



UNITED NATIONS  
ECONOMIC  
AND  
SOCIAL COUNCIL



Distr.  
LIMITED

E/CONF.69/L.41  
30 June 1977

ENGLISH ONLY

---

THIRD UNITED NATIONS CONFERENCE ON THE  
STANDARDIZATION OF GEOGRAPHICAL NAMES  
Athens, 17 August-7 September 1977  
Item 11 of the provisional agenda.  
Automated data processing (ADP):  
(a) Coding and abbreviation

DATA PROCESSING FOR THE PREPARATION OF THE  
GAZETTEER: FEDERAL REPUBLIC OF GERMANY

Paper presented by the Federal Republic of Germany

## 1. Introduction

Aim of the Information System for the Capture of Geographical Names from Official Topographic Maps of the Federal Republic of Germany

The names of geographical features are of no interest as far as they are not explained by information on the kind and function, dimension, and location of the feature concerned. The "Gazetteer Federal Republic of Germany", being prepared as a contribution to the "United Nations Series of Gazetteers" is to comprise all names with explanatory information in alphabetical order and official spelling, as shown on the official maps of the Federal Republic of Germany. For the time being, the gazetteer is limited to the about 30 000 names shown on the General Map 1 : 500 000 (ÜK 500) of the "World Series 1404".

If later on the names shown on maps of larger scales will be included, the volume of the gazetteer will considerably increase. So we expect that about 180 000 names must be captured from the General Topographic Map 1 : 200 000 (TÜK 200).

For the features indicated by names the following descriptive details are given

- kind and function by feature codes,
- location by map sheet number, coordinates, height above mean sea level,
- dimension by number of inhabitants, area, length,
- administrative area (relation) by statistical key number.

A gazetteer, comprising all names shown on the map series at a certain scale and giving such additional information, can be used for different purposes:

- as a general source of information on geographical features,
- as a base for the representation of administrative structures,
- as a compilation of names, from which those are selected by kind and function, location and dimension, which are to be shown on a map sheet to be prepared,

- as a compilation of criteria for the selection of kind and dimension of map lettering and its placement on the map.

However, in order to make such a various use possible, it is necessary that the data, as e. g. the number of inhabitants, be updated continuously and that they can be selected and sorted according to different characteristics. Electronic data processing (EDP) is the best means to solve this task. The Institut für Angewandte Geodäsie (IfAG), charged with the preparation of the "Gazetteer Federal Republic of Germany", has considered, already at a very early date, the use of EDP. While doing so, we have not restricted ourselves to the preparation of a "gazetteer" only, but proposed to establish an Information System for the Capture of Geographical Names from Official Maps of the Federal Republic of Germany in order to serve all purposes mentioned above.

We have particularly considered the plan to use the information, stored together with the names, for the automatic preparation of the printing plates for map lettering. The result has been that in addition to the coordinates of the feature concerned, the coordinates of the base line of the name on the map, as a rule those of the left bottom corner of the first letter, have also been captured.

Without the coordinates of the base line, letter type and letter size can already be determined by means of kind, function, and dimension of the topographic feature. In the Institute we plan the following procedure which considerably exceeds the scope of the above project:

The names stored are displayed, together with the symbols of the appertaining features and their surrounding, in their correct size on the screen of an interactive graphic system by means of the coordinates of the base lines. By displacing the names on the display to the most favourable position, the coordinates of the base line are changed at the same time. In an automatically controlled photo-setting device, the names are "written" on a film in the correct type and size of letters at the positions determined by the corrected coordinates of the base line. After its development this film serves the production of the printing plate for map lettering.

## 2. Structure of the Information System for the Capture of Geographical Names from Official Topographic Maps of the Federal Republic of Germany

The "Information System Geographical Names", as we will abbreviate it hereinafter, consists in its core of a data bank system which admits at any time to input new data, to modify existing ones, or to erase them, and finally to select a great number of data according to certain characteristics. This system is supported by several programs: Examination of data groups at a given format with regard to selection criteria and interrelations, storage of the characteristics and relations found in a suitable way, classification of the search characteristics in case of calls to the information system in such a way that the data bank system can find the data records in question by means of the stored characteristics and relations and make them available in an output data file.

The data bank system DATAS has proved very suitable for the purposes of the Information System Geographical Names. It has been described in detail by Wigand Weber in "Ein Datenbanksystem für Geographische Namen" (A Data Bank System for Geographical Names) (Nachrichten aus dem Karten- und Vermessungswesen, Series I, No. 69). In the meantime, DATAS has been installed in the computer TR 440 of the "Gesellschaft für Mathematik und Datenverarbeitung mbH (Society for Mathematics and Data Processing), Bonn, Bereich Darmstadt (GMD)". However, it can also be operated on a smaller computer.

With the computer TR 440 the following processes are also performed: (a) The data bank input, i. e. the capture of names and additional information on data carriers (magnetic tape, punch cards) and the preparation of the data up to the provision of the data records for the data bank, (b) the production of the gazetteer, i. e. the sorting of the data records in the alphabetical order of the names, writing of the numerically coded information in clear text, and the printed output. Planning, programming, and computation is done by the group of operators in the IfAG.

### 2.1 Data Bank Input

The data bank of the information system stores the following data:

- the name of the feature, names other than place names are followed by the grammatical gender (m), (f), (n), (pl);

/...

- a second name of the feature or the name of a special superior administrative unit below "Kreis" (district) level;
- the coordinates of the central point of the feature (Gauß-Krüger coordinates, geographical coordinates, and UTM coordinates);
- the Gauß-Krüger coordinates of the base line, necessary for placing the name on the map;
- the height of the feature above mean sea level;
- the dimension of the feature (number of inhabitants of towns and communes, length of rivers, surface area of lakes and landscapes);
- the statistical key number, indicating the political affiliation to superior administrative units as Land (Federal State), "Regierungsbezirk", "Kreis" (districts), town or commune;
- the hydrographic code number of rivers, lakes, canals;
- geographical and hydrographic code numbers;
- feature codes, taken from the feature catalogue, attached to the "Standard-Daten-Format für den Austausch kartographischer Daten" (Standard Data Format for the Exchange of Cartographic Data) which explain the kind and function of the feature by four digits;
- numbers of the official map sheets on which the name is shown.

The data are input by two steps:

- Capture of the data on data carriers, e. g. magnetic tape or punch card;
- preparation of the captured data in order to supply the complete data records in a "Data File of Geographical Names" which are taken over into the data bank system.

#### 2.1.1 Data Capture

When establishing the Information System Geographical Names, the most time-consuming factor is the data capture, i. e. the transfer of the data compiled from different sources to the data carrier. Therefore we strived from the very beginning to avoid the repeated input of information, to take over the data already captured elsewhere, and to admit the least possible number of intersteps, as e. g. the preparation of separated capture documents. For this reason, we have developed

/...

and tested a procedure by means of which the coordinates of the central point and the base line of each feature are measured at the digitizer, the names and all additional information are input on a keyboard of the digitizer, so that all data are stored together in one data record on magnetic tape.

#### 2.1.1.1 Data Capture at the Digitizer

The off-line digitizer Bendix-Aristogrid used consists of a table with a cursor movable on it, both being coupled to a control unit with operation field and output installations. The coordinates of the cursor with reference to a starting point, the table coordinates, are continuously stored electronically. In the same way the input data, - as name, height, key numbers, feature codes - can be stored. After pressing a record key, all stored values and, in addition, a preset header are output on magnetic tape or also on punch tapes and lists.

For reasons of intelligibility and accuracy, the table coordinates of the central points and base lines of the features on 1 : 200 000 maps are measured at the digitizer. Therefore, a preparatory working group must mark all named features to be captured as shown on the 1 : 500 000 map on the corresponding map sheets of the General Topographic Map 1 : 200 000 (TÜK 200) before they are captured at the digitizer. This group must also assemble the documents showing the additional information.

At the digitizer, the data records of the following kind are input one after the other and written on magnetic tape:

Header 1111: Label: Input of number and scale of map sheet, number of magnetic tape, names of map sheet and operator;

Header 2222-

Header 7777: Sheet corner record: Clockwise measurement of the table coordinates of the corners of the inner map margin, starting at the left bottom;

Header 0000: Feature data record: Measurement of the table coordinates of <sup>the</sup> central point and base line of the feature and ensuing input of name, feature code, height, statistical key number, and number of inhabitants or code number with indication of length or surface area.

(Erroneous fields or records can be repeated after input of an error sign before being output on the tape.

Umlauts and ß have to be labelled by placing a "+" in front of the basic vowel, small letters at the beginning of a word being labelled by several interspaces in front of the word);

Header 7777: Terminal sheet corner record: In case of interruption of the capture of feature data records, above all in case of vibration of the measuring instrument, and after the last feature data record, additional sheet corner records are to be inserted;

Header 8888: Terminal section record: After that, the data of a new capture section (map sheet) can be stored on the same magnetic tape, starting again with header 1111;

Header 9999: Terminal tape record: With the capture section the magnetic tape is also concluded and can be transferred to the computer for editing.

#### 2.1.1.2 Data Capture via Punch Cards or Display Terminal

The data capture by digitizer is particularly suitable when a great number of features with central point and base line coordinates have to be captured. We plan the input in punch card format for correction and supplements of already captured features - additional feature code: new statistical key numbers, and officially determined central point coordinates in the national system in case of newly created communes and for the adding of certain features not yet captured, the coordinates of which have been determined elsewhere. The data can be punched in the prescribed format in cards and input into a data file, or in the same format be input into this data file from a display terminal. The different input modes (1-3) to be applied are conditioned by the effects of the modifications on other fields within the data record or also on other data records, as e. g. modification of map sheet numbers when the feature coordinates are modified, modification of the sorting field when names are changed.

Mode 4 is applied in case of a new capture of features with all additional information.

When the input is made via punch cards or terminal, only Gauß-Krüger

coordinates, geographical coordinates, or UTM coordinates must be input as feature coordinates, but no table coordinates.

## 2.1.2 Data Editing for the Input into the Data Bank of the Information System

### 2.1.2.1 Editing of the Data Captured at the Digitizer on Magnetic Tape

At the digitizer, data records with 1536 characters each in the EBCDIC-Code are written on tape. For the processing on the computer TR 440, the data records of a tape are stored on magnetic disk by translation into the central code ZC1 of the TR 440 and reduction to 359 characters in a data file with random access.

These data are edited for the transfer to the data file in a program system called "GEONAM", the run of which can be controlled in interactive mode from the data remote processing station of the Institute, interrupted in case of gross errors, and continued after correction of small errors.

In detail, the following operations are executed:

- Processing of name strings by insertion of the central code data for umlauts, small letters, and ß, recording of the length of strings and position of umlauts, establishing of a sorting field for alphabetizing the data records;
- presentation of a test drawing at the plotter of the computation centre. This drawing shows the features and names by means of the measured table coordinates of the central points and base lines of the features. By comparison with the original maps one can find out, whether features have been forgotten or names and features do not match;
- output of lists of the captured data in order to check the data (names, key numbers, heights, feature codes) input via keyboard;
- determination of sheet corner coordinates by means of the number of the measured map sheet, and computation of the elements of a conformal or affine transformation of the table coordinates into Gauß-Krüger coordinates of the national system, using the sheet corners as control points;
- transformation of all table coordinates of central points and base lines of features into Gauß-Krüger coordinates, transformation



- of Gauß-Krüger coordinates into geographical coordinates, and transformation of geographical coordinates into UTM coordinates;
- determination, by means of its geographical coordinates, of the numbers of map sheets, pertaining to the series from the Topographic Map 1 : 25 000 (TK 25) to the IMW 1 : 1 000 000 (IWK 1 000), on which the feature is shown;
  - assignment of the commune name to parts of populated places by means of the statistical key number.

The data, edited in this way, are included in the data file of geographical names for transfer to the data bank of the information system.

Errors in the data can be corrected as follows:

- By repeating the capture at the digitizer and the ensuing run of the above program;
- by input via punch cards or terminal according to one of the four input modes.

#### 2.1.2.2 Editing of the Data not Captured at the Digitizer

After the first capture, as well as at a later date, the data stored in the data file of geographical names can be modified or corrected and supplemented. A program system for the manipulation of feature data of the gazetteer (GNOM), going back to those parts of the program, which are also used for editing the data captured at the digitizer, effects the different operations corresponding to the input mode in question, as e. g. the processing of the name string (capitals and small letters, umlauts, sorting field), transformation of the input Gauß-Krüger, geographical or UTM coordinates into the respective two other coordinate systems, determination of map sheet numbers by means of the geographical coordinates. The feature data, captured at the digitizer, as well as those input via punch cards or at the terminal, are available at the end of editing in the data file of geographical names for the transfer to the data bank of the information system. In this data file, a record number is definitely assigned to each feature, enabling random access to the data record in question.

## 2.2 Data Output for the Printing of the Gazetteer Federal Republic of Germany

For the Information System Geographical Names, the essential data of features - kind and function of features, administrative and geographical area - have been coded numerically. The data records are numbered and stored in the sequence of their capture. The Gazetteer Federal Republic of Germany shall list the geographical names in alphabetical order, as well as feature codes and administrative or geographical area in clear text. This requires, besides a sorting process, a considerable expenditure in storage capacity and computation work.

The alphabetical sorting of the data records is performed by means of the sorting operator of the computer TR 440. The sorting field has been established when processing the name strings after their capture. It comprises 32 characters.

The information on the length of name strings, positions of umlauts, sorting field etc., as also stored in the data file of geographical names when editing the data, are only required for the output of names in order to print the gazetteer. As they can at any time be reconstructed from the edited name strings too, they are not taken over into the data bank of the Information System Geographical Names in order to save storage capacity. At the time being, therefore only the data file Geographical Names is used for the preparation of the Gazetteer Federal Republic of Germany. The use of the information system is, at present, restricted to cartographic applications as e. g. the above-mentioned preparation of map lettering.

The further processing of the sorted data records is done by means of the GEODRUCK program system. It comprises, among others, the following individual operations:

- Insertion of mnemotechnical abbreviations of feature codes;
- classification of the statistical key numbers by Land (Federal State), Regierungsbezirk, Kreis (district), town or commune;
- classification of the code for geographical features according to the control number of the structure of natural regions and/or hydrographic codes;

classification of the hydrographic codes according to the receiving hydrography concerned (Rhine, Main, Nidda, Wetter).

In order to provide the correct spelling of the character sequences corresponding to the key numbers, as well as that of the names themselves, they had also to be processed by capitals, small letters and umlauts, and to be stored in such a way that quick and unequivocal access is possible by means of the key numbers.

At present, the Gazetteer Federal Republic of Germany is being output on a high-speed printer equipped for spelling of capitals and small letters and installed in the computation centre of the "Gesellschaft für Mathematik und Datenverarbeitung". Before being printed, the text pages must still be reduced.

In the meantime, tests have been performed to avoid the high-speed printer and to write directly on microfilm, and then to enlarge the images for reproduction. For the future, it is also planned to use an automatically controlled phototyping installation.

-----