# Automated Spatial Data Processing and Refining

Marion Simon[✉] and Hartmut Asche

Geoinformation Research Group, Department of Geography,
University of Potsdam, Potsdam, Germany
`{marion.simon,hartmut.asche}@uni-potsdam.de`

**Abstract.** This paper focuses on the definition of a method for data processing by means of automated professional map generation. For this, initially services have to be identified that represents cartographic rules and recommendations. In order to link those services with respect to their cartographic content and to control the process within a component, a set of rules has to be designed. This is explained by examples and can be used as a template pattern for other services. Individual services and modules within the process will be arranged hierarchically on the basis of the cartographic visualisation pipeline. Its consequent graphical classification is presented. The aim is to prepare the theoretical cartographic basis in a formal way, which should enable technical implementation without cartographic technical expertise.

**Keywords:** Geovisualisation · Rule-based map construction · Mapping assistant service · Cartographic visualisation pipeline · Cartographic services

## 1 Introduction

Since 1972, when the first operational remote sensing satellite provided digital data of our environment, the amount of geospatial data has been increasing at an exponential rate. To process and analyse such geodata, dedicated ICT systems (information and communication technology) systems called geographic information or, in short, geoinformation systems (GIS) have been developed. Geodata modelling and analysis can be considered their prominent strength, visualisation in cartographic modelling quality their major weakness. In contrast, graphic-oriented map design systems provide extensive drawing and design functions but lack non-graphic data management and analysis capabilities. Recent years have seen a constantly growing demand for maps combining cartographic visualisation for visual analysis and interaction with the underlying geodata for non-graphic data manipulation and analysis. At the same time, map production and geodata processing technologies employed by commercial map producers do not facilitate an integrated production process for the generation of maps in cartographic modelling quality from geospatial databases. This situation poses a major economic problem which has caused a significant number of map producers to cease operations in the last decade. The research presented here addresses this problem. It aims at developing a technology concept allowing for the integrated processing and construction of non-graphic and graphic cartographic products from existing geodata.

The key objective is to design and implement a modular, scalable workflow that processes geospatial input data into application-specific quality geodata and map products. It is obvious that professional and economic expertise of a commercial map producer is essential for the success of the research.

In the past decades, extensive research has been carried out on GIS, geodata acquisition and filtering, geodatabases and geodata management as well as on digital map construction and geovisualisation [1, 2]. Map production processes from geodata stocks have received much less attention [3]. Projects relevant to the research dealt with here include the development of a now operational process for the automated construction of quality map graphics [4] and the ongoing development of a map construction assistant facilitating a rule-based map construction process for cartographic visualisation of statistical mass data [5]. These projects and other related research work provide an appropriate basis for our research. Combining available geodata processing and map production functions in one process will, however, not bridge the gap between professional geodata modelling and professional map construction. What is required is an integrated, software-driven process in which dedicated data and map modelling modules interact to produce application-specific geodata sets or cartographic maps or interactive data-based geovisualisations.

The contribution of this paper is presenting a method for building up a rule-based modularised process for automated quality map construction that can be implemented as a mapping assistant service. The next section of this paper describes the overall approach in the field of developing a process for automated map construction. Section 3 focuses on the characteristics of one selected map representation type as an example for showing developed method of process definition. Finally, in Sect. 4, some conclusions and open questions are discussed.

## 2    Materials and Methods

A process for automated data processing is the focus of several professional disciplines. Automated map production researchers are dealing with this issue since decades without having arrived at a conclusive solution yet. In this work the perspective comes from geoinformatics and deals with a method for automated map construction. Building on the classical visualisation pipeline [6], general working steps in map production are integrated [5, 7]. For automation within a service those modules are split up in submodules based on identified rules, definitions and recommendations from literature. Therefore a collection of rules in natural language is built up and classified to matching module components. A visualisation of relations and dependencies between modules is done with jABC (Java Application Building Center) framework [8]. The visualisation of graphs precedes a development of a typology for subsequently creation of fomalised rule-sets. This paper shines a light on the visualisation process in general (Fig. 1) and how it could be used for developing a process chain for automated map production. Additionally a rule set for setting case sensitive sequential process definitions will be discovered. Both could be used for developing a so called Mapping Assistant Service.
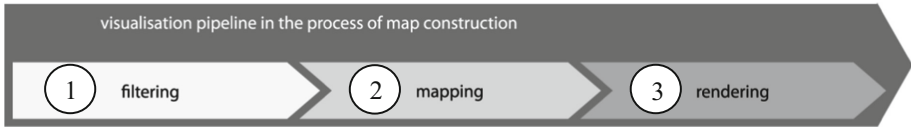
**Fig. 1.** Visualisation pipeline in the process of map construction by Mapping Assistant Service [5]

In this work the model of the visualisation pipeline provides the base for embedding additional sub-processes. The visualisation pipeline for automated professional map generation consists of the core processes filtering, mapping and rendering at the upper hierarchical level. These core processes are first explained below in general and then partly broken down to the necessities of a Mapping Assistant Service.

For the prototype, the presented sequence is used as a basis for embedding sub-processes and services which are presented as sub-models in different hierarchical levels.

Filtering means data processing which can be carried outfor example by completing, reducing or filtering the amount of raw data. This core process includes acquisition or editing metadata equally. The following step of mapping comprises mapping of the pre-processed geometry data including their attributes. Transformation of presented data into image data and/or into a digital image is called rendering. Following Kucharczyk [7], the three featured core processes are partly further split by means of map production.

## 2.1   Filtering Module

Correction of erroneous values and changes in the data basis as well as calculation of new values out of the existing should be provided by the service (Fig. 2).



**Fig. 2.** Cartographic processes within filtering module

These functionalities increase output quality on the one hand. On the other hand a representation of relations between data could generate new knowledge to the user. Basically, an error-free and complete data set should be available for use. Since the mapping assistant service is a software for representation of statistical mass data with its spatial reference, it should be possible to calculate statistical characteristics (stored as new values). In the preparation process of statistical information, the correlation and regression analysis, determination of frequency distributions, standard deviations and confidence intervals is recommended by Hake et al. [9].

Scale-dependent options are possible in further process flow. An output map scale (target scale) could be fixed at the very first beginning setup of using the assistant,

but after this point within the process it has to be specified by the user. Since the output parameters (e.g. layout design, map-scale, medium, title of map) could in theory also be queried at this point by the user, this process step has not been taken from the graph. Kucharczyk [7] distinguishes between determining spatial reference and allocation of projection. Here, a georeferencing module comprises both mentioned sub-processes. It includes checking spatial extent (geographical bounding box defined by north-, south-, east-, west-limitation), geocoding or transformation of geodata as well as choosing and application of optimal projection. An automated selection of an appropriate carto-graphic representation method necessitates a preceding data analysis module for examining present meta data. The analysis covers geodata on dimension (D), semantic information (S), scaling niveau (SN), attribute value (A), attribute value course (AV) and time relation (T). Model generalisation is according to Schürer [10] the transformation of an object model "with respect to its semantic and/or geometric res-olution, its data model and its structure simplified or in a new digital object model".

## 2.2   Mapping Module

To proceed with the selection of a cartographic representation method the system gets the outcomes of the data analysis. Within the step of object sign referencing, it comes to the linking of symbols and geo-objects. Finally the process step map sign con-struction complements the mapping part. This predetermined sequence is used to design an entry for the prototype. Because of the differentiation of the last two steps it can be assumed that either the appropriate cartographic symbol for a geo object does not exist, or that the user is free to generate symbols. A module for map symbol construction ensures a gap closure and requires a return loop for object sign referencing for assignment of the new cartographic symbol to related spatial data.

   A renouncement of mentioned map symbol construction modules for prototype development is feasible (Fig. 3). For the prototype, development should be dispensed with map symbol construction first. Instead, the symbol library should have sufficient symbols for topographic base maps as well as thematic content representation in stock.



**Fig. 3.**  Cartographic processes within mapping module

## 2.3   Rendering Module

The third core process of the visualization pipeline (Fig. 4) begins with the represen-tation of referenced map symbols. Thereafter cartographic generalisation for visual optimisation of a cartographic representation follows. The map production in carto-graphic quality depends on this sub-process. It contains "necessary respectively applicable processes and types of generalisation and abstraction that leads to a transfer of differentiated and detailed reality into an expressive map representation" [11]. Objective (e.g. map border, map field, map frame) and semantic map components

(e.g. map content, map title, map grid) all together make up the marginal data. The map composition includes the creative round off to finalise the cartographic result and should be, according to Kucharczyk [7], partially left to the user. For this last process step within the visualisation pipeline, an editor is conceivable, which includes components for printing, exporting or saving the map product.



**Fig. 4.** Cartographic processes within rendering module

## 3   Mapping Assistant Service

A process has been conceptualised starting from customer-centred geodata input to enterprise-based data filtering, analysis of visualisation options communicated to and endorsed by the customer, automated map or geodata generation according to the specifications agreed on, prototype production, again communicated to and approved by the customer, to the customised final product and delivery to the customer.

For implementation of cartographic representation methods within a mapping assistant service their specific characteristics must be identified. It contains features of spatial data as well as design elements and its possible variations depending on data basis.

### 3.1   Set of Rules

A modularised implementation of the components should be considered in the context of a rule-based system. According to Uthe [12] a rule-based system consists "of a database with valid facts, the rules for the derivation of new facts in the knowledge base and the rule interpreter for controlling the efficient selection and execution of rules that perform certain actions". A set of rules is based on the defined requirement formulation of the specification. A module-based structure of the system brings the guarantee of clarity of the regulatory framework, since the number of rules are placed per module. The importance of rules is exemplarily constituted for the sub-process (second hierarchical level of visualisation pipeline) selection of cartographic representation method, its design elements and the variations thereof. For the development of the regulations, a typology is always created, which can be applied in their scheme to other sub-processes or modules.

The presented rule set uses results generated in the filtering-module *data analysis* and stored as variables D, S, SN, A, AV and T (see Chap. 2.1). For building up a rule set structure a method of Germany's Federal Ministry of Transport and Digital Infrastructure (BMVBS) [13] served as a basis. Following BMVBS the "elements of a rule-based system are existing rules composed of a condition part (premise) and an action part (conclusion)" [13]. From the cartographic point of view.

1. Selection of an optimal cartographic representation method (phase 1),
2. Linking of design elements and cartographic representation methods (phase 2),
3. Linking of variations and design elements (phase 3).

**Selection of Cartographic Representation Method.** Phase 1 involves the selection of the optimal cartographic representation method (here: RM) for given data to be visualised. The available methods are summarised in the corresponding typology (Fig. 5). Based on the analysis results from the data pre-processing (data analysis), the selection-default is embedded in the set of rules and the results are stored for further processing, in the cache, for example, in a variable RM.

| level 1 | | | level 2 | | |
|---|---|---|---|---|---|
| code | abb. design. | designation | code | abb. design. | designation |
| M1 | RM | cartographic representation method | M11 | RM_MDS | method of diagram signatures |
| | | | M12 | RM_MDM | method of diagram maps |
| | | | M13 | RM_MPS | method of positionsignatures |
| | | | M14 | RM_CMG | method of choropleth mapping related to geometric referenced area |

**Fig. 5.** Typology of cartographic representation methods (RM)

The set of rules to be implemented rests upon the results of the requirements analysis for a prototype of the assistant as well as on representation methods identified as necessary based on the component data analysis within the filtering process, in order to reflect any kind of data basis. In order to meet the above condition, the results of data analysis (data preparation) have to be read out in a first step by the system and to be complied with the subsequent election of the representation method in Fig. 6 listed specifications.

**Linking of Design Elements.** Phase 2 specifies the representation of the selected method by assigning the design element(s) (area, diagram, signature line and line with arrow), which are treated as separate modules in the system. The result of this phase is for further system access, here in the variable DE (design element) cached. The assigned notational conventions define the appearance of visualisation in principle without defining the unique appearance. This separation of design elements and variables is used for the sake of clarity with respect to the portion of dependencies modules to each other. From a business perspective the assignation of design elements as realisation of specifications is necessary, whereby said module parts are inferior to the hierarchical level of the module design element.

**Linking of Variables.** In Phase 3 the linking of design elements and the modules for the variation takes place. Variations (size, shape, direction, colour, brightness and pattern) can be limited for certain design elements. A basic allocation of variations to

| code | cartographic representation method (phase 1) | specification (related to data basis) | premise |
|------|------|------|------|
| 1 | RM_MDS | selection for data with point-dimension in combination with quantitative semantic information | D = „point" AND S = „quantitative" |
| 2 | RM_MDM | selection for data with area-dimension in combination with quantitative semantic information | D = „area" AND S = „quantitative" |
| 3 | RM_MPS | selection for data with point-dimension in combination with qualitative semantic information | D = „point" AND S = „qualitative" |
| 4 | RM_CMG | lection for data with areal- ~ombinat' | „area" AND |

**Fig. 6.** Rule set – Selection of optimal cartographic representation method (RM)

design elements is excluded due to professional dependencies. A variation in the brightness is at least in the design element signature applicable, if the scale is ordinal and the RM method of position signatures is used. Accordingly, it is at this point necessary for the system, as in the selection of cartographic representation method, to again access the results from the data analysis.

## 3.2   Process Definition by Example

This use case aims at an automated cartographic production of an economic map by assistant. The chosen digital dataset contains feature classes relating economic data of Germany in the year 2006 [14]. The amount of working population, differed by economic sector, describes the economic structure for each administrative district (NUTS 2-level). For orientation, further topological data (e.g. river, administrative borders, capital cities of federal states) are contained. This use case focuses on processing data to thematic map representation under consideration of given and processed topological data representation.

Based on the described dataset process, steps of the filtering module can mostly be skipped. There is no need for correction of erroneous values or calculation of new values. Map scale depends (in this case) on spatial extent and size of output medium. For this use case a map scale 1:2,500,000 is calculated and suggested to the user by the assistant automatically. The georeferencing sub-module is skipped as well. The given dataset contains spatial metadata for projection. These metadata will be compared with default reference coordinate systems held by the service. Loaded data will be checked for delivered coordinate systems and projection parameters by the Mapping Assistant Service. A data analysis will be done automatically. It determines topological spatial

structure, data dimension and semantic information. Depending on the outcome further data attributes were determined. For this case an extended analysis is not necessary. This analysis module cooperates with the mapping module selection of cartographic representation method (RM). This prevents calculations that are not needed for dedicating the optimal RM. Furthermore the data analysis tool determines the need for data classification. In this use case the mapping assistant service will initiate a classification within the following module *model generalisation*. This is necessary for the cartographic generalisation module within the rendering sub-process, described below.

The selection of the cartographic representation method realises the first step in the mapping part of the visualisation pipeline. It uses the outcomes of the data analysis-module to modify the production process for this certain case. Asche et al. [15] describe the complex selection process for an appropriate cartographic representation for each possible combination of data characteristics method by UML (Unified Modeling Language). Going through the mentioned process the following outcome will appear for this example of an economic map of Germany 2006: The topological spatial structure is discrete. The amount of working population is given by the year 2006. There is no continuous value representation. Data dimension is area, because the values are related to an administrative unit. The amounts of working population are data with quantitative semantic information. This leads as seen in Fig. 7, to the choice of method of diagram maps (RM_MDM).

An object sign reference module contains the rules for a connection of data to libraries containing design elements and related variations. In our sample case, the design element *diagram* is allowed for representing the given data. The user can choose the diagram style (circle, bar, square chart). For this paper diagram style circular chart is chosen by the user. A variation of this chosen design element size and colour is allowed. For a monochrome representation grey is chosen. The user is also allowed to define if the size of the circular chart should also represent the total amount of working population. A static size only represent the portions of economic sectors.

The basemap must be brighter than the represented theme to guarantee the semantic hierarchy and the readability of the map. Depending on brightness, saturation and colour of a basemap, the assistant offers through rules the only possible range of grey variation to the user. A module for construction of map symbols will be passed by, because standard options are sufficient to represent the amount of working population of Germany in 2006 related to federal states. Size of diagram style circle is proven by the system. Sizes are oriented to smallest and biggest administrative area and amount of working population (in case of dynamic diagram size) and are not allowed to contact or overlap borders of related areas.

A graphical output like Fig. 8 on the whole dataset is first given in the first module of the rendering process. The application of map symbols is done, so visual impression of the potential map is possible. Conflicts in graphical output can be analysed. For correction of cartographic conflicts the following module is implemented. Cartographic generalisation is one of the most important modules in high quality map production process. One problem of a representation of single quantitative data related to administrative borders by diagram could be on side of theme: position and size of diagram-circles (overlaying boundaries) as well as minimum distances that makes modification necessary. On side of basemap simplification of boundaries (terrestrial as
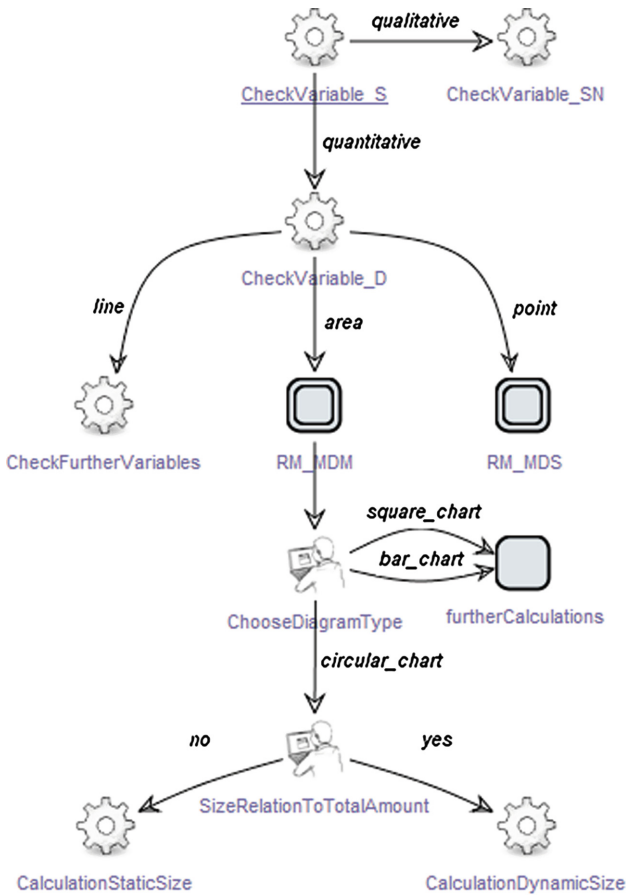
**Fig. 7.**  Implemented process for RM selection in jABC

well as land-ocean-borders) could be necessary. Overall goal of cartographic gener-
alisation is to maximise expressivity, effectivity and adequacy.

Sub-processes, as a result of splitting a complex process, and their further
sub-processes are interrelated. A graphical representation of these divided connections
leads to a hierarchical structure. Before process elements will be implemented by
algorithms, this division into sub-processes will be continued down to service level.
After structuring of logical relationships activity diagrams can arise, that can be shown
by means of UML notation. Sub-processes of the rendering were used exemplarily to
demonstrate the embedding of nested processes in its various hierarchical levels. The
amount of hierarchical levels is not fixed. Within the rendering four hierarchies resulted
(not sure what this means). The detailed design of the map construction process by
means of a map construction assistant is now initiated. The technical interpretation
clarifies the hierarchical construction of the services. A cartographic interpretation will
not be stated in this paper.
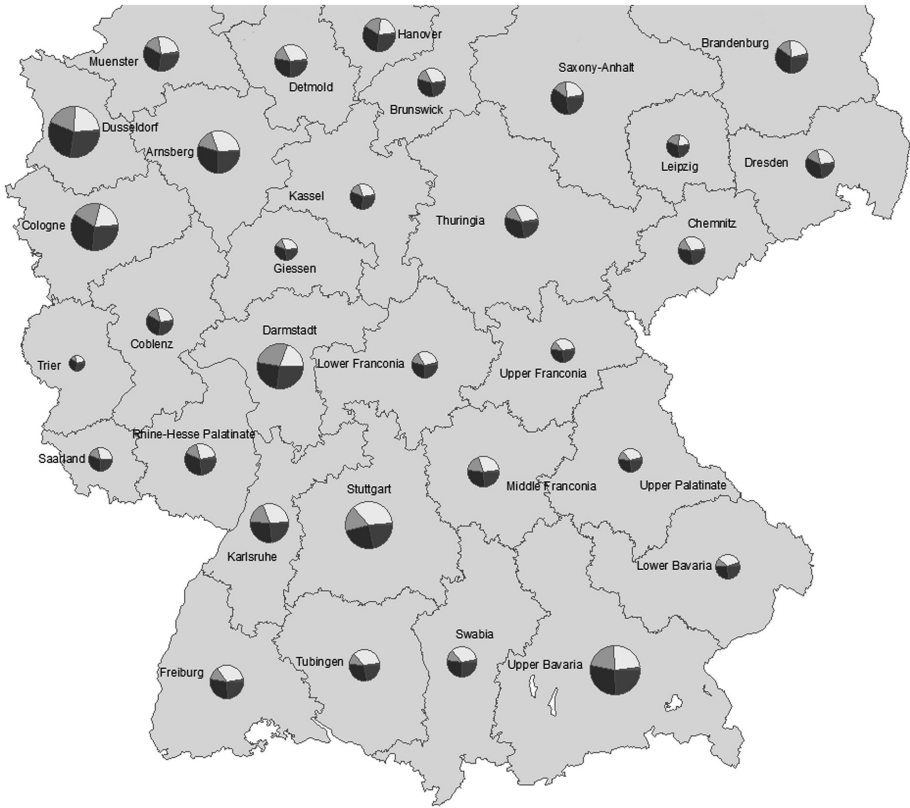
**Fig. 8.** Output of process on given dataset on economic structure of Germany 2006, subdivided into five sectors in relation to working population (excerpt, not scaled)

## 4   Conclusion

The technological concept outlined in this paper is, to our knowledge, novel in the following aspects. First, it integrates the customer into the process via an internet-based portal enabling her to influence the generation of the end product within the scope of professional geodata processing and visualisation principles implemented in the respective modules. Second, a double-track process will be developed to generate either or both graphic or non-graphic geodata products. Due to the software limitations mentioned in chapter 1, such process is not operational in commercial production environments. Third, an automatic data-to-map transformation procedure will be provided. Such rule-based transformation of a filtered geodata model into a corresponding map model by data-to-localised-symbol mapping and generalisation procedures is rarely implemented in commercial map production. Fourth, development and implementation of the production process will be based on software components, such as the jABC (Java Application Building Center) framework. At present, the majority of

geodata production processes is based on commercial software products with the strengths and weaknesses outlined above.

In the early stages of our research, a generic map construction process based on the jABC framework was developed. This needs to be elaborated and extended to cover the complete geodata and map generation process. On a conceptual level, the geospatial data processing and map generation processes will have to be broken down into elementary sub-processes to finally allow for a greater degree of scalability and automation. This will be complemented by an examination of potential system environments. The resulting preliminary process defined and implemented on a working basis will then be evaluated for its functionality and economic viability with real-world sample data from the collaborating cartographic enterprise. Further development will be based on a mix of methods of which software engineering methods, rule-based map construction and geovisualisation principles are the most important.

# References

1. Hardy, P.: High-quality cartography in a commodity GIS: experiences in development and deployment. In: ICA Symposium on Cartography for Central and Eastern Europe [CD-ROM], 16–17 February. Technische Universität Wien/International Cartographic Association, Wien (2009)
2. Buckley, A., Frye, C., Buttenfield, B.: An information model for maps: towards cartographic production from GIS databases. In: Proceedings of the 22nd International Cartographic Conference A Coruna (ICC 2005), Mapping Approaches into a Changing World, [CD-ROM], 9–16 July. International Cartographic Association, A Coruna (2005)
3. Engemaier, R., Asche, H.: CartoService: a web service framework for quality on-demand geovisualisation. In: Murgante, B., Gervasi, O., Iglesias, A., Taniar, D., Apduhan, Bernady, O. (eds.) ICCSA 2011. LNCS, vol. 6782, pp. 329–341. Springer, Heidelberg (2011). doi:10. 1007/978-3-642-21928-3_23
4. Asche, H., Stankute, S., Mueller, M., Pietruska, F.: Towards developing an integrated quality map production environment in commercial cartography. In: Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) ICCSA 2013. LNCS, vol. 7974, pp. 221–237. Springer, Heidelberg (2013). doi:10.1007/ 978-3-642-39649-6_16
5. Simon, M.: Automatisierte Konstruktion thematischer Karten. Kartentypen, Prozessdefinition und Prozesssteuerung. Unpubl. Master thesis, University of Potsdam, Potsdam, pp. 2–24 (2014)
6. Haber, R.B., McNabb, D.A.: Visualization idioms: a conceptual model for scientific visualisation systems. In: Shriver, B., Nielsen, G.M., Rosenblum, M. (eds.) Visualization in Scientific Computing, pp. 74–93. IEEE Computer Society Press, Los Alamitos (1990)

7. Kucharczyk, C.: Konzeption, Entwicklung und Implementierung eines regelbasierten Kartenkonstruktionsassistenten zur fachgerechten Visualisierung statistischer Massendaten. Unpubl. Master thesis, University of Potsdam, Potsdam. pp. 80, 89, 96 (2013)
8. Steffen, B., Margaria, T., Nagel, R., Jörges, S., Kubczak, C.: Model-driven development with the jABC. In: Bin, E., Ziv, A., Ur, S. (eds.) HVC 2006. LNCS, vol. 4383, pp. 92–108. Springer, Heidelberg (2007). doi:10.1007/978-3-540-70889-6_7
9. Hake, G., Grünreich, D., Meng, L.: Kartographie: Visualisierung raum-zeitlicher Informationen. 8., vollst. neu bearb. u. erw. Aufl. Walter de Gruyter & Co., Berlin (2002)
10. Schürer, D.: Modellgeneralisierung, p. 30, 10:17 (2002). http://www.geoinformation.net/lernmodule/lm10/download/vgl_le5.pdf. Accessed 30 Sept 2015
11. Ogrissek, R. (ed.): abc Kartenkunde. 1. Aufl., p. 170. VEB F. A. Brockhaus Verlag, Leipzig (1983)
12. Uthe, A.-D.: Stichwort "regelbasiertes System". In: Bollmann, J., Koch, W.G. (eds.) Lexikon der Kartographie und Geomatik. A bis Z. CD-ROM. Spektrum Akademischer Verlag GmbH, Heidelberg, Berlin (2002)
13. BMVBS Bundesministerium für Verkehr, Bau und Stadtentwicklung; BBR Bundesamt für Bauwesen und Raumordnung (eds.) Automatische Ableitung von stadtstrukturellen Grundlagen und Integration in einem Geographischen Informationssystem, p. 21. Abschlussbericht. Schriftenreihe: Forschungen. Heft 134, Bonn (2008)
14. Breyer, J. (ed.): Haack Weltatlas. GIS-Unterricht mit Atlas und ArcGIS von ESRI. Buch mit CD-ROM, p. 32. Klett Verlag (2010)
15. Asche, H., Kucharczyk, C., Simon, M.: Geodata discovery assistant: a software module for rule-based cartographic visualisation and analysis of statistical mass data. In: Gervasi, O., Murgante, B., Misra, S., Gavrilova, M.L., Rocha, A.M.A.C., Torre, C., Taniar, D., Apduhan, B.O. (eds.) ICCSA 2015. LNCS, vol. 9157, pp. 566–575. Springer, Heidelberg (2015). doi:10.1007/978-3-319-21470-2_41