links and exchanges between Data Banks

This topic was discussed at greater length than any other. From the discussions it was possible to distinguish three types of interaction between Data Banks according to the relative nature of their respective scientific or technical fields, i.e. between:
- data Banks dealing with the same or similar data within the same or similar scientific or technical field;
- data Banks dealing with different but related fields within the same discipline. Examples are Data Banks concerning different aspects of cattle rearing such as artificial insemination, breeding, and milk yield;
- data Banks dealing with distantly related or apparently unrelated fields but where links might be necessary to facilitate research or establish a trend by comparing data. An example is the possible dependency of food composition or contamination upon soil composition or fertilizer treatment.

The optimum level, link or exchange between Data Banks is different according to which of these three situations prevail. Various levels were mentioned:

- awareness of the existence of other Data Banks with correspondence as occasion might demand;
- exchange of data, not necessarily on a two-way basis;
- exchange of data on a regular basis for insertion into both Data Banks;
- inter-terminal arrangement.

It was clear that all delegates were aware of the advantages of links but very few links existed among those Data Banks represented at the present time. Those that did exist were mostly of the simpler kinds and no terminal arrangements were cited. The difficulties that need to be overcome in order to promote greater co-operation between Data Banks were discussed in some detail and emerged as:

- incompatibility of hardware;
- use of different software;
- use of different terminology and classifications making data incompatible;
- different approaches to editing (validity, confidentiality etc.) making data incompatible;
- cost;
- unavailability of skilled workers.

It was agreed it is unrealistic to expect different workers to standardize equipment but that differences in hardware could generally be circumvented. Software differences might be dealt with by exchanges of software or by developing suitable exchange formats. However before this can be done it will be necessary to agree terminology and develop glossaries. This will be a costly endeavour and involve many workers. The cost and labour could be high and not always justifiable; few of the organisation represented admitted to being in a position to meet these extra demands.

In spite of these arguments a view emerged that as an essential first step to promoting greater co-operation between Data Banks the following action might be taken:

- the compilation of a list of existing Data Banks with details of operation, staffing, hardware, software etc. and the state of development as described by Tomlinson (1970) UNESCO/IGO;
- the development in various fields of standard technology;
- the examination of ways and means to provide finance and personel.

In view of complexity of the existing situation delegates expressed strongly that it is premature to envisage at the present time networks of Data Banks. Instead, encouragement should be given by the Commission and by Governments of Member States to setting up bilateral links which would require the minimum in finance and manpowers.

Role of government and industry in support of Data Banks

Although instances were quoted where the costs of individual data banks were partially met by sales of publications or charges for access to data, in general organisations were dependent upon the financial arrangements made by their parent organisations. Support is more often than not barely adequate and does not permit development. It was felt that industry and government should be made more

aware of the importance of well established Data Banks and the necessity of providing them with a firm foundation. The major oil companies already give some support to agriculture data banks in which they are interested. Links between Data Banks are costly to set up and this would perhaps be the major difficulty encountered in achieving the agreed aims. Future developments should take heed of the costs involved at all stages.

Training of experts and users of data banks

It was acknowledged that training of experts and users is desirable for several reasons. Training of experts would facilitate greater efficiency in the running of individual data banks and enable links between data banks to be more easily set up. Training of users would demonstrate the usefulness of Data Banks, increase their use and make possible feedback from users to data banks organisers, thus leading to improved Data Banks. Most delegates agreed that such training should be given priority in future programs.

Group discussions

Group 1. Animal production and food technology

The Data Banks represented covered the fields of:
- animal disease;
- milk production;
- cattle registration and artificial insemination;
- feedstuff information and commercial information on feedstuff companies;
- food composition and contamination;
- food, physical properties (projected)

Only one, (INFIC, feedingstuffs) is international although the project COST data bank of the physical properties of foods will also be so. It was observed that only two delegates at the conference are concerned with food technology as against purely agricultural matters. The Group came to the following conclusions:
- With a view to the establishment of a clearing house for information about Data Banks, the Community should build up a classified directory of Data Banks based upon a clear definition of the term 'Data Bank' and a classification of different types of Data Banks.
- The establishment of compatible methodologies and multilingual terminologies to facilitate the interpretation and interchange of data between different Data Banks is essential.
- The desirability and feasibility of establishing links between Data Banks which deal with subjects transcending local, national or narrow sectional interests should be investigated.
- The community should ensure the continued existence of those Data Banks of proven value to the Community.
- Since several Data Banks use scientific literature as a source of data, the Community should establish contact with relevant abstracting services with a view to their providing input for certain data banks.
- The Community should emphasize potential users' needs in order to

facilitate and/or promote Data Bank development and maintanence.
- The Community should promote access to European and International Data Banks by developing countries.
- The establishment of Data Banks containing integrated data on farm animals e.g. milk production, milk composition, feeding, breeding, fertility, disease etc., should be encouraged within the Community.
-Since animal disease is not limited by national boundaries and since information on the spread of animal disease is not easily collected, a clearing house for relevant information should be set up. After processing, the data would be presented in documentary form as suggested in contract A04 and eventually in machine readable form. (it was realized that this subject was under discussion in the VETEC group but the desirability of this concept was endorsed).
- Consideration should be given to the establishment of an information service on national and international legislation on agriculture and food products, additives and contaminants, with special reference to import/export.

Group 2. Plant production

The data banks represented covered:
- soil sciences;
- genetics;
- cereal production.
It was noted that this was far from representative of the Data Banks existing within the Group's area and in particular the following subjects were not covered:
- pesticides;
- hydrology;
- forestry;
- tropical crops.
Within this subject there is some co-operation and links at international level e.g. in animal husbandry and genetics respectively.

International co-operation is projected within the RITOS network the ISDEF exchange format (Belgium, Canada and France) and the FAO international soil Data Bank project released the previous month. In Italy there are links between data banks concerning fruit trees and cereal production.

It was noted that many Data Banks exist outside the official organisations within the Member States. Major industrial organisations have data banks of their own as well as making use of publicly owned data banks and it was regretted that representatives of the larger agricultural business concerns could not be at the Symposium.
The group came to the following conclusions:
- There is a need for a full survey of existing agricultural data banks in the Community. The survey should include the major agricultural companies and cover both factual data and the availability of know-how. It should be based upon clear definitions of terms such as data bank, system, package etc. and include details of software, hardware as well as actual factual content. The state of development of the data bank should be recorded according to the IGU/UNESCO recommendations.

- The Commission might stimulate activities by setting practical goals and favouring a step by step approach. Existing Data Banks of wide interest should be developed rather than new ones opened although the Commission could open a new Data Bank in an area not yet served if a clear need could be demonstrated.
- The Commission might set up a 'Referral Centre' of Data Banks. The first stage might be a multilingual users guide to Data Banks, a directory of their addresses, an indication of their contents and the name of the person in charge.
- In order to exchange data high-level personnel will be required to consider terminology. Standardization of input and processing (especially software) may be desirable.
- There is a need to recognise the independence of existing national Data Banks. Full co-operation between them will require trained personnel to act as experts. Development of these Data Banks would benefit from Community support and the Commission should be asked to investigate the possibility of such support for on-going activities. This might include training of personnel who will be dealing with enquiries from Community users.
- The content of Data Banks should include details of the origin of the data and the way it has been collected. Careful investigation should be made of the legal and economic (commercial) implications necessitating confidentiality.
- Before procedures for co-operation can be laid down it is desirable to discuss priorities with users. Cost will be an important factor.
- When information is available (from the proposed survey) the Commission should convene a Conference of owners and users of Data Banks to consider what future developments are necessary and to agree priorities for future actions.

Group 3. Agricultural Management Information System (AMIS)

The Data Banks represented covered the following types of data:
- macroeconomic;
- microeconomic;
- information and documentation;
- operational research;
- socio-economic.

Numerous Data Banks are in existence in this field in the Member States. Those described at the Conference are predominantly regional in nature, only MEDISTAT and IDAIC having international data connections. There is some promise of international co-operation concerning design via direct contact between those concerned (e.g. documentation KTBL Data Bank). International use of these Data Banks is made difficult by the regional nature of the stored data, problems of software and hardware compatibility and the use of different terminology.

The group came to the following conclusions:
- The dialogue on AMIS Data Banks initiated at this symposium should be continued.
- A list of AMIS Data Banks should be drawn up including those in operation or in the planning stage.

- Contacts concerning Data Banks should be established between the Community and governments of the Member States bearing in mind costs and time factors.
- There are no plans for setting up joint Data Banks, even if joint documentation of AMIS Data Banks is desirable.
- At present additional close co-operation within the Community concerning AMIS Data Banks does not seem essential. For this reason it is not possible to lay down priorities for co-operation.

List of participants

Mr M. Allaya
Institut Agronomique Méditeranéen, CIHEAM - B.P. 1239, 3191 Route de Mende, 34011 Montpellier CEDEX, France

Prof. S. Andersen
Den Kgl. Veterinaer-og Landbohøjskole Landbrugets Plantekultur, Thorvaldsensvej 40, 1871 København V, Denmark

Dr F. Ascanti
Ministerio dell'Agricoltura e delle Foreste, DGPA Div. VII, Via XX Settembre, 0100 Roma, Italy

Mr. C.C. Balch
National Institute for Research in Dairying, Shinfield, Reading RG2 9AT, United Kingdom

Prof. Bimbenet
Ecole Nationale Supérieure des Industries Agricoles et Alimentaires, 91305 Massy (Essonne), France

Mr. H. Buntrock
Directorate-General Scientific and Technical Information and Information Management, Commission of the European Communities, Bâtiment Jean Monnet, Kirchberg, Luxemburg

Mr. D. Byk
Statistical Office, Commission of the European Communities, Bâtiment Jean Monnet, Kirchberg, Luxemburg

Mr Caruso
Directorate-General Agriculture, 84 rue de la Loi, 1049 Brussels, Belgium

Prof. S. Cooney
The Agricultural Institute, Headquarters, 19, Sandymount Avenue, Dublin, Ireland

Prof. D. Dagnelie
Service des Statistiques, Faculté des Sciences Agronomiques de l'Etat, 5800 Gembloux, Belgium

Dr. A. d'Ambrosio
Ministero Agricultura e Foreste, DGPA Div. VII, Via XX Settembre, 0100 Roma, Italy

Mr. G. Davies
Central Veterinary Laboratory, Ministry of Agriculture, Fisheries and Food, New Haw, Weybridge, Surrey, United Kingdom

Mr. F. Delcour
Centre d'Etude des Sols Forestiers de la Haute Belgique IRSIA, Laboratoire de Science du Sol, Faculté des Sciences Agronomique de l'Etat, 5800 Gembloux, Belgium

Dr Deschner
Bundesministerium für Ernäherung, Landwirtschaft und Forsten, Rochusstrasse 1, 5300 Bonn-Duisdorf, Federal Republic of Germany

Mr Dieudonne
Réseau de Documentation en Economie agricole (RESEDA), Ministère de l'Agriculture, 78 rue de Varenne, 75700 Paris, France

Mr Fischbach
18 route d'Esch, Luxembourg

Mr Gay
Directorate-General Agriculture, Commission of the European Communities, 84 rue de la Loi, 1049 Brussels, Belgium

Mr R. Haas
Directorate-General Scientific and Technical Information and Information Management, Commission of the European Communities, Bâtiment Jean Monnet, Kirchberg, Luxemburg

Dr H. Haendler
Dokumentationsstelle der Universität Hohenheim, Stuttgart, Federal Republic of Germany

Dr J. Haimerl
Bayerisches Staatsministerium für Ernährung, Landwirtschaft und Forsten, Bonn-Duisdorf, Federal Republic of Germany

Mr R. Kayl
Chargé d'Etudes au Service d'Economie Rurale, 113-115 rue de Hollerich, Luxemburg

Dr W.P. Kirchner
Bundesforschungsanstalt für Fischerei, Palmaille 9, 2000 Hamburg 50, Federal Republic of Germany

Mr H. Kuehner
Kuratorium für Technik und Bauwesen in der Landwirtschaft, Postfach 120142, 6100 Darmstadt 12, Federal Republic of Germany

Drs Krolis
IMAG, Mansholtlaan 10-12, Wageningen, the Netherlands

Mr J. Leger
Chargé de Mission, Sous-Direction des Affaires européennes et multilatérales, Ministère de d'Agriculture et du Développement rural, 78 rue de Varenne, 75700 Paris, France

Mr J.-P. Legros
Not present, paper presented by R. van den Driessche

Dr S. Leinert
Not present, paper presented by L. Seidewitz

Dr L. Lloyd
Ministry of Agriculture, Fisheries and Food, Salmon and Freshwater Fisheries Laboratory, Whitehall Place (East), London SE1A 2HH, United Kingdom

Mr Lorigny
Sous-Directeur des Industries Agricoles, Service Central d'Enquêtes et d'Etudes Statistiques, Ministère de l'Agriculture, 4 Avenue de Saint-Mandé, 75012 Paris, France

Mr E. Mann
Editor, International Food Information Service (IFIS), Lane End House, Shinfield, Reading, RG2 9BB, United Kingdom

Mr Mildon
Directorate-General Agriculture, Commission of the European Communities, 84 rue de la Loi, 1049 Brussels, Belgium

Dr E. Muller
Zentralstelle für Agrardokumentation und -information, 53 Bonn-Bad Godesberg, Heerstrasse 110, Federal Republic of Germany

Mr Onidi
Directorate-General Agriculture, Commission of the European Communities, 84 rue de la Loi, 1049 Brussels, Belgium

Miss L. Porges
Directorate-General Scientific and Technical Information and Information Management, Commission of the European Communities, Bâtiment Jean Monnet, Kirchberg, Luxemburg

Mr L. Rechaussat
I.E.M.V.T., Service Documentation, 10 rue Pierre Curie, 94700 Maisons-Alfort, France

Ir F.H. Rijsdijk
Landbouwhogeschool, Afdeling Phytopathologie, Binnenhaven 8, Wageningen, the Netherlands

Prof. E. Romagnoli
Istituto di Diritto Agrario Internationale e Comparato, Piazzo Massimo d'Azeglio 39, 50121 Firenze, Italy

Mr. A. Rutgers
Centrum voor Landbouwpublikaties en Landbouwdocumentatie (Pudoc), Prinses Marijkeweg 17, Wageningen, the Netherlands

Dr ir J. Schelling
Netherlands Soil Survey Institute, Staring Building, Marijkeweg 11, Wageningen, the Netherlands

Mr R. Schiel
Directorate-General Scientific and Technical Information and Information Management, Commission of the European Communities, Bâtiment Jean Monnet, Kirchberg, Luxemburg

Dr U. Schützsack
Institut für Dokumentationswesen, 5 Herriotstrasse, 6 Frankfurt/ Main, Federal Republic of Germany

Mr L. Seidewitz
Institut für Pflanzenbau und Saatgutforschung, Braunschweig-Völkenrode, Federal Republic of Germany

Mr D.D. Singer
Ministry of Agriculture, Fisheries and Food, 65 Roomney Street, London SW1P 3RD, United Kingdom

Dr R.D. Smith
Royal Botanic Gardens, Kew, United Kingdom

Mr Strickland
Ministry of Agriculture, Fisheries and Food, Hatching Green, Harpenden, Herts. AL5 2 BD, United Kingdom

Prof. Thomsen
Økonomisk Institut, Thorvaldsensvej 40, 1871 København V, Denmark

Ing. G. van de Werken
IMAG - Instituut voor Mechanisatie, Arbeid en Gebouwen, Mansholtlaan 10-12, Wageningen, the Netherlands

Mr R. Van den Driessche
Office de la Recherche Scientifique et Technique Outre-Mer, Services Scientifiques Centraux, 70-74 route d'Aulnay, 93140 Bondy, France

Mr J.H. Verheijen
GIR - Stichting Gemeenschappelijke Informatieverwerking voor de Rundveehouderij, Eusebiusbuitensingel 38, Arnhem, the Netherlands

Miss Waibel
Directorate-General Agriculture, Commission of the European Communities, 84 rue de la Loi, 1049 Brussels, Belgium

Dr R. Webster
Agricultural Research Council, Weed Research Organization, Soil Survey of England and Wales, Bagbroke Hill, Yarnton, Oxford, United Kingdom

Mr R. Weydert
Attaché économique au Service d'Economie rurale, 113-115 rue de Hollerich, Luxemburg

Mr P. Willeberg
Royal Veterinary and Agricultural University, Bülowsvej 13, 1870 København V, Denmark

ISBN 90 220 0674 3