Werk

Titel: Managing and marketing of urban development and urban life
Untertitel: proceedings of the IGU-Commission on "Urban Development and Urban Life", Berlin, August 15 to 20, 1994
Jahr: 1994
Kollektion: fid.geo
Signatur: XX
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN1030505985
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN1030505985
OPAC: http://opac.sub.uni-goettingen.de/DB=1/PPN?PPN=1030505985

LOG Id: LOG_0143
LOG Titel: DUVA - a concept for a metadata driven statistical production and information system
LOG Typ: article

Übergeordnetes Werk

Werk Id: PPN1030494754
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN1030494754
OPAC: http://opac.sub.uni-goettingen.de/DB=1/PPN?PPN=1030494754

Contact

Niedersächsische Staats- und Universitätsbibliothek Göttingen
Georg-August-Universität Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen
Germany
Email: gdz@sub.uni-goettingen.de

Abhandlungen - Anthropogeographie
Institut für Geographische Wissenschaften, FU Berlin
Band 52, Berlin 1994

DUVA - A CONCEPT FOR A METADATA DRIVEN STATISTICAL PRODUCTION AND INFORMATION SYSTEM

Norbert Kopp
Statistical Office of Berlin,Germany

The DUVA-concept is an integrated base, which contains the whole process control for statistics production from the survey about plausibility, storing, confidentiality, editing and dissemination.

Every step in the process of production will be standardized independently of a specific statistic. With that it is possible to realize in the production independent modules for every step. About definition of the different interfaces, this concept can be developed step by step and in a division of labour.

For storing data there are two models. Our input interface to the system are standardized basic files, which can contain micro data or aggregated data. The second model are macro files with precompressed aggregated data. Because both data models are standardized it is possible to use a macro data generator.

All needed descriptions for basic and macro files like structure of records, variables, categories of attributes are stored in a metadata information system. All available information in that system will be used in an automatic way for further steps in the process of production.

Transformations in the statistical production process are realized on the basic of rules. Therefore the metadata information system knows different types of rules (reference tables, algorithms, lists of conditions, hierarchies) which can be used in an isolated or combined way.

For classifications the DUVA-concept has an own key concept. It guarantees, that it will be possible to combine horizontal or vertical different hierarchical levels from one or more classifications. About a system of reference tables you can combine them always in the same way.

User interface for the DUVA-system is a thesaurus. About that thesaurus the user can communicate with the system in his natural language. In the thesaurus there is no redundancy. The problem of homonyms, which normally appears in thesauruses, are solved, since the words are self-explanatory thanks to the contextual approach to the references. All words are found in that context they were used in the system to describe a given situation.

In DUVA there is an approach to solve the statistical confidentiality on the level of basic files. With our program SAFE (standardized anonymization functions for individual data), which is realized as a prototype, all known problems like primary and secondary confidentiality and also the problem of dominance can be solved.

DUVA contains many export interfaces to other systems, for example to SPSS, SAS, PC-AXIS, MS-Excel and so on. DUVA is conceptualized as an open system so that it is possible to integrate more standardized products in this system.

The Aims of a Statistical Information System

1. The superior aim of a statistical information system (of official statistics) should be to satisfy the statistical needs of the economy and the society in the best way possible. For that, it would be necessary to adjust the statistical information system consistently to user needs.

However, such user orientation can be made operational only indirectly, since the needs for statistical information are not constant, but undergo permanent changes, depending on the processes of social and economic changes. In addition, the users of statistics represent a broad range of sometimes quite different needs, so that it is practically impossible to adjust the planning involved in the conception of a statistical information system accordingly.

2. In order to be able to further pursue the first aim mentioned, however, without depending on changing needs of the various users, it is necessary to pursue a second aim, namely to make the entire data stock of official statistics the basis on which the different wishes of users are satisfied.

However, taking into account the manifold possibilities of evaluation, such data-oriented approach is possible only on the basis of micro data. This inevitably raises problems of data protection and of ensuring the statistical secret.

3. Thus, a third aim has to be mentioned, i.e., while implementing this data-oriented approach, it is necessary to strictly observe the statistical secret. The assurance given to respondents, that their data will be used for statistical purposes only, is not only prescribed by law, but also a necessary condition for them to accept the collection of statistics. If the respondents obliged to present information are not ready to supply truthful data, official statistics would hardly be in a position to provide the public with reliable information. It would be counterproductive if the data-oriented approach endangered the observance of the principle of scientific neutrality, objectiveness and neutrality regarding the collection and presentation of statistical results.

However, to keep the statistical secret and at the same time to try to satisfy all the wishes of users regarding the extent, the form and the required quality of data, is too costly and too time-consuming. So far, because anonymization efforts have been taken as a rule with regard to concrete evaluations and because tailor-made solutions are more expensive than general statistical publications, this often collides with the capabilities of statistical offices in terms of capacities available.

4. Thus, with a view to cost-benefit aspects it is necessary to look for possibilities to satisfy individual user requirements in an economical way. An economical means to achieve this aim is the multiple use of the data once collected, based on consistent standardisation of the whole stock of data and of the instruments used, since the value of statistical information is not lost in the course of utilisation, but, as a rule, even increases, when this information is used repeatedly. Consistent standardisation of the whole data stock and of the instruments to be used would make it possible to achieve this aim.

However, almost all statistical techniques have developed over the time and are characterised by a broad range of incompatible applications and organisational forms. Thus, in view of the abundance and complexity of tasks to be fulfilled by a statistical office, together with the capacity problems already existing and constraints by concrete deadlines to be observed, it seems to be almost hopeless to achieve standardisation.

5. That leads to a fifth aim: it is necessary for the statistical information system also to support statistical offices in coping with the traditional tasks of evaluation by providing means for their rationalisation in order to ease the burden on current statistical production and to set free the resources required for standardisation purposes. This should not be mixed up with the automation of traditional work processes, since solutions, once they have been developed, can be transferred to other applications only after appropriate standardisation. Such independence from data and programs is the only means to achieve that different work fields can profit from each other.

However, it will be seen quickly that this cannot be done without complete and standardised descriptive information. But so far such metadata have neither been standardised, nor been elaborated in a form generally suited for further processing.

6. Thus, the advantages of repeated uses, which are achieved thanks to such independence from data and programs, will not become effective, unless we strive for a standardised approach to the storage of metadata existing in the most different forms in data files, registers or even just in the minds of the employed staff. In view of the capacities available, this can only be done successfully, if metadata are recorded without redundancy and, if possible, at the very moment, when they are created, and if they can automatically be used later on in the statistical production process.

However, this requires looking at the entire statistical production process in its totality. But in view of the extent and the complexity of the range of tasks a statistical office is confronted with, there is a lack of transparency preventing even insiders from seeing all the structures and interrelationships of the various elements in the entire statistical production process.

7. Thus, in order to reduce the complexity of the overall system, an attempt should be made to elaborate roughly outlined structures of a statistical production process and to define the interfaces between its various components. The different components can only be refined as isolated elements, after this has been done.

Such project naturally raises scepticism, since, on the one hand, roughly outlined structures are too abstract for many critics to accept them as a strategic approach to possible improvements. On the other hand, concrete steps which might rapidly lead to visible success risk to produce failures in development, such as the deficits of presently existing statistical information systems.

8. For that reason the total concept to be developed should be set up so that a frame is created which makes it possible to start implementation in sub fields immediately. Thus, on the principle of trial and error it is then possible to develop prototypes suited on the one hand to collect practical experience as quickly as possible, and, on the other hand, to better assess and simultaneously to concretize the total concept and to protect it from possible future deficiencies.

As this work has to be done in parallel with current statistical production, which, however, in view of the problems of time and capacity, which exist anyway, must not be burdened additionally, such research-induced approach will never be free of

friction in a statistical authority. On the other hand, one has to recognise that in view of the many interdependencies and the lack of transparency such work, if it is only done outside the statistical offices, without their active participation, can hardly have good chances of success.

9. For that reason it is particularly important that in statistical offices it should be possible to pursue an open systems approach, which allows to approach these aims gradually, perhaps restricted to several sub fields only, but at the same time ensures co-operation, excludes duplication of work and guarantees that partial steps, which are at first implemented in isolation from each other, are compatible and that elements which are unknown at a given moment or have to be newly developed can be integrated into the system later on without difficulty.

An orientation towards hardware and software standards is not helpful here in view of the methodological problems involved. In order to solve, in particular, the semantic problems, it is necessary to standardise the contents of the subject matter, which, in our opinion, can be achieved if a thesaurus is implemented as a user interface providing access to all metadata.

As no experience has so far been collected in official statistics with the use of thesauruses and as, basically, only linguists and librarians have traditionally dealt with this field of work, the said efforts might increase the risk that in addition to already existing communication problems in the field of information experts, data processing specialists, statisticians, lawyers, miscellaneous users etc. with their different technical terms, new difficulties for mutual understanding will directly be programmed.

10. In spite of these problems it is absolutely necessary to establish co-operation, because without interaction between offices and institutions and, above all, without interdisciplinary interaction it will not be possible for us to cope with the tasks involved in the implementation of a metainformation system. This can only be done on the road of dialogue, of learning jointly, of reaching agreement on an absolutely voluntary basis.

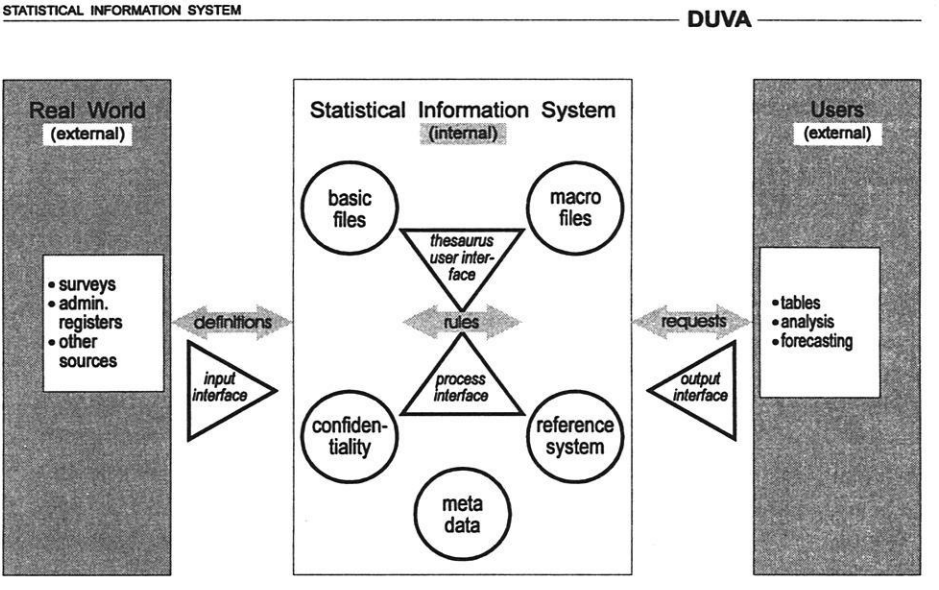
One should not be too optimistic in assessing the readiness of many to cooperate in such project. Mostly, there will be no co-operation, because the various interests are too different. It is indeed easier to stick to the currently applied methods of work, despite obvious deficits in the structure and operation of the present statistical information systems. In addition, most systems were developed in respect of concrete hardware and software solutions so that a new approach will be rejected as long as the new techniques available will not abolish all the deficits immediately and completely.

However, in view of the positive experience we gathered in the co-operation project of large cities in Germany, we are optimistic and believe that co-operation is possible and can be successful.

Elements and Interfaces of a Metadata Driven Statistical Information System

On this background and taking into account the profile of requirements described above (Figure 1), which should be met by a statistical information system, we should like to present below the major components which we found out as well as the possibilities of subject-related standardisation.

Figure 1 Outline of a Metadata Driven Statistical Information System
Definition of Standards and Interfaces



1. All plausibilised micro data from statistical surveys, automated administrative registers and statistically relevant results from other sources are potential data inputs for the statistical information system.

Theoretically, data input might begin with the survey itself, even with the planning for a survey. However, since for many years statistical offices have dealt with a broad range of very different statistics, partly involving highly complicated techniques, such approach would without necessity complicate the setting-up of a metadata-driven statistical information system or even render it impossible.

2. The system's input interface is the data model for those basic files, during the formation of which data collected and already plausibilised are stored in a standardised form and described in the metadata system, with regard to populations, objects, attributes and variables, categories, units of measurement and additional information.

3. All access operations in the system are made via a thesaurus (user interface), so that it is possible to proceed ahead step by step, to avoid redundancies, to recognise interrelations immediately and to integrate new words at any time without difficulty.

4. Relationships of any kind between objects, attributes and categories in the various basic files are established at this input interface by separately maintained classifications, references and reference tables, rather than by the basic files themselves. The semantic and delimitation problems, which are sometimes very difficult, have already been solved by others in advance and, thus, need not be regarded as an additional burden on the input process.

5. In order to ensure statistical confidentiality, all basic files are duplicated and anonymized in a standardised form. Thus, a data base, which is accessible by standard instruments, is available for evaluation purposes and, nevertheless, the occurrence of confidentiality cases is excluded. The implementation of this module is of utmost importance for efficient performance of the total system.

6. Evaluations as such, including the derivation of new attributes, are done in a standardised way, using the rules of the metainformation system. Data model No. 2 includes macro files which can be further processed flexibly in interchangeable formats.

7. All metadata regularly generated in the production process are stored automatically. Cumulatively they lead to permanent extension of the metainformation system and extend (and improve), with each further evaluation, the knowledge base for future uses.

8. Transparency for all elements of the statistical information system is provided by a registration or reference system, in which all components of the statistical information system are documented, which helps to avoid duplication of work and supports economical fulfilment of orders.

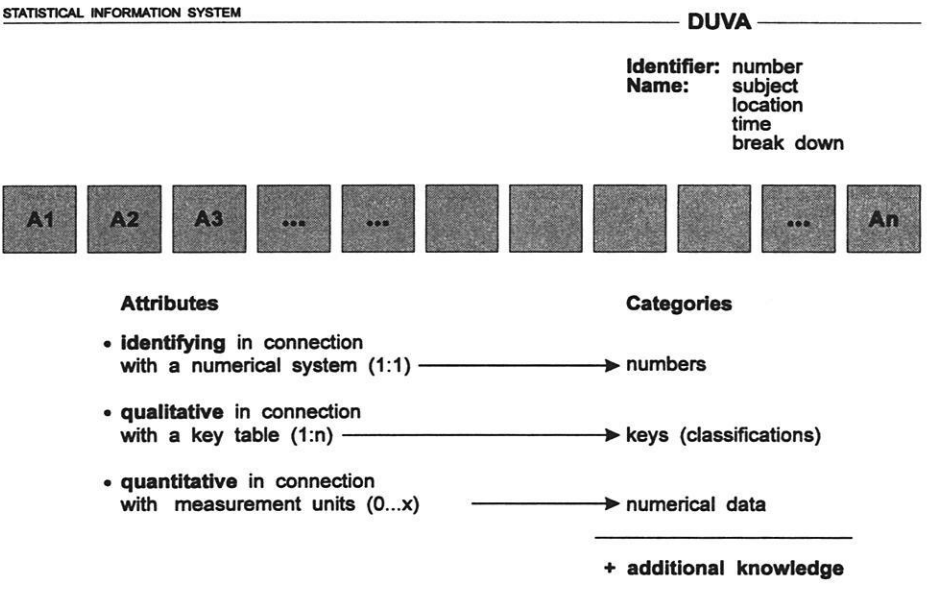
9. Thanks to the thesaurus, all queries for data can be drafted in natural language and be used for the collection of additional use-related metadata with the help of an electronic order processing system (output interface).

10. The electronic order processing system makes it possible to further develop and optimise the statistical information system permanently, in accordance with user needs. On the one hand, the user needs not grope his way through the specifications of the system, on the other hand, thanks to the integration of requirement-induced metadata the system is gradually and automatically adjusted by the users themselves according to their specific data needs.

How this can concretely be done shall be described in more detail below, by specifying the modules - basic files, statistical confidentiality, macro files, reference system and metainformation system - and the aspects of standardisation will be examined more closely by describing again 10 topics for each individual complex.

The data model for data input is a flat file used for the storage of plausible micro data (Figure 2). We believe that, if the following standards are observed, it will be possible to store all initial data of the statistical information system in a standardised form.

Figure 2 Data Model I: Basic Files



The standards for generating basic files are:

1. All data of a basic file must originate from one single survey or data source.
2. All data of a basic file have to belong to a population of objects unambiguously defined in terms of subject, space and time.
3. All attributes of a basic file have to be stored per each object in a data record of fixed length, arranged in an identical order.
4. All data of a basic file per each object have to be complete and free of redundancy.
5. All data of a basic file per each object have to have the same rank of plausibility.
6. All records of a basic file have to contain an unambiguous identification mark (record identifier) in the first field.
7. All qualitative attributes of a basic file have to be coded identically for identical categories.
8. All quantitative variables in a basic file have to use identical units for their value data.
9. All information required for the generation and utilisation of the basic file has to be described in the metainformation system.
10. All additional information, which is relevant for later interpretation of the basic file, must be recorded and stored separately as a non-formalised, free text.

Since the various statistical surveys in statistical offices cannot be seen in isolation, but should be regarded as components of a total system of mutually harmonised and interlaced surveys (full surveys, sector statistics, samples and the data from administrative registers), it is very important for the integrity of the statistical information system that it is possible to link all those different statistics to each other, which have to be stored in separate basic files.

This requirement is met, for example, by key tables. Reference tables are drawn up in the metainformation system, which are used, for example, to assign items within classifications, to define relationships of superiority and inferiority or to combine different classifications.

Figure 3 shows schematically how these reference tables permit that in the case of qualitative attributes the corresponding data in a basic file can be substituted without difficulty by corresponding other keys, so that it is possible to implement any sort of combination or grouping on a one-to-one basis.

There is no doubt that the generation of basic files is not an easy task. But it should be seen that the work required is done by those who are professionally engaged in statistics and so are suited best to solve the substantial problems that may arise.

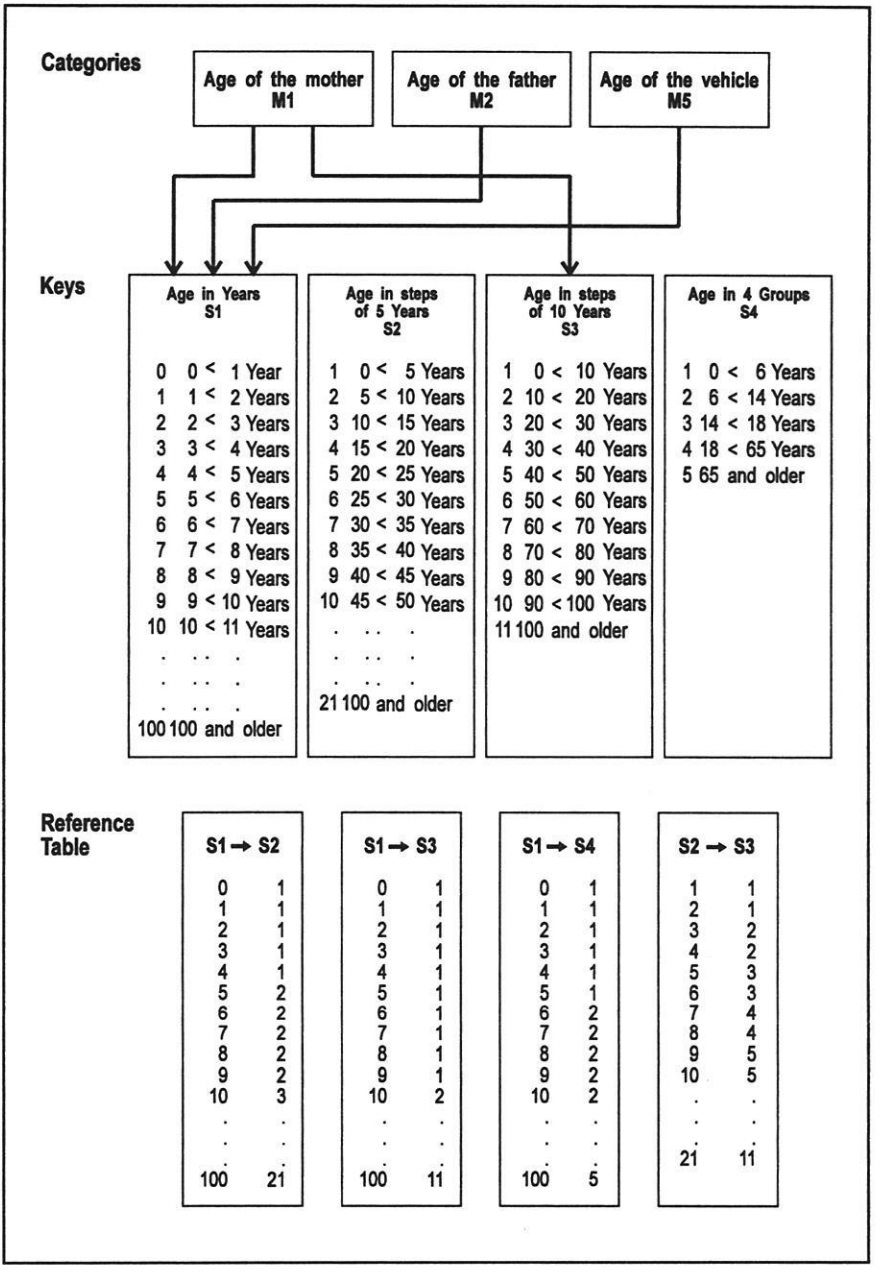
Statistical Confidentiality

To anonymize data for confidentiality reasons, traditionally, means for all statistical offices to satisfy conflicting aims, i.e. to find a sound compromise between confidentiality, on the one hand, and adequate data quality, on the other hand. We have developed ten criteria, which would have to be observed, if this conflict in aims, with which official statistics are confronted as a consequence of the confidentiality issue, is to be solved with standard instruments.

Due to the fact that confidentiality measures are taken at the end of the statistical production process and always involve individual expenditure, it was necessary, in order to establish standardisation, to look for solutions which would make it possible to deal with confidentiality in an early phase of the statistical production process, similarly as in the case of plausibility checks.

Reference tables are suited to ensure a standardised approach to the confidentiality issue, because it is possible to replace keys which, in a certain combination, raise confidentiality problems by other corresponding keys.

Figure 3 Key Concept



So far, standardised basic files offer a good basis, on which it is possible to organise this work so that all records with their concrete combinations of categories occur at least three times in the basic file. At first, this seemed to be unrealistic, in our opinion, but, in fact, it is principally feasible, because, though individual data are actually required to provide a data base which can be used flexibly to satisfy the most different evaluation needs, the statistical results themselves, however, always relate to statements about sets and subsets and to their structures and trends. But this is something for which just a few segments of the data base are utilised.

With our program system SAFE (standardised anonymization functions for individual data), which is now under development, we pursue an approach to confidentiality which uses copies of the basic files to exchange various keys of categories so that, on the one hand, the data stock includes only those records which occur at least three times, but, on the other hand, the frequencies of the various categories and also those of combinations of categories remain largely intact.

The perturbation technique used for the said purpose cannot be applied in the case of current economic statistics. Here it is better to solve confidentiality problems by using grouping techniques. And in order to avoid cases of predominance, it has turned out to be advisable to use the technique of suppression, because it does not prevent the results from being further processed by external users.

Doubtless, our SAFE confidentiality approach involves considerable manipulations with the initial statistical data, but the advantages connected with such standardised approach are substantial. For that reason, to arrive at an assessment of possible constraints and data quality, it is planned to also provide regular information about the extent to which evaluations made on the basis of anonymized data deviate from those based on original data.

But data quality also means, in our opinion, that the provision of data should not be delayed without necessity or be made more expensive by anonymization measures that may be required, which, however, is actually unavoidable with the present ways of production. In this sense data quality is worst in such cases where for considerations of data protection the supply of data, which would otherwise be possible, is either delayed or even rejected entirely.

Macro Files

With regard to statistical evaluations it is not advisable to operate directly on the level of basic files, even though they may be anonymized. According to experience, only some attributes are required for evaluations, and not the entire set of data, so that it is possible to raise the efficiency of work considerably, if the data are pre-compressed. This relates even more to standardised cases. How this can be done shall be demonstrated on a simple evaluation, for example, a file containing data on 100,000 individuals.

Population:	all inhabitants (100 000 persons) of the region Berlin-Mitte at a defined date.	**Subject** **Place** **Time**

Objekt: individuals

Serial No.	Attribute	Category of attributes	Key
M0	Identifier	serial numbers	1, ..., n
M1	Economic activity	economically active economically inactive	1 2
M2	Sex	male female	1 2
M3	Nationality	German non-German	1 2
M4	Marital status	single married widowed divorced	1 2 3 4

If all records in this file are sorted by attribute and category, it is comparatively easy to total the successively arranged identical records into one single record which is identified by the keys of the categories contained in a given combination so that it is necessary to just add the case number of the records totalled.

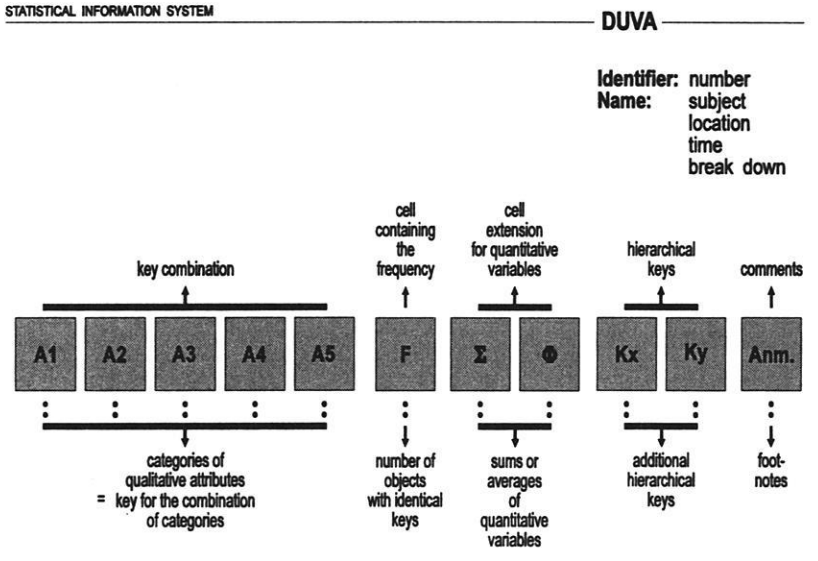
A large number of records were totalled in this way without any loss of statistical information. The theoretically imaginable number of data records in such a macro file is the result from the multiplication of the number of categories for all attributes, where the number of data records in the macro file can never be larger than the number of records in the intermediate file. In our example a maximum of 32 (2 x 2 x 2 x 4) data records may occur.

With this data model a standardised data base is available, which permits evaluations in any combination for all attributes included in the basic file, without making it necessary to use individual records. Its structure can easily be described completely, since it contains the same names as the basic file.

The structure of this macro file is determined entirely by qualitative attributes (Figure 4). It is, of course, also possible to transfer quantitative variables into macro files. However, they have no influence on the structure of the macro file, because they have always only one dimension, namely the unit of measurement. That is why quantitative values can be added to the frequency field as a sort of cell extension and also be stored in this macro file in the form of value sums or averages. In the

same way it is also possible to store various hierarchical keys in this cell extension and even to include, if necessary, footnote references or also additional relevant information, which may be important for the interpretation of the cell contents.

Figure 4 Data Model II: Macro Files



Since basic files must not include data that can be determined in the statistical production process, it is necessary to ensure that all operations, which are characteristic of statistical evaluations, can be made during macro file generation by using the metainformation system. On the level of files, objects, attributes and categories, this includes operations of selection, combination, computation, etc.

The macro file data model can be evaluated at will, together with the corresponding metainformation. An isolated evaluation of the macro files may be done, for example, by using PC axis developed in Sweden. But the normal case, probably, is that macro files are generally used as a standardised intermediate product for further processing in the most different forms both inside and outside the statistical offices.

Reference System

Traditionally, the classical output of a statistical office is the transfer of tables, statistical reports, charts and comments. In addition, there are internal supply tables as well as on-line storage capabilities, the transfer of diskettes, CD-ROM, geographic and other specialised data banks in various fields.

A statistical information system should take account of this variety of uses and of the specific requirement profiles of the various users, without compelling the user to

read through thick catalogues or to acquire specific knowledge in order to receive the statistics he needs.

It is important to grant all users equally easy access to the information potential and, at the same time, to use techniques which, as a feedback, analyse experience collected during the processing of user requirements and, in particular, their modifications in order to further develop the components of the system.

This is the purpose that our concept of a reference system, for which the following ten standardisation requirements were drawn up, is to serve:

1. Everything available in the statistical information system should be transparent to the user. The metainformation system with its descriptive information, to which each user should have access, serves this purpose.
2. All users should have access to the statistical information system at any time and in any way suited for them. There must not be barriers or unreasonable obstacles resulting from different states of knowledge or the use of different technologies. This easy access is realised by a thesaurus, since the use of natural language enables everyone to utilise the statistical information system. This approach makes it even possible to solve the problems raised by the use of different languages.
3. Data requirements should be met as inexpensively as possible. For that purpose the registration system provides information about the availability of final or semi-finished products (for example publications or macro files) or about the possibility of compiling results out of basic files. This makes it possible to find the financially most favourable solution.
4. This approach should be supported by electronic order processing to document and trace electronically all queries to the system irrespectively of whether they come from outside or inside the statistical office. By that, it is possible at the same time to implement a dynamic, use-related extension of the metadata, because new queries lead to new metadata and implementations in a form, different from that practised up to now, should simultaneously be seen as a service rendered in advance for future evaluations.
5. Electronic order processing can also be used systematically for dynamic system adaptation, by replacing parts which have proved to be less satisfactory in practice by better solutions so that the system is optimised continuously. For that purpose, the weak points are analysed periodically and appropriate improvements are made. In our opinion, this approach to optimization is a promising field for the use of expert systems.
6. Further possibilities to increase the efficiency of the system are provided by the thesaurus, which makes it possible to integrate into the metainformation system additional requirements which cannot be satisfied on the basis of a data-oriented approach, since the system can evaluate only those data which already exist in the system. Thus, unsuccessful queries can be evaluated systematically, in order to find out, for example, if missing data can either be replaced approximately by alternative

data or if an attempt should be made to open up new input-output sources. This would lead to a demand-driven extension of the data stock available.

7. As it is always extremely difficult for a statistical office to make outsiders understand the usefulness of its work, in particular that of a statistical information system, it is necessary to document systematically - in the sense of controlling - how the system is utilised. This can be achieved by interlacing the electronic processing system with an electronic accounting system, which combines the queries entering with the expenses for replying to them and automatically draws up invoices for those data supplies which are not free of charge.

8. Statistical offices have not only to take account of presently existing information needs, but they are also obliged to deal with quantitative historiography so that future clients, for example historians, can later receive statistics from the statistical information system. The appropriate means to take account of such future information needs is to register all activities of the system consistently in archives.

9. The service of a statistical information system should also provide for the possibility to establish links and to grant access to external data. For that purpose, it is necessary not only to supply descriptive information about the data available and the corresponding possibilities for processing them, but also to keep in the system a directory of sources and literature, which should be as complete as possible and continuously updated, so that access to text comments and external data sources is ensured at any time.

10. Last but not least, apart from the man-machine interface the system should also offer an individual form of service. For that it is continuously necessary to have organisation charts and to define responsible employees so that it is possible quickly to contact, if necessary, competent experts of the statistical offices who are able to give individual advice at any time.

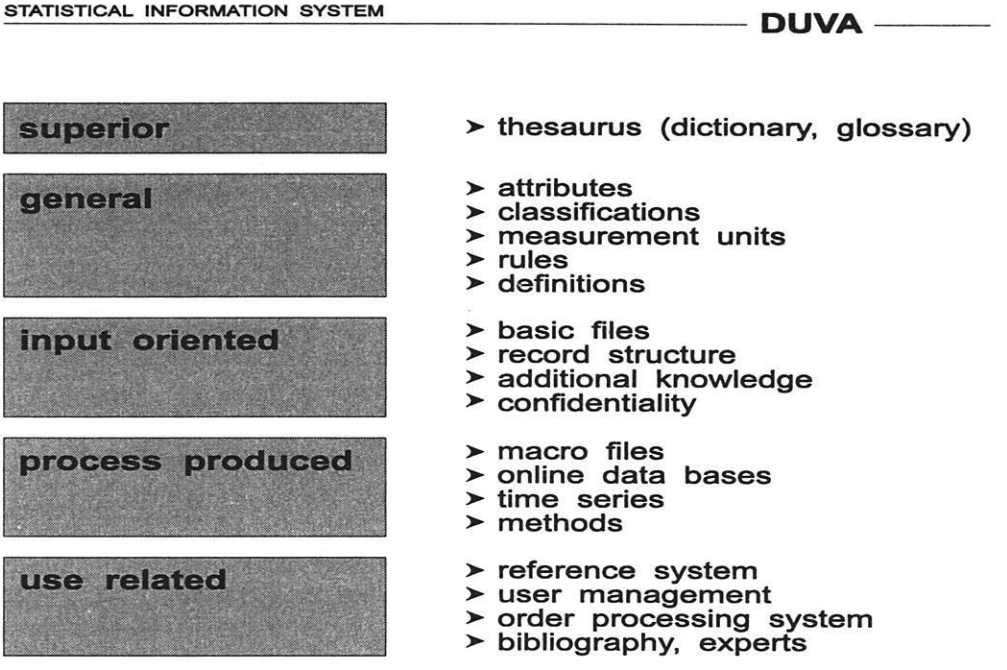
If these standards are taken into account, it will also be possible, in our opinion, to fulfil the expectations linked with a statistical information system, namely rendering optimal service to the clients of a statistical office, better than up to now, because, in addition to the supply of data, users are granted a possibility to exert influence in the individual components of the system by their concrete demands, and likewise it will be possible gradually to build up the metainformation system required for this purpose. How this can concretely be done will be described in the section that follows.

Metainformation System

What was said above was a description of the components needed for an operational metainformation system and of the possibilities to build it up gradually without running the risk that newly emerging components and demands cannot be integrated later into the existing system.

In this context, a superior element of such a system is the thesaurus, which at first is restricted to the inclusion of individual words used during the generation of individual basic files and the storage of the respective classifications. Then, links between these words and to the various elements will be established gradually (Figure 5).

Figure 5 Data Model III: Meta Information System



Attributes, classifications, units of measurement, rules and definitions have to be described as general components of this metainformation system. Input-oriented metadata include descriptions of basic files, of record structures and of the relevant extra knowledge required. Here the issue of confidentiality protection is dealt with in our project, too.

All metadata which are generated in the production process are stored with the process in which they are generated. This is done during metadata-driven production of macro files, during further processing operations, for example, in an on-line storage or for the production of time-series, and during the application of methods of mathematical statistics, which are used in the system.

Use-related metadata originate with the uses of the registration system, in particular, with electronic order management, supplemented by a bibliography and a list of organisational units and responsible staff members.

Final Remarks

The abundance of requirements with regard to subject-related standards of a metainformation system shows clearly that a large package of tasks has to be mastered here. It is certainly too much for an individual institution to cope with. However, in our opinion, promising fields of co-operation present themselves, in particular, with science. In addition, it is possible to approach the problem step by step.

As has been said, it is advisable to start by developing prototypes in selected sub fields, which can be used for new findings and later improvements.

In conclusion, it should be stressed that, in our opinion, without systematic elaboration of a theoretically founded concept for the solution of the metainformation problem, which is of relevance to the entire statistical information system, it will not be possible in future, either, to arrive at generally satisfactory statistical information systems, which means that, from this point of view, efforts in this field are justified in any case.