ORIGINAL PAPER

METHODOLOGY OF EVALUATION AND CORRECTION OF GEOMETRIC DATA TOPOLOGY IN QGIS SOFTWARE

Anita Kukulska✉, Tomasz Salata✉, Katarzyna Cegielska✉, Marta Szylar✉

Department of Spatial Planning and Landscape Architecture, Faculty of Environmental Engineering and Land Surveying, University of Agriculture in Krakow,
ul. Balicka 253c, 30-149 Kraków

ABSTRACT

Geographical Information Systems (GIS) has revolutionised the process of collecting and processing data, therefore, more and more data recorded in an analogue form are transformed into the digital format. However, the process of generating vector models poses a risk of appearing defects of different types. A methodology of correcting common geometric and topological errors that appear in the manual vectorization of a raster model was presented in the paper. The research material was the vector layer including the digitized version of several dozens of drawings of spatial development plans. The paper also presents a procedure for creating a vector model of spatial data with attention paid to potential sources of errors which could be incurred at the stage of its creation as well as indicates methods for their prevention. The tools and plug-ins for evaluation and revision of geometric and topological correctness of a vector model implemented in QGIS software were mainly used in the survey. Elaborated algorithms are aimed at acceleration of data processing to allow their usage during that process. Indeed, proper conducting of spatial analyses needs to administer a data set which is free of errors. Only then, is it possible to obtain proper results and draw appropriate conclusions.

Keywords: geometric data, topology, vectorization, QGIS

INTRODUCTION

The traditional language of cartography that is used to present geographic space is being gradually displaced by Geographical Information Systems (GIS). According to Ślusarski (2012), the demand for spatial information has now intensified. Data recorded in digital form are highly valued due to the possibility to utilize them to perform diverse analyses (Kistowski 1997, Gotlib et al. 2008), both statistical and spatial, with the possibility to their visualization at the same time (Próchnicki 1998). For this reason, GIS also play a increasing role in spatial planning and decision making (Plümer and Gröger 1997). However, to start working with GIS first needs creation of a proper database (Panecki 2015) which means

compiling the obtained information and their proper arrangement (Jucha 2015). As Wolski (2012) emphasizes, there is a fundamental rule *Garbage In, Garbage Out* (GIGO) that forms the basis for spatial analyses. So, the usability of data shall be determined by their credibility and completeness (Feuchtwanger 1989). Data that are fundamental for the conducted analyses should be free of errors, otherwise, they will cause improper results and, as a result, wrong conclusions (Johansson 1990, Kienast 1993, Ubeda and Egenhofer 1997).

Most generally, an error means a mistake, differences between an obtained value and a 'real' or 'expected' one (Kendall and Buckland 1986). Scientific and technological knowledge is affected by errors which influence the accuracy of results. The size of

✉e-mail: a.kukulska@urk.edu.pl, rmtsalat@cyf-kr.edu.pl, cegielska_katarzyna@wp.pl, szylarmarta.kgpiak@gmail.com

such errors depends on the technique of gaining data (Shi 1998), measuring instruments and also experience of an operator (Przewłocki et al. 1994, Jagielski 2003). In case of spatial data, there is also a term of uncertainty (Feuchtwanger 1989). It results from awareness of existing differences between obtained data and 'real' environment because of, among other things, limited precision of measurements or record generalization (Prokop 1998, Felcenloben 2010, Wolski 2012). Gaździcki (2001) indicates that the quality of spatial data (the ability to fulfill needs defined by users) is decided by: origin, compatibility, completeness, and semantic, temporal, positional and attributive accuracy. Only free-of-defects data can fully realize their purpose (Redman 2001). Commonly accepted and used guidelines of spatial data quality are Digital Cartographic Data Standards Task Force (Moellering 1985). Main components of spatial data quality were also determined by international standardization bodies: ISO/TC 211, OGC – Open Geospatial Consortium and FGDC – US Federal Geographic Data Committee. These standards take: lineage, positional accuracy, attribute accuracy, semantic accuracy, temporal accuracy, logical consistency and completeness for an indicator of numerical bases evaluation (Wang 2008). The lineage is a very important feature associated with the history of the data, which directly affects their quality (Prokop 1998). The occurrence of errors in the source material is in fact impossible to eliminate (Goodchild and Gopal 1989, Foody 2001).

Numerical maps that are created by vectorization of cartographic materials are affected above all by topological errors the main sources of which are methods of processing as well as quality of operator's work (Burrough 1986). In order to increase the reliability and the accuracy of GIS analyses, models with improper topology need thorough correction of detected errors (Servigne et al. 2000, Maraş et al. 2010). For this reason, many authors (Laurini and Milleret-Raffort 1994, Ubeda and Egenhofer 1997, Maraş et al. 2010, Derwisz and Zygmunt 2011, Zygmunt et al. 2013) have worked on topological error correcting for ages. As Derwisz and Zygmunt (2011) noticed, manual searching and correcting of such errors is a burdensome and long-lasting process and for that reason there is the necessity to elaborate solutions to make these activities automated.

The aim of the paper is to present the methodology of evaluation and correction of geometrical data topology using tools implemented in QGIS software. In this paper, a procedure for creating a vector model from the analogue maps with attention paid to potential sources of errors as well as indicates methods for their prevention is also shown. As Ubeda and Servigne (1996) emphasize, visualisation is not the main function of GIS these days. Since GIS has become a part of decision making process, spatial queries and reasoning become much more important. The lack of topologically correct database forbids spatial reasoning. Essential operations such as geographical object area calculation are often impossible or their results are not reliable. Creating a vector model with correct structure is aimed at improving the results of data processing.

GEOMETRY AND TOPOLOGY

A map is a set of information – both descriptive and spatial – that present the arrangement of displayed phenomenon (Urbański 2012, Panecki 2015). What testifies to the quality of these set of information is spatial data consistency (Herring 1989). Inconsistencies of a vector model of data which is a digital record of space (Gaździcki 2001) can be of geometric or topological nature (Plümer and Gröger 1997). Several attempts to describe properties of geometric data have been made (Egenhofer and Franzosa 1991, Cui et al. 1993, Ubeda and Servigne 1996, Plümer and Gröger 1997).

In general, standard types of geometry are: point, line – set of segments (segment links two points) and polygon (area, field) – closed sequence of segments (Hodgson et al. 1989, Ubeda and Servigne 1996, Shi 1998, Bouhadjar 2014). Geometry defines objects' location in the accepted reference system and also their shape (Feuchtwanger 1989) when there are lines or polygons. By means of them, the elements of the real world are represented in the spatial model (Izdebski 2015). Duplication of points, overlapping lines and looping – crossing of lines (or contours) with each other are the typical errors of geometry.

On the other hand, topology determines spatial relations between objects (Egenhofer and Franzosa 1991, Urbański 2012). It is a set of rules, which do not depend on inaccurate measurements or imprecise

concepts (Plümer and Gröger 1997). Topology define data correctness and integrity (Mandel et al. 2016). It also describes vicinities, overlap and contain of objects (Chung et al. 1995, Maraş et al. 2010, Bouhadjar 2014). These relations are of qualitative character as well as being permanent during transformations (rotation, scaling) (Gotlib et al. 2007).

The most common topological errors in vector data listed by Maraş et al. (2010) are:
- floating or short lines,
- overlapping lines,
- overshoots and undershoots,
- unclosed and weird polygons.

In the case of simple vector models (non-topological data structure) also called *spaghetti models* (Bouhadjar 2014)*,* structural dependencies between objects do not occur and every object is described separately by means of points in the accepted coordinate system (Medyńska-Gulij 2015). However, the necessity to record the same coordinates in a large number of objects contributes to unfavourable data redundancy (Werner 2004). On the other hand, the topological vector model allows use of a single record of geometrical information which means that data are not duplicated (Iwańczak 2013). Moreover, as Maraş et al. (2010)

and Ładysz (2015) emphasize, topology enables more effective data processing and storage, optimizes using analytical functions, as well as increases spatial analysis opportunities.

CREATING SPATIAL DATABASE

The research material was the vector layer containing the drawings of local plans of spatial development that included over 31 thousand objects (over 177 000 ha). A diagram of creating the layer is presented in Figure 1. QGIS software operating under the license GNU-GPL was used in the paper. The data were obtained as a result of vectorization of the raster layer and they were also imported from other computer services.

Due to the fact that many communities have local plans in analogue form only, a part of the numerical record of plans was obtained by earlier scanning of their paper versions. Currently, the scanning of analog maps into raster formats and then digitizing into vector models is the most common method of producing vector maps (Grimshaw 2000). However, the process of scanning, due to a risk of material's movement, almost always impacts deformation of the original image (Wolski 2012) which significantly complicates the
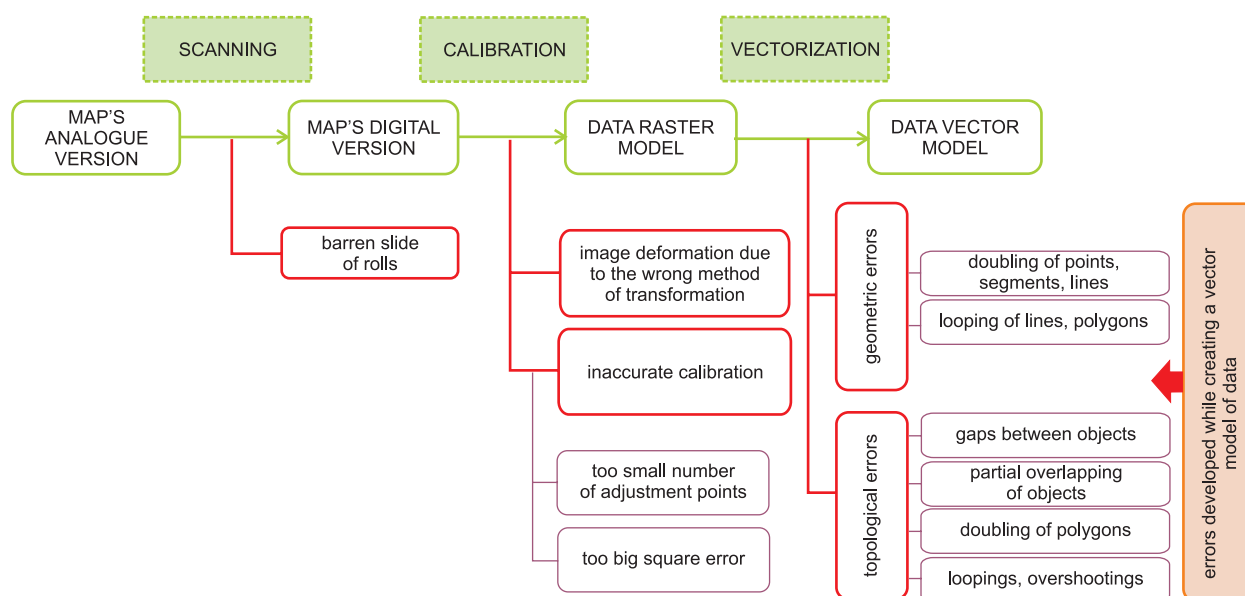
Fig. 1. The diagram of creating a vector model with errors accompanying the individual stages

Source: own study

calibration of such material (see: Fig. 2) so it is also a source of additional errors.

In the further stage of works, in order to obtain a raster model, a digital version of plans was calibrated and they were given georeferences. It means that they were fitted into the vector referential model. The model can represent a numerical record of coordinates of plots, situational details such as buildings' corners, points of measurement control network etc. or the net of coordinate system (in case of the presented survey – the National System of Geodetic Coordinates, in Polish: Państwowy Układ Współrzędnych Geodezyjnych 1992, code EPSG: 2180). The purpose of the calibration was to transform raster images to nominal sizes (Kamocki 1998) and to remove geometrical deformations. Taking data quality into account, two kinds of transformation were used depending on the character of deformations. The linear isotropic Helmert transformation that scales and twists the scan by some an-

gle was used for materials with small deformations (Ładysz 2015). Data obtained by scanning and with larger deformations were subjected to transformation based on spline or rubbersheeting that uses polynomials of higher grades (Jakubczyk 2008, Jaskulski et al. 2013) which transforms the map locally (Iwańczak 2013). The tolerance accepted for root mean square errors (RMS) was 10 pixels.

The raster model prepared in such a way was then subjected to vectorization also called vector digitizing, vector digitization (Medyńska-Gulij 2015), digitization on the screen (Izdebski 2015) or polygonisation (Jucha 2015). As a result, the numerical record of local plans of spatial development were obtained in the vector layer which was subjected to geometrical and topological control.

As Maraş et al. (2010) and Wolski (2012) emphasize, the quality of vectorization is influenced by the operator first of all, accuracy of their work as well as ex-
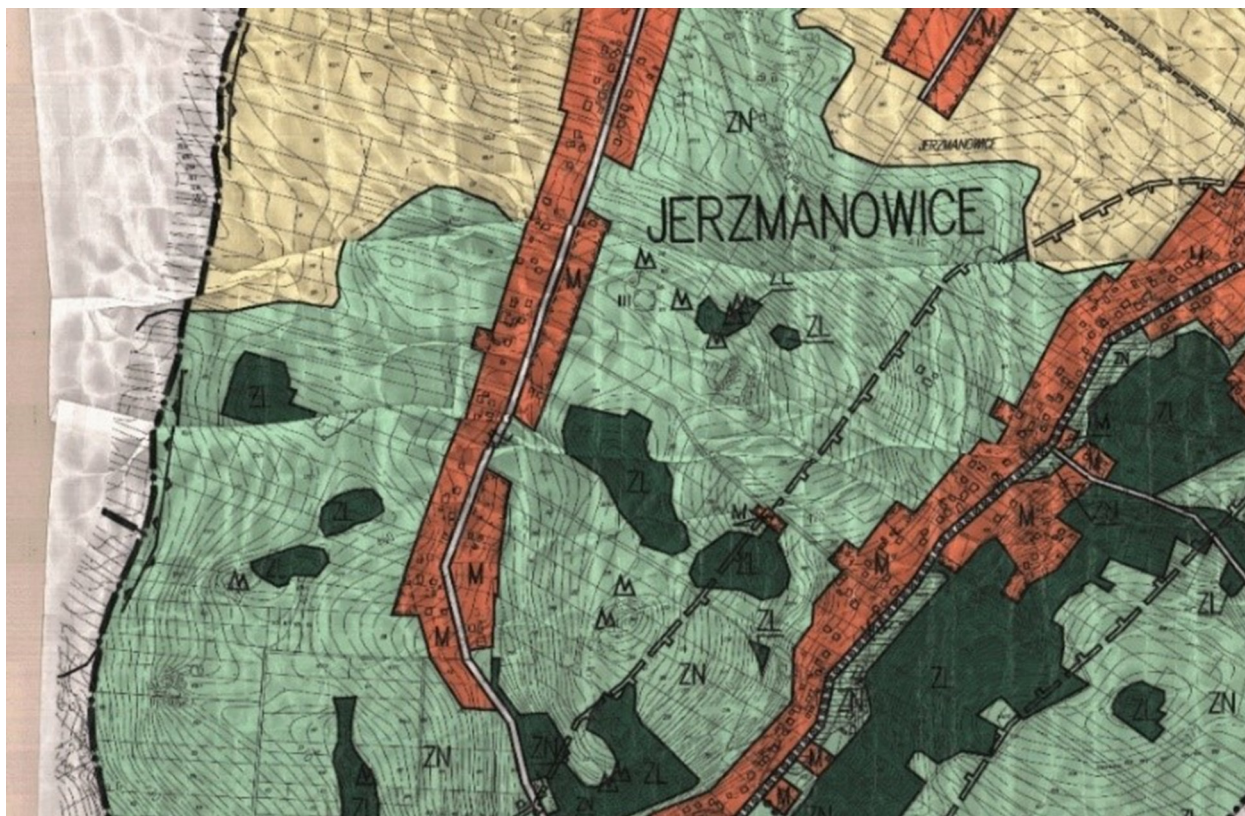
Fig. 2. Deformations of a scanned drawing of the local plan of spatial development

Source: The Municipal Office of Jerzmanowice-Przeginia

perience, which is necessary to proper interpretation of objects and phenomena. At this stage of creating a digital model, they have to define unequivocally the borders of introduced objects and consider topological relations between them (Longley et al. 2008, Felcenloben 2010). So, appearing of errors is almost unavoidable.

The basic errors of digitization listed by Iwańczak (2013) are:

- loopings i.e. inserting double nodes or crossing the line with itself;
- lines not retightened to already existing objects;
- overshoots which means drawing a line through the existing object.

These errors result in objects in a topological model that do not have common edges and imperceptible by the 'bare eye' microgaps appearing from time to time between them.

Using automatic tools that provide some topological correctness of a drawing allows avoidance of some errors. The QGIS programme in the *Snapping options* enables to arrange automated snapping to the vertex or segment (Nowotarska 2009). Consequntly, the new nodes will be located in the place of already existing ones. Work is also easier with the function *Avoid intersections* that performs a function so that new polygonal objects that overlap already existing objects are automatically trimmed to their borders which eliminates the necessity of double digitizing of the same edge and partly automates time-consuming vectorization (Czapiewski 2013). Also, including the function *Enable topological edition* in the window of *Option of pulling* when editing the given node will cause its location in both objects will change at the same time which preserves topological correctness.

THE CONTROL OF GEOMETRIC DATA CORRECTNESS

Every error of vectorization complicates operations on layers and significantly deforms the results of spatial analyses (Wolski 2012). Therefore, correctness of a drawn model should be verified. There are numerous tools and plug-ins (mini-applications) which increase the operative possibilities of QGIS (Iwańczak 2013).

The simplest tool of controlling geometry is the function *Check geometry validity* in the tab *Geometry tools*. It reveals mistakes in geometry of linear and spatial objects such as: doubling nodes, crossing and overlapping lines, generating automatically layers free of these errors. The *Topology Checker Plugin* serves for more advanced topological analyses so it considers not only geometric correctness of objects but most of all relations between adjacent objects. This plug-in allows the user to determine parameters for the analysis of topology, tolerance and relations between layers (Szukała 2015) due to the type of a vector layer. In the case of polygonal layers, the following rules of topology can be checked using this plug-in:

- *must contain* – the polygonal layer has to include at least one element of geometry of the other layer;
- *must not have duplicates* – the objects of the same layer cannot have identical geometries;
- *must not have multi-part geometries* – none of the objects of the layer can be composed of several parts;
- *must not have gaps* – adjacent objects cannot have gaps between them, they have to be adjacent to each other and possess a common border;
- *must not overlap* – the objects from one layer cannot overlap each other;
- *must not overlap with* – the objects from two layers cannot overlap each other;
- *must not have invalid geometries* – all polygons must have proper geometry (QGIS User Guide).

After determining the rules of topology control, the errors are displayed with possibility of individual verifying each of them (see: Fig. 3).

The QGIS programme also operates a tool that undertakes complex evaluation of both geometry and topology of a given vector model. The tool anable to choose a wide range of checking settings. The plug-in is hidden under the same name as the standard function of QGIS *Geometry tools → Geometry Checker Plugin*. In contrast to the other functions of controlling geometry and topology, it verifies errors of not only the whole layer but also the marked objects, exclusively. In the range of geometry control, the user is able to determine many rules of checking from pointing the type of acceptable geometry to other more advanced settings such as: minimal length of a polygon or minimal angle between segments. In the range of topology control, this plug-in will find duplicates and objects located inside other objects and also objects overlapping each other as well as gaps between the objects. There
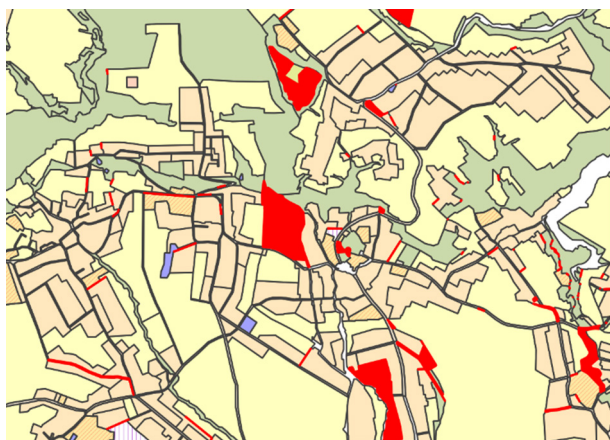
Fig. 3. Results of using the tool *Topology control*

Source: own study

is also the possibility of determine the surface of such overlap or gap.

In progressing the analyses and using the above-mentioned tools, it was revealed that the most frequent errors which appeared during vectorization of the model by inexperienced operators were: objects with improper geometry and gaps between objects or mutual partial overlapping of objects from the same layer.

The most frequent error of geometry is duplicated lines in the polygonal model (see: Fig. 4). Admittedly, these errors often do not disturb the model's visualization but they significantly complicate further spatial analyses. Their identification is also not easy. The basic plug-in of geometry *Check geometry validity* in the tab *Geometry tools* does not read such objects as improper.

Results are only delivered after using the *Geometry Checker Plugin* which by setting the condition of geometry where the minimal angle between segments has to be bigger than 0˚ reads objects constructed in such a way as incorrect. Locating such irregularities is also possible with the option *must not have invalid geometries* which is serviced by the plug-in *Topology control* which is really the tool that should implicitly search for only topological errors.

ALGORITHM OF THE REVISION OF GEOMETRIC DATA ERRORS

Checking of geometric correctness should be performed in the first stage of the revision. Occurrence of objects which are improper geometrically often excludes them from the topological analysis – which should be performed in the second stage: manually, half-automatically or automatically (see: Fig. 5). In the case of QGIS programme, automated reparation of geometric errors is possible by means of the function *Check geometry validity* in the tab *Geometry tools* which generates the geometrically proper layer but it does not catch the errors presented in the Figure 4. Possible automated correction of errors is allowed by the plug-in *Geometry tools → Geometry Checker Plugin* which suggests the option of self-acting reparation for found errors. Simple topological errors can be removed automatically by means of the other tools implemented in the software. Eliminating every duplicate is possible with
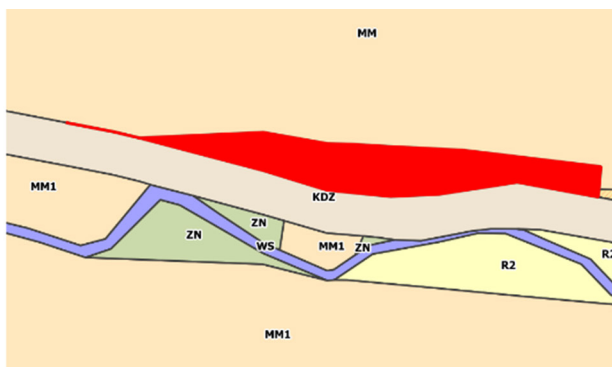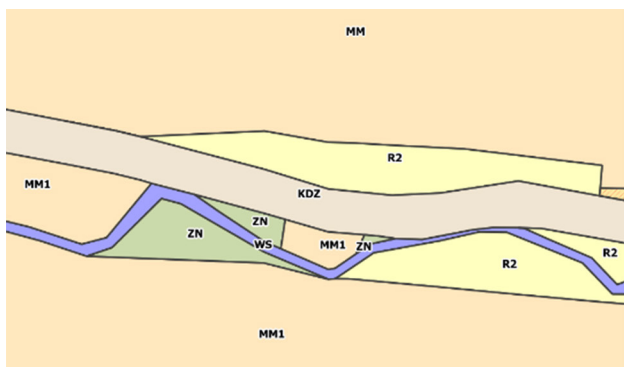
Fig. 4. Error of geometry that does not disturb visualization

Source: own study
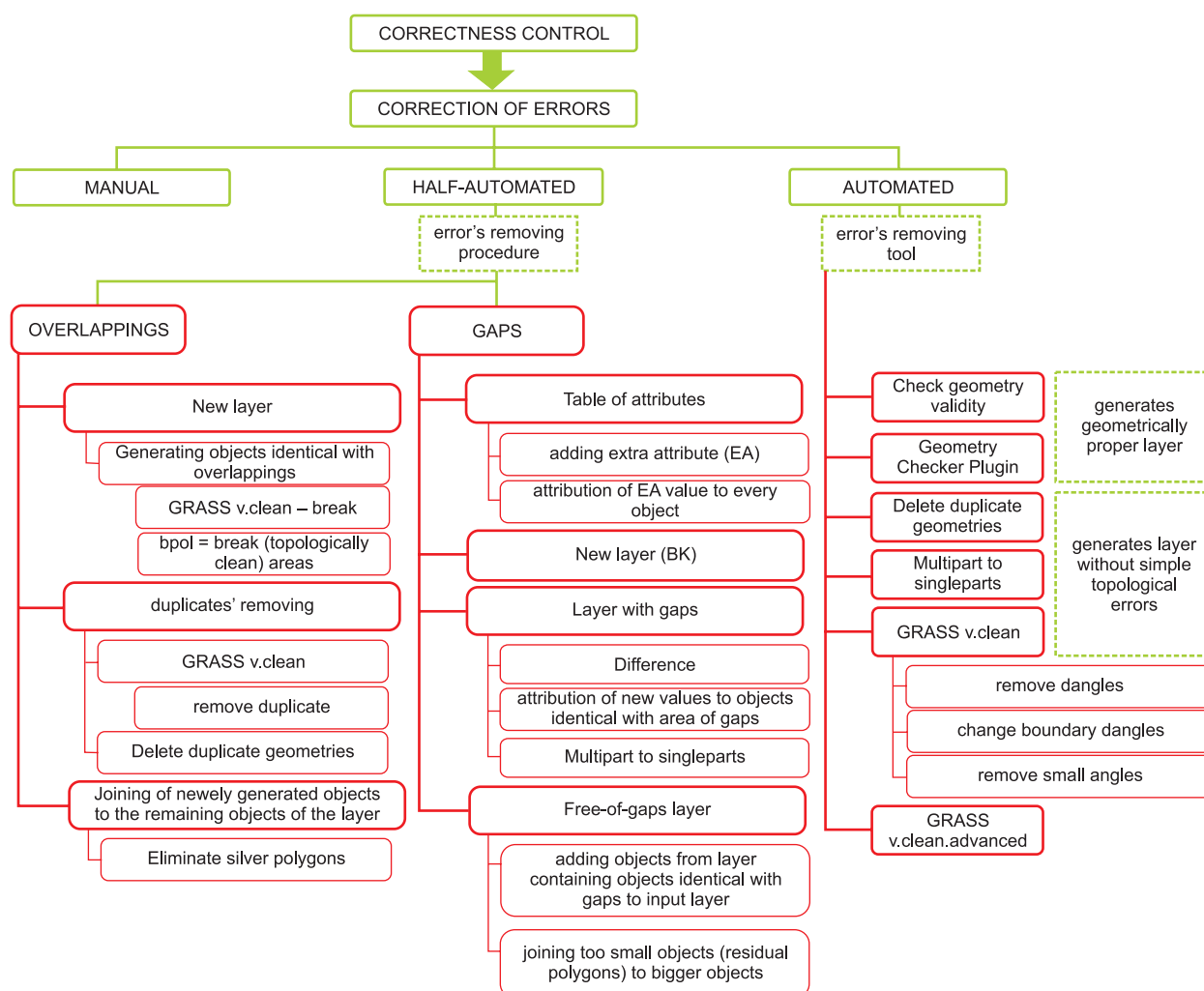
Fig. 5. Methodology of correction of vector model error

Source: own study

the plug-in *Delete duplicate geometries*. Using the function *Multipart to singleparts* allows the user to dispose of geometries that consist of several parts for one-piece objects with the same attributes as the source multi-piece object.

Elimination of the error of geometry presented in Figure 4 is possible by first choosing of the option *Fix checked errors by means of implicit solution* or *NFix checked errors, ask about the answer of answer* in the window of the plug-in *Geometry Checker Plugin*. As a result, the node with an angle equal 0˚ is removed so the overlapping segments of the errant polygon are eliminated. In the next stage, the double node in the

place where the overlapping segments originated is removed. Similar results can be achieved when using the plug-in *GRASS v.clean*. In this case, there are two operations on hand. The first one is *rmdangle* (*remove dangles*) which removes 'hanging' (overlapping) segments. There is also a possibility of determining with the length of segments that should be eliminated (function *Treshold*). *Chdangle* (*change boundary dangles*) is a very similar operation which acts exclusively on polygons (QGIS User Guide, Szukała 2015). This error can be eliminated by means of the algorithm *smsa* (*remove small angles*) which removes zero angles between segments.

The algorithms of half-automated correction of the most popular topological errors that occur during vectorization (i.e. occurrence of gaps between objects and partial overlapping of objects) were presented above (see: Fig. 5).

The source of errors of such type, apart from incorrect vectorization, can be also joining of data sets that come from various sources. The authors paid attention at the possibility of eliminating these errors as there is no specific tool to repair them all and when they are numerous, their manual correction is extremely time-consuming. Despite the genesis of these errors is the same, their correction occurs quite differently.

Gaps between objects

In order to remove the gaps between the objects half-automatically, it is worth starting the work by adding an additional attribute (further: EA – *extra attribute*) to the table of attributes of the adjusted layer (further: FL – *faulty layer*) (see: Fig. 6a) which will serve to classify so called gaps in the next stage. All objects from the layer FL have the same EA value, e.g. 0 assigned. In the next stage, the extra layer (further: BK – *background*) should be generated which will be the background for the input layer filling in the existing gaps. In order to obtain the vector layer which includes only gaps between objects of FL layer, the function *Difference* from *Geoprocessing Tools* should be used. This option subtracts from one another common areas of vector layers. The effect of this action is that the resultant layer includes exclusively these areas of the input layer which were not common with the cutting layer, leaving data from the input layer only in the table (Nowotarska 2009). In that case, the input layer is the layer BK and the cutting one – FL. Using the function *Difference*, the layer that includes the objects with geometry that is identical with the geometry of gaps of FL layer was obtained (see: Fig. 6b). It is worthwhile to ensure at this stage that the value EA should be different from the value EA for the FL layer. EA equal 1 was accepted for the purpose of the present algorithm. Because as a result of using the function *Difference* the resultant layer contains one multi-part object, it should be fragmented into one-piece objects. To this end, the option *Multipart to singleparts* from the standard packet of *Geometry tools* was used. This algorithm generates a new vector layer in which all geometries include only one part (see: Fig. 6c). Next, these objects are added to FL layer and the model without gaps between objects is obtained as a result (see: Fig. 6d).
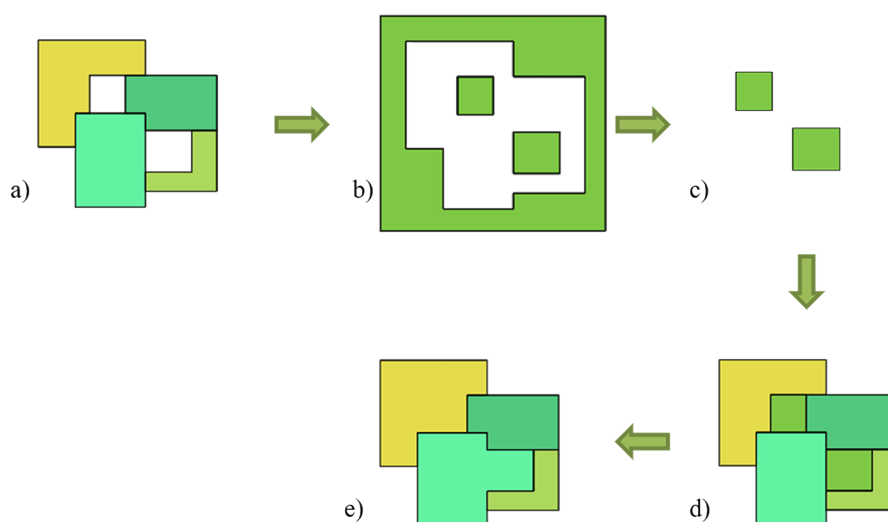
Fig. 6. The procedure of half-automated mass elimination of gaps in the polygonal spatially continuous layer: a) FL layer containing gaps between objects; b) the layer which is a difference between BK and FL layers; c) the layer containing objects – 'gaps' as single objects; d) FL layer with added objects – 'gaps'; e) FL layer with eliminated gaps between objects

Source: own study

Since such gaps have usually very small areas, the model should be arranged in the last stage of works by joining these gaps to the primary objects of FL layer. The function *Eliminate sliver polygons* from *Geoprocessing tools* implemented in QGIS software was used for this purpose. This algorithm improves topology by removing polygons with big disproportions of area in relation to the remaining objects. The user defines the character of objects which should be eliminated by joining them to the adjacent object with the biggest or the smallest area or with the longest border (Chmaj 2014). In the case of the described algorithm, joining the objects that represent the gaps was performed by using the EA attribute of selection and pointing out the value of this attribute for polygons to joining, the value 1 in this case. Thereby, FL layer was deprived of gaps between objects and residual polygons (see: Fig. 6e).
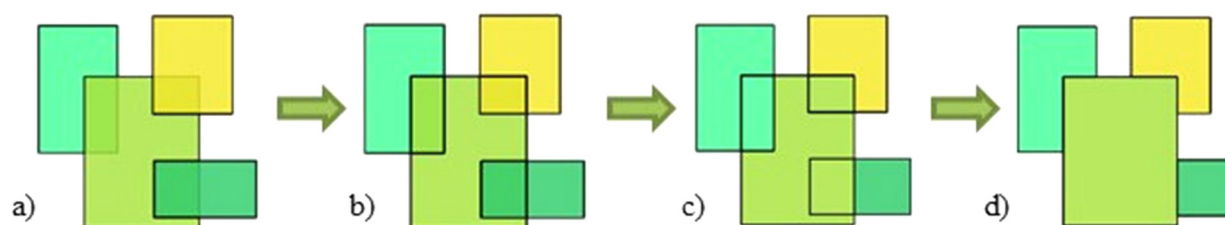
Overlapping objects

The procedure of partial elimination of overlapping objects from one vector layer (see: Fig. 7a) is started by generating objects with geometry that is identical to the geometry of these impositions (see: Fig. 7b). The option *break* serviced by the plug-in *GRASS v.clean* enables this operation. This function breaks the line in every place of cross-cut and also in the loops creating new objects. The function *bpol = break (topologically clean) areas* works similarly as well, except that it does not break the closed loops. As the lines of two objects were intersecting with each other in the place of overlappings, two objects with identical geometry (so called duplicates) were generated in this place. That is why they should have been removed in the next step. These operations can be performed using the next

function *GRASS v.clean* i.e. *rmdupl* (remove duplicate) or the other QGIS tool – *Delete duplicate geometries*. Thereby, the input layer with overlapping objects includes single objects without partial impositions (see: Fig. 7c). The last stage is joining the newly generated objects to the other existing objects of the layer (see: Fig. 7d). These new objects, like the gaps due to the same genesis, have usually small surface. The option *Eliminate sliver polygons* is used for this purpose.

The proposed procedures of half-automated mass elimination of gaps and overlapping objects in the polygonal spatially constant layer were applied on the vector layers including local spatial development plans of: (1) Słaboszów, (2) Sułoszowa and (3) Kocmyrzów-Luborzyca communes. The results of corrections are presented in Table 1.

The applied procedure allows eliminate form 96% to 100% of existing errors, which gives an average of 98% efficiency. The group of correction algorithms in some cases cannot be effective. These are the following reasons:
- there are 'NULL geometric' objects;
- there are polygons with a number of nodes smaller than 3;
- existing multipoligons with rings may have an unfavorable position relative to other layers. This can lead to:
  – creation of geometrically erroneous objects;
  – generating an object with 'NULL geometry';
  – lack of possibility to create an object;
- as a result of processing objects, incorrect objects are created (e.g. cutting, simplifying);
- as a result of determining a non-zero tolerance, the location of close nodes is modified.

Fig. 7. Procedure of half-automated mass elimination of overlapping objects in the polygonal spatially constant layer: a) the vector layer including partially overlapping objects; b) the vector layer with objects with geometry identical with the geometry of 'impositions'; c) the vector layer with removed duplicates; d) the vector layer with the eliminated error of partial overlapping of objects

Source: own study

Table 1. Effectiveness of the procedures of half-automated mass elimination of gaps and overlapping objects in the polygonal spatially constant layer

Type of topological errors	Number of objects on the layer			Number of errors on the layer						The effectiveness of the algorithm		
				before correction			after correction			%		
	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]	[1]	[2]	[3]
Gaps	1192	674	1541	90	40	37	4	0	0	96	100	100
Overlapping				93	112	26	0	2	1	100	98	96

Source: own study

In the last mentioned case, the algorithm may cause overlapping of nodes and the creation of geometrically erroneous objects. This may prevent the correction of the original error in certain cases.

THE OTHER TOOLS FOR TOPOLOGY CORRECTION

The mentioned plug-in *GRASS v.clean* offers numerous options for revealing and correcting errors which allows the user to improve automatically the vector layer's topology. The results similar to these obtained using the tool *Eliminate sliver polygons* from *Geoprocessing tools* can be achieved by means of the function *rmarea* (*remove small areas*), except that it joins the areas only according to the given area expressed in the layer's units (for metric functions it is m^2) to the adjacent area with the longest common border. When the object does not adjoin to any object, it is completely removed. So the function *Eliminate silver polygons* gives more operational possibilities.

The other function from the packet *GRASS v.clean* tj. *snap* pulls the vertices to adjacent nodes which are located within the given distance. It allows elimination of some types of topological errors which manifest by causing gaps between objects. However, using the tool called *snap* can severely damage the topology of a vector model when the parameters are improper. The next algorithm *prune* (*remove vertices*) simplifies the geometry of objects (levels) by removing the vertices located within the given range with preserved topology, at the same time. The plug-in *v.generalize* is characterized by similar functionality, however, having much more possibility to simplify geometry, it does not preserve topology of objects. In the set of tools *GRASS v.clean,* there is also the function *rmline* (*remove lines*) which can eliminate lines with zero length (QGIS User Guide, Szukała 2015).

The plug-in GRASS *v.clean.advanced* allows much more automation of the process of topology correction. In that case, the user is able to determine the number of orders to reveal and correct errors of topology from the packet of options *GRASS v.clean* which will be performed by the programme automatically one after another without choosing each operation separately as it is in the standard version *v.clean*.

The next tool used for topology correction is the script *Fill holes, which* fills in the empty areas in the range of rings, exclusively. The lack of choice of an object to which the filled element would be matched also seems significant in that case.

The function *geometry snapper* from the packet of tools of the plug-in *Geometry Checker Plugin* automatically matches the edges and vertices of one vector layer to the edges and vertices of the other layer using the tolerance defined by the user (QGIS User Guide). However, using this plug-in to improve local plans needs great caution from a user. When the reference layer will be the one that includes the record borders of a property, it should be noted that areas with different purposes do not overlap perfectly with the borders of the plot i.e. several areas with various purposes can exist within one plot. That is why careless use of this plug-in can modify formal arrangements of local plans.

Removing the gaps between objects is also possible using the option *Polygons to lines* from the packet of *Geometry tools* and then the tool *Polygonize* created by Rajmund Szostak. This tool generates polygons from all polygonal shapes limited with lines. Unfortunately, the content of tables of attributes will be

lost as a result of using the combination of these two tools, which in case of the base of local plans of spatial development is unacceptable due to the purpose of land recorded there. Better effects would be possible to achieve if the function *Polygonize* worked also for polygons and not only for linear objects.

The additional tools of analytical geometry of *v. clean* type offer a wide range of operations that improve geometry and topology of objects. However, as Zygmunt et al. (2013) emphasize, topological error elimination methods cannot decrease quality of the processed data. For this reason, users have to use them carefully and properly set the parameters of control and reparation. Otherwise, they will obtain errors which do not disturb visualization but are not compliant with data geometry (Laurini and Milleret-Raffort 1994). This can significantly complicate performing relevant spatial analysis (Servigne et al. 2000).

RECAPITULATION AND CONCLUSIONS

There is no doubt that it is possible to produce vector maps from the analogue data through the QGIS software. The manual digitizing of analogue cartographic data is still in progress all over the world (Zygmunt et al. 2013). However, due to the high possibility of adding extra errors, working on geographic data manually is often avoided (Maraş et al. 2010).

Topological errors are one of the critical issues in GIS researches (Cui et al. 1993), because the topology is a powerful tool for advanced spatial queries (Bouhadjar 2014). Unfortunately, the significant part of existing data sets is vitiated by the lack of proper geometric and topological structures. This problem is noticed, among others, by Ubeda and Servigne (1996), Servigne et al. (2000) as well as Zygmunt et al. (2013). These errors cause results to be unreliable (Ubeda and Egenhofer 1997). Furthermore, the data collection is very costly and time-consuming process. If minimal standards on the quality of such data are violated, the effort spent for their acquisition is futile (Plümer and Gröger 1997). So, having in mind care for the final quality of numerical maps, it is necessary to introduce specific mechanisms of control, evaluation and elimination of errors that automate work at the same time.

Presented algorithms of half-automated mass elimination of gaps and overlapping objects in the polyg-

onal spatially constant layer allow to improve significantly the process of 'repairing' the vector model. QGIS software does not contain tools for automatic correction of the topological errors. Furthermore, manual modifying of models that consist of several dozen thousands of objects is unproductive. Using the combination of available tools of QGIS software, it is possible to increase the efficiency of work and treat the vector model to further usage. Half-automated correction methods also have the advantage over automatic methods that they allow for a certain control of the correctness of performed activities. This enables us to verify the stages of the correction. Such possibilities do not give us an automatic tools. Moreover, the proposed algorithms, with the correct geometry of the polygon layer, guarantee full effectiveness and efficiency of the corrections made. The disadvantage of half-automated methods are their multistage, which is associated with a greater time-consuming than in the case of automatic methods.

Appearance of errors, especially during vectorization of a raster model by an inexperienced operator, is almost unavoidable. According to Chrisman (1991), errors of any kind are an integral part of processing spatial information. That is why they should be treated as the basic indexes of evaluation of set usefulness and not as parameters which disqualify further usability of the given model.

ACKNOWLEDGMENTS

REFERENCES

Bouhadjar, M. (2014). Quality Assessment of Geospatial Data. In: B. Zagajewski, M. Kycko and R. Reuter, eds. European Association of Remote Sensing Laboratories Symposium Proceedings 34th EARSeL Symposium European remote sensing - new opportunities for science and practice. Warsaw 16–20 June 2014. EARSeL & Warsaw University Faculty of Geography and Regional Studies, Germany, 2.1–2.4.

Burrough, P.A. (1986). Principles of geographical information systems for land resources assessment. Monograph

on Soils and Resource Survey, 12. Oxford University Press, Oxford, 193.

Chmaj, K. (2014). Tutorial QGIS: 7. Geoprocessing. Urbnews, http://urbnews.pl/tutorial-qgis-7-geoprocessing (Accessed 15.05.2017).

Chrisman, N. R. (1991). The error component in spatial data. [In:] D.J. Maguire, M.F. Goodchild, D.W. Rhind (eds.). Geographical Information Systems. Longman, Harlow, 165–174.

Chung, M., Cobb, M., Shaw, K., Arctur, D. (1995). An object-oriented approach for handling topology in VPF products. [In:] GIS/LIS '95 Annual Conference and Exposition Proceedings 1. Nashville 14–16 November 1995. American Society for Photogrammetry and Remote Sensing, Tenessee, 163–174.

Cui, Z., Cohn, A. G., Randell, D. A. (1993). Qualitative and topological relationships in spatial databases. [In:] D.J. Abel, B.C. Ooi (eds.). Advances in Spatial Databases – Third International Symposium on Spatial Databases (SSD'93), Singapore 23–25 June. Springer, Berlin, Heidelberg, 296–315.

Czapiewski, S. (2013). Rysowanie – czyli tworzenie obiektów na warstwie wektorowej cz. 2. Krajobraz – GIS – analiza przestrzenna, http://krajobraz.ukw.edu.pl/?p=630 (Accessed 14 May 2017).

Derwisz, J., Zygmunt, M. (2011). Implementacja algorytmu korekty danych wektorowych (Implementation of algorithm of vector data correction). Infrastr. Ekol. Ter. Wiej., 4, 227–232.

Egenhofer, M.J., Franzosa, R.D. (1991). Point-set topological spatial relations. Intern. J. Geographical Information System, 5(2), 161–174.

Felcenloben, D. (2010). Niepewność danych przestrzennych w systemach informacji geograficznej (GIS) (Uncertainty of spatial data in geographic information systems GIS). Acta Sci. Pol., Geod. Descr. Terr., 9(3), 3–12.

Feuchtwanger, M. (1989). Geographic logical database model requirements. Proceed. Auto Carto, 9, 599–609.

Foody, G.M. (2001). GIS: the accurancy of spatial data revisited. Progress in Physical Geography, 25(3), 389––398.

Gaździcki, J. (2001). Leksykon geomatyczny [Lexicon of geomatics]. Polskie Towarzystwo Informacji Przestrzennej, Warszawa, 24.

Goodlich, M., Gopal, S. (1989). Accuracy of spatial databases. Taylor & Francis, London.

Gotlib, D., Iwaniak, A., Olszewski, R. (2007). GIS – Obszary zastosowań. PWN, Warszawa.

Grimshaw, D.J. (2000). Bringing geographical information systems into business. John Wiley & Sons, New York.

Herring, J.R. (1989). A Fully Integrated Geographic Information System. Proceed. Auto Carto, 9, 828–837.

Hodgson, M.E., Barrett, A.L., Plews, R.W. (1989). Cartographic data capture using CAD. Proceed. Auto Carto, 9, 406–415.

Iwańczak, B. (2013). Quantum GIS: Tworzenie i analiza map. Helion, Gliwice.

Izdebski, W. (2015). Wykłady z systemów informacji o terenie, http://www.izdebski.edu.pl/index.php?akcja=pokaz_kat&kat=18 (Accessed 1.02.2017).

Jagielski, A. (2003). Geodezja II. Wydawnictwo P.W. Stabil, Kraków, 173 – 174.

Jakubczyk, K. (2008). Interpolacja funkcjami sklejanymi, http://kaj.uniwersytetradom.pl/prace/Splines.pdf (Accessed 14.05.2017).

Jaskulski, M., Łukasiewicz, G., Nalej, M. (2013). Porównanie metod transformacji map historycznych (Comparison of methods for historical map transformation). Roczniki Geomatyki, XI, 4(61), 41–57.

Johanson, L.B. (1990). Analyzing spatial and temporal phenomena using geographical information systems. Landscape Ecology, 4(1), 31–43.

Jucha, W. (2015). Tworzenie bazy danych do projektu GIS – źródła danych i założenia wstępne (Making of databases in GIS projects – data sources and preliminary assumptions). Prace Studenckiego Koła Naukowego Geografów Uniwersytetu Pedagogicznego w Krakowie, 4, 70–79.

Kamocki, A. K. (1998). Technologia numerycznego opracowania przyrodniczych map tematycznych w oparciu o oprogramowanie Intergraph. [In:] M. Kistowski (ed.). Systemy Informacji Geograficznej w badaniach środowiska przyrodniczego. Gdańsk, 21–23 May 1998, DJ, Gdańsk, 157–162.

Kendall, M.G., Buckland, W.R. (1986). Dictionary of statistical terms. Państwowe Wydawnictwo Ekonomiczne, Warszawa.

Kienast, F. (1993). Analysis of historic landscape patterns with GIS – a methodological outline. Landscape Ecology, 8(2), 103–118.

Kistowski, M. (1997). Systemy Informacji Geograficznej jako narzędzie informowania i wspomagania pracy w nauce i administracji na przykładzie „Cyfrowego atlasu środowiska przyrodniczego województwa gdańskiego” (Geographical Information Systems as a helpful tool of information for research and administration institutions (the case project: A Digital Atlas of Natural Environment of Gdańsk Voivodeship)). [In:] S. Żynda (ed.). Kartografia w ochronie środowiska przyrodniczego i zagospodarowaniu przestrzennym (Cartography

in environmental protection and spatial development). Poznań 13–15 November 1997. Bogucki Wydawnictwo Naukowe, Poznań, 124–134.

Laurini, R., Milleret-Raffort, F. (1994). Topological reorganization of inconsistent geographical databases: a step towards their certification. Computers & Graphics, 18(6), 803–813.

Longley, P.A., Goodchild, M.F., Rhind, D.W. (2008). Geographic information systems and science. PWN, Warsaw.

Ładysz, T. (2015). Technologia GIS w inżynierii bezpieczeństwa. Wyższa Szkoła Oficerska Wojsk Lądowych we Wrocławiu, Wrocław, 33–39.

Mandel, A., Ferrero, V. O., Graser, A., Bruy, A. (2016). QGIS 2 Cookbook. Packt Publishing Ltd, Birmingham, 350–354.

Maraş, H.H., Aktuğ, B., Maraş, E.E., Yildiz, F. (2010). Topological error correction of GIS vector data. Internat. J. Physical Sci., 5(5), 476–483.

Medyńska-Gulij, B. (2015). Kartografia – zasady i zastosowania geowizualizacji. PWN, Warszawa.

Moellering, H. (1985). Digital Cartographic Data Standards: An Interim Proposed Standard. National Comittee for Digital Cartographic Data Standards. Ohio, USA: Numerical Vartography Laboratory. The Ohio State University.

Nowotarska, M. (2009). Wprowadzenie do Quantum GIS. QGIS Polska, http://qgis-polska.org/_media/czytelnia/wprowadzenie_do_quantum_gis.pdf (accessed: 15 May 2017).

Panecki, T. (2015). Modelowanie obiektów topograficznych w bazach danych historycznych (Modelling of topographic objects in historical databases). Acta Univesitatis Lodzensis. Folia Geographica Socio-Oeconomica, 21, 37–51.

Plümer, L., Gröger, G. (1997). Achieving integrity in geographic information systems – maps and nested maps. GeoInformatica, 1(4), 345–367.

Prokop, P. (1998). Jakość danych a błąd i niepewność w systemach informacji geograficznej z punktu widzenia badań środowiska przyrodniczego (Data quality, error and uncertainty in the Geographical Information Systems from environmental research point of view). [In:] M. Kistowski (ed.). Systemy informacji geograficznej w badaniach środowiska przyrodniczego. Gdańsk, 21–23 May 1998, DJ, Gdańsk, 33–42.

Próchnicki, P. (1998). Środowisko przyrodnicze gminy Narewka w postaci map numerycznych. [In:] M. Kistowski (ed.). Systemy Informacji Geograficznej w badaniach środowiska przyrodniczego, Gdańsk, 21–23 May 1998, DJ, Gdańsk,165–171.

Przewłocki, S., Czochański, M., Kowalski, G. (1994). Geodezja i kartografia dla inżynierii środowiska i architektury. Wydawnictwo Politechniki Łódzkiej, Łódź.

QGIS User Guide, https://docs.qgis.org [accessed 1 February 2017].

Redman, T.C. (2001). Data Quality. The Field Guide. Digital Press, Boston, 73.

Servigne, S., Ubeda, T., Puricelli, A., Laurini, R. (2000). A methodology for spatial consistency improvement of geographic databases. GeoInformatica, 4(1), 7–34.

Shi, W. (1998). A generic statistical approach for modelling error of geometric features in GIS. Internat. J. Geographical Information Sci., 12(2), 131–143.

Szukała, P. (2015). Poprawa topologii warstw wektorowych, http://www.pro-gis.pl (accessed: 1 April 2017).

Ślusarski, M. (2012). Propozycja oceny geoportali internetowych poziomu lokalnego (The proposal of evaluation of web geoportals on local level). Infrastruktura i Ekologia Terenów Wiejskich, 1 (III), 109 – 115.

Ubeda, T., Servigne , S. (1996). Capturing Spatial Object Characteristics for Correcting and Reasoning. [In:] M. Rumor, R. McMillan, H.F.L. Ottens (eds.). Geographical Information. From Researh to Application through Cooperation. Joint European Conference and Exhibition on Geographical Information. Barcelona, 27–29 May 1996, IOS Press, Amsterdam, 24–33.

Ubeda, T., Egenhofer, M. J. (1997). Topological error correcting in GIS. [In:] M. Scholl, A. Voisard (eds.). Advances in Spatial Databases – Fifth International Symposium on Large Spatial Databases, July 1997 Springer, Berlin – Heidelberg, 281–297

Urbański, J. (2012). GIS w badaniach przyrodniczych, http://ocean.ug.edu.pl. (accessed: 1 February 2017).

Wang, F. (2008). Handling Data Consistency through Spatial Data Integrity Rules in Constraint Decision Tables. Thesis (PhD). Universitätsbibliothek der Universität der Bundeswehr München.

Werner, P. (2004). Wprowadzenie do systemów geoinformacyjnych. Uniwersytet Warszawski. Wydział Geografii i Studiów Regionalnych, Warszawa, 48–53.

Wolski, J. (2012). Błędy i niepewność w procesie tworzenia map numerycznych (Errors and uncertainty in digital mapping). Prace Komisji Krajobrazu Kulturowego, 16, 15–32.

Zygmunt, M., Siejka, M., Ślusarski, M. (2013). Correction of topological errors in geospatial databases. Internat. J. Physical Sci., 8(12), 498–507.

METODYKA OCENY I KOREKTY TOPOLOGII DANYCH GEOMETRYCZNYCH W OPROGRAMOWANIU QGIS

ABSTRAKT

Systemy Informacji Geograficznej (ang. GIS – Geographical Information System) zrewolucjonizowały proces gromadzenia i przetwarzania informacji, w efekcie czego coraz więcej danych zapisanych w postaci analogowej transformuje się do postaci cyfrowej. Proces generowania modeli wektorowych obarczony jest jednak ryzykiem pojawienia się różnego rodzaju wadliwości. W pracy przedstawiono metodykę korekty popularnych błędów geometrycznych i topologicznych powstałych głównie podczas ręcznej wektoryzacji modelu rastrowego. Materiał badawczy stanowiła warstwa wektorowa zawierająca zdigitalizowaną wersję kilkudziesięciu rysunków miejscowych planów zagospodarowania przestrzennego. Artykuł przedstawia także procedurę tworzenia wektorowego modelu danych przestrzennych, zwracając uwagę na potencjalne źródła błędów mogących powstać na etapie jego tworzenia oraz wskazuje metody służące zapobieganiu im. W badaniu wykorzystano głównie zaimplementowane w oprogramowanie QGIS narzędzia i wtyczki oceny oraz korekty poprawności geometrycznej i topologicznej modelu wektorowego. Opracowane algorytmy mają na celu przyspieszenie procesu uzdatniania danych oraz umożliwienie ich wykorzystania w procesie przetwarzania. Prawidłowe wykonanie analiz przestrzennych wymaga bowiem dysponowania zbiorem danych wolnym od błędów. Tylko wówczas możliwe jest otrzymanie prawidłowych wyników i wyciągnięcie stosownych wniosków.

Słowa kluczowe: dane geometryczne, topologia, wektoryzacja, QGIS