DUVA – Information Management System:

collecting - processing – presenting Information

Dietrich Bangert, Senatsverwaltung für Stadtentwicklung Berlin
Holger Pietschmann, Statistisches Landesamt des Freistaates Sachsen, Germany

The goal of a decentralized, regional statistical data collection and -evaluation can only be achieved if a common base on statistical data content and methodology. In other words, starting point, also every regional statistic is the creation of standardized metadata. This process is simpler, if it succeeds, one accepted by all stakeholders to create common methodology for metadata management. This knowledge is now generally accepted.

20 years ago, the DUVA project started with the both aims: to develop the methodological basis for a standardized metadata management as well as the necessary software tools. Regional statistical requirements played, from the outset a major role, for it was the institution of the project and is essentially the German local statistics.

The aim of DUVA is the cooperative development of a cost-effective, reliable and easy to use information management system. The DUVA-project was founded by a nationwide group of representatives from communal statistical offices in 1989. It's first objective was to develop a software tool for a comprehensive and flexible evaluation of the data of the German 1987 census. Due to the successful completion of this project, DUVA became a continuous organization. The portfolio of DUVA meanwhile covers the field of data entry, a common database and means to aggregate and to analyse data. All data is listed with a complete set of metadata. Existing metadata is used to control the entire process of data entry, data mining and data analysis.

DUVA is part of the KOSIS-network. The KOSIS-network was founded in Germany in 1982 as a common platform for different self-help projects in the field of statistical information management. Currently, more than 150 communes and other public institutions are associated to the KOSIS-network; 57 members such as the city of Berlin, the city of Frankfurt and the cities of Nuremberg and Dusseldorf as well as several federal institutions belong to the DUVA-section of the KOSIS-network at the moment.

Meta information and physical information

The statistical processes described above must be considered as a process chain rather than isolated, single steps. The objective of statistical investigations beyond the field of technical issues and statistical computations is to collect and maintain all process-related information and make this set of information available for comprehensive retrieval processes.

In this context two information qualities must be taken into consideration:
- meta information (all descriptive information) and
- physical information (measurement or survey results in the form of codes, values, frequencies or textual data).

Meta information is derived early on during initial reflections on research objectives. All subsequent steps lead to an increase in the amount of meta information. As to the amount of physical data the situation is different.

Following data collection, physical information is available in the form of measurement or survey results. The information load comprises all attributes collected for each statistical unit (e.g. each respondent). Since this unit-level information load is unsuitable for conveying knowledge, the body of data is reduced by means of several processes of abstraction. For example, primary data are first compounded into aggregate data (e.g. frequency tables) or into statistical measures (e.g. means, measures of dispersion, contingency measures, etc.).

The DUVA approach

Statistical work is not a static process chain. It requires maximum flexibility to address content-related issues, data source issues (primary, secondary statistics) as well as logistic (e.g. data flow, data communication), formal (e.g. data formats, etc.) and technical issues (e.g. data processing systems, programs for further processing, etc.). A comprehensive conceptual approach to data management of statistical information systems is needed.

A possible minimum manual solution to data management must involve detailed and systematic written specifications of every single step (from defining a research objective to data communication in the Internet via dynamic tables of results). Databases, spreadsheet programs and statistical analysis programs are very efficient means for maintaining and processing physical information. An obvious disadvantage of such minimum solutions is the fact that meta information and physical data are stored on different media (media splitting). Attempts to integrate both types of information require searches on the medium concerned and a time-consuming integration process at every single step.

The basic idea underlying DUVA system development was to provide a management system for meta- and physical information without splitting between different media and to provide meta information interfaces for all program elements forming part of the statistical production process (data bases, aggregation and analysis programs, tabulation, graphics, plotting programs, etc.). The content-related flexibility of the system is extended to include technical flexibility. Furthermore, the system should offer exhaustive search functions (e.g. search for individual data, linked information, analytic results, etc.).

The referral system (Metadata Repository) is the core element and control center of the DUVA information management system. The referral system collects and maintains meta information and uses this set of information for system control purposes. Although meta- and physical data are held by separate media, the NWS referral system provides access to physical information. System access and individual access rights are controlled by the user administration module.

Meta information input follows the principle that ‘metainformation is systematically collected at its source and is subsequently reused and augmented repeatedly throughout the statistical production process’.

Software modules of the DUVA metadata system

The DUVA metadata system is modular. All these modules use the same, centrally managed metadata. It consists of the following modules:

- Metadata Repository (referral system)
  All the metadata of Duva metadata system are centrally managed in a metadata repository. The kernel of the DUVA metadata are descriptions of files, which are arranged by sections. Furthermore, values and classifications will be maintained and their definitions. And finally, it can be described rules, which are used for processing of the data.

- Target file generator
  Using the target file generator will be produced from source data files new, so-called target data files. To this will be used the metadata, in particular the rules, stored in the central metadata repository.

- Form generator
  The Form Generator allows the use of the metadata repository to almost completely automated creation forms for data collection in the PC network and the Internet or intranet.

- Mdirekt und Internet-Assistant
  These two modules are used for data analysis and presentation of results. The module Mdirekt is a data analysis too for the PC network. The Internet Wizard provides the ability to perform flexible data analysis on the Internet.

Examples for using DUVA Metadata Approch and software modules

The DUVA approach to metadata management and the DUVA software tools are used in many local and regional statistical projects. Some examples are briefly presented below:

- Crossborder Friedship Database
- Data collection for Urban Audit in Germany
- PRISMA Berlin.

Crossborder Friendhip Database

The intensive international co-operation has been concentrated in EU harmonisation process recently. It involves new challenges for statistical offices, e.g. to offer datasets of comparable statistical indicators and to establish shared presentation systems of statistical data and corresponding metadata.

One of the projects of this aspiration is the Cross-Border Friendship Database (CBFD). It will be made accessible for public via Internet on occasion of the Conference after several years of preparation works. It has been focused on both laic and professional public. The database has arisen in co-operation of 3 statistical offices around the frontier of Germany, the Czech Republic and Poland. There are several hundred indicators involving demographic, economic, social and environmental statistics in the database. They are classified into 10 thematic domains. Predefined data tables were carefully chosen and assessed from the point of view of comparability and accessibility of corresponding data. All tables have been offered in 4 languages. Wide range of metadata involving description, definition and methodology of finding out the indicators are at disposal, too. The database contains data on the spatial level of NUTS. Data describing Euroregions around the common borders have been prepared, too. The users can access the database via standard Web browsers at the Internet-Adress www.CrossborderDatabase.org.

In this project be used the basic concepts of Duva for metadata management. The reason is that the metadata repository can not be maintained over the Internet. Furthermore, no multilingual metadata can be managed. Both are important aspects for the further development of DUVA metadata system.

Data collection for Urban Audit in Germany

Using the DUVA software tools and using a common set of metadata, data collection for Urban Audit is conducted in a standardized manner. This uniform data description has been developed and stored in the metadata repository DUVA. It was then that the forms for collecting data automatically generated. Via the Internet now can all participating municipalities in Germany enter their data in the online forms.

Planning Related Information Service for Spatial Monitoring and Analysis (PRISMA) - The DUVA System in the City of Berlin

Berlin –a metropolitan area in transformation

Since the fall of the Wall Berlin has been in a continuous comprehensive process of transformation including all aspects of urban, economic and social life. This transformation process included the challenge to change from an old fashioned industrialised society to a society based on services, technology development and science. Despite the heavy burdens of the past Berlin is on the way to a modern multicultural metropolitan city.

Berlin administration – tackling with deficiency

One of the burdens of the City of Berlin are the enormous debts which restrict the options to develop the city enormously. In the focus of the expectations of politicians to transform administration into a slim customer oriented organisation public authorities are confronted with a huge loss of administration staff and - in the end - the loss of knowledge in every field of administration action. As a consequence corporate governance has to be adapted: administration action has to be based on modern governance principles like target-oriented transparent proceeding and project oriented action as well as on public participation to communicate administration actions and make it comprehensible and acceptable for the public.

Organisational transformation: one strategy for social urban development

Within this context an urban development strategy has been evolved that targets to transform sectoral oriented planning processes into integrated concerted planning actions based on “Living Environment Areas” (LEA) for which area specific planning targets have been developed. By making these planning actions transparent for all stakeholders of the LEA means can be combined and concerted, sectoral planning targets be harmonized and investment activities be bundled. Another aspect is that means and the allocation of funds need not to be inefficiently spread over the whole city but can be concentrated in identified areas of high priority.

PRISMA: data analysis and distribution for integrated planning action

For this new and powerful integrated planning approach comprehensive current information on a small-scale basis is needed. Together with all city administrations including eight senate departments, twelve boroughs of Berlin and the State Statistical Institute Berlin-Brandenburg a “Data Pool” has been built up to serve as the central information source and platform. About 60 data stocks with ca. 500 variables have been identified and will be kept in the data pool and provided by a multi-tenant information system which will allow to allocate specific data for specific user groups. This data pool will be combined with an already existing geo information system (“Map Pool”) for spatial analysis purposes using - on the basis of international standards - components of the Geo Data Infrastructure Berlin-Brandenburg which is in progress. To satisfy user needs a user-friendly Data Navigation Assistant (DNA) will be developed as part of the “Planning Related Information System for Spatial Monitoring and Analysis” (PRISMA). For these purposes DUVA software components will be used.