The geographic information systems spatial data quality evaluation in the Metamodel context

T V Shuchkevich
Southern Federal University, Institute of Nanotechnologies, Electronics and Equipment Engineering, 2, Shevchenko str, Taganrog, 347922, Russia
E-mail: tvshushkevich@sfedu.ru

Abstract. The article describes a model of geographic information system developed on the basis of metasystem approach. The problem of the spatial data quality evaluation is considered for this model. An example of the evaluation descriptors selection for a particular dataset case is given.

Introduction
Quality evaluation of spatial data in the metamodell context is described in relation to geographic information systems (GIS) of cadastral registration. Spatial data can be contained in information funds of various levels according to the Federal law N 431-FL “About geodesy, cartography and spatial data and about modification of separate legal acts of the Russian Federation”. Accordingly, there are problems of assessing the quality of data or data sets that are in information funds, as well as the issues of adequate representation of the subject area in the funds at various levels.

Theory
In most cases, a complex multicomponent problem can be solved by the decomposition method. The process of obtaining data, assessing their quality, storing, updating, representing the subject area within the framework of GIS, maintaining and upgrading systems, etc. can be represented as a metamodell that unites subsystems characterizing various aspects of the objects and systems life cycle. The principle of the metasystem approach (System of system - SoS), described in [1], makes it possible to form a common system set-theoretic model [2]:

\[ M = \{M_O, M_E, M_{OE}, M_D, M_{GIS}, Q, M_{DM}, M_U, M_H, A\}, \]  

(1)

where: \( M_O \) is the identifying model of the system (object); \( M_E \) – is the surrounding medium model; \( M_{OE} \) – is the model of system and environment interaction; \( M_D \) – is the behavioral model of system implemented in a simulation of disturbances and control actions; \( M_{GIS} \) – is the geographic information system model; \( Q \) – defines the disturbance/control actions; \( M_{DM} \) – defines the decision making model; \( M_U \) – defines the control system model; \( M_H \) – defines the model of the observer (engineer, expert, operator); \( A \) – is a set of rules of the combined models and the choice of the object modification process.
Model development

From the point of view of the task, the GIS $M_{GIS}$ model is of interest. Representation of $M_{GIS}$ within metamodels, as a rule, does not explicitly include an assessment of data or data sets [3]. The model is described as a set of models measuring the state of the system and the environment.

To assess the data quality, it is proposed to supplement the model with an evaluation subsystem based on the requirements [4]. A tuple of measures for the spatial data can be represented as:

$$M_{s}= \langle DQ_{C}, DQ_{LC}, DQ_{TA}, DQ_{TQ}, DQ_{PA}, DQ_{UE} \rangle,$$

(2)

where: $DQ_{C}$ – is a subsystem of measures characterizing the completeness of the dataset; $DQ_{LC}$ – is a subsystem of measures characterizing the logical consistency; $DQ_{TA}$ – is a subsystem of measures characterizing thematic accuracy; $DQ_{TQ}$ – is a subsystem of measures characterizing temporal quality; $DQ_{PA}$ – is a subsystem of measures characterizing positional accuracy; $DQ_{UE}$ – is a descriptor for opportunity of the correct data use.

All the subsystems specified in (2) are tuples. For example, the $DQ_{C}$ subsystem is represented as

$$DQ_{C} = \langle DQ_{CC}, DQ_{CO} \rangle,$$

(3)

where $DQ_{CC}$ characterizes data redundancy and $DQ_{CO}$ characterizes the omission of the required data items or datasets.

The $DQ_{LC}$ subsystem is represented as the tuple, which combines measures of conceptual ($DQ_{CCon}$), domain ($DQ_{DCon}$), format ($DQ_{FCon}$) and topological ($DQ_{TCon}$) consistency:

$$DQ_{LC} = \langle DQ_{CCon}, DQ_{DCon}, DQ_{FCon}, DQ_{TCon} \rangle.$$  

(4)

All the second-level subsystems have a set of descriptors for the data quality evaluation. The set of descriptors may vary depending on the purpose of the evaluation and spatial data selection.

A collection of measures to evaluate the data obtained during the land survey can be an example of an evaluation descriptors set associated with a dataset (Table 1). This set could be a data provider for cadastral GIS or be extracted from GIS to confirm compliance with the universe of discourse. Table 1 shows the descriptors for the two evaluation subsystems, being only a part of the ongoing assessment of the dataset quality. Descriptor indicators correspond to [4].

The subset can be extracted from the dataset depending on the data evaluation purpose.

The rows in italics in Table 1 are some aggregating data classes. The set of measures to evaluate classes as evaluation objects is usually wider than the set of measures for a single element (item).

The set of critical values corresponding to a set of measures is formed in accordance with the quality evaluation purpose. In determining the critical values, it is advisable to use the approach outlined in [5].
Table 1. Descriptor set

| Class/item | DQ_C | DQ_LC |
|------------|------|-------|
|            | DQ_CC | DQ_CO | DQ_CC, | DQ_DCon | DQ_FCon | DQ_TCon |
| 1. Data on the formed land plots boundary turning point: | 1, 2, 3, 4 | 5, 6, 7 | 8, 9, 12, 13 | 14 - 18 | 19, 20 |
| 1.1. Name of boundary turning points | 4 | 5, 6, 7 | |
| 1.2. Coordinates of boundary turning points of the formed land plots X, Y | 4 | 5, 6, 7 | 8, 9, 10, 12, 13 | 14 - 18 | 19, 20 |
| 1.3. Root mean square error of the boundary turning point position [m] | | 5, 6, 7 | | 14 - 18 | |
| 1.4. Description of a point fixation | | 5, 6, 7 | |
| 2. Data on parts of the formed land plots borders | 1, 2, 3, 4 | 5, 6, 7 | 8 – 10, 12, 13 | 14 - 18 | 21 - 26 |
| 2.1. Name of the borders parts: starting point | 4 | 5, 6, 7 | |
| 2.2. Name of the borders parts: end point | 4 | 5, 6, 7 | |
| 2.3. Horizontal distance [m] | 5, 6, 7 | | 14 - 18 | |
| 2.4. Description of the border part course | 5, 6, 7 | |
| 3. Data on the land plots formation by redistribution | | | | | |
| 3.1. Source of forming: | 1, 2, 3, 4 | 5, 6, 7 | 8 – 10, 12, 13 | |
| 3.1.1. The land plot cadastral number (the cadastral district registration number) | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 3.1.2. Plottage [m²] | | 5, 6, 7 | | 19, 20 | |
| 3.2. Data on the land plots parts (territory) included in the formed land plot structure: | 1, 2, 3, 4 | 5, 6, 7 | 8 – 10, 12, 13 | |
| 3.2.1. Name | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 3.2.2. Plottage [m²] | 5, 6, 7 | | 19, 20 | |
| 4. General information about the formed land plots | | | | |
| 4.1. The land plot name | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.2. The land plot address | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.3. Location of the land plot in a structured form in accordance with the Federal | 1, 2, 3, 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| Information/Address System (in case of the address absence - of the land plot) | 4 | 5, 6, 7 |
|---|---|---|
| 4.4. Additional information about the land plot location | 4 | 5, 6, 7 |
| 4.5. Land category | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.6. The permitted use type | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.7. Plotage ± the value of the error in determining the plotage | 5, 6, 7 | 14 - 18 |
| 4.8. The minimum and maximum size of the land plot [m²] | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.9. Cadastral or other numbers of the real estate objects located on the land plot | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.10. The land plot arbitrary number | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.11. Registration number of the land surveying project | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.12. The number of borders registered | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 4.13. Other data | 5, 6, 7 |
| 5. Data on the land plots adjacent to the formed land plots | 1, 2, 3, 4 | 5, 6, 7 | 8 – 10, 12, 13 | 14 - 18 | 19, 20 |
| 5.1. Name of the boundary turning point or the border part | 4 | 5, 6, 7 |
| 5.2. Cadastral numbers of the land plots adjacent to the formed land plot | 4 | 5, 6, 7 | 14 - 18 | 19, 20 |
| 5.3. Data on the adjacent land plots’ owners | 4 | 5, 6, 7 | 19, 20 |

**Summary**
The article complements the metamodel in the context the SoS-approach for solving the spatial data quality evaluation problems for the cadastral GIS. The approach allows formalizing the evaluation model, creating tuples of the evaluation descriptors measures for spatial data collections evaluation standard problems.

**References**
[1] James E Campbell, Dennis E Longsine, Donald Shirah, and Dennis J Anderson 2005 *System of Systems Modeling and Analysis* (Springfield, VA).
[2] Ginis L A, Gorelova G V, Kolodenkova A E 2016 *Cognitive and Simulation Modeling of Development of Regional Economy System* (International Journal of Economics and Financial Issues) 6 (5S) 97-103.
[3] Gordienko L V, Dmitrieva I A, Popova O V 2016 *IT in the Educational Programs in the Field of: “Safety of Person’s Activity”* (2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT): proceedings, Baku), pp. 764-767, Oct 2016.

[4] ISO 19157:2013 *Geographic Information - Data Quality*.

[5] ISO 3951-1:2013 *Sampling procedures for inspection by variables -- Part 1: Specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection for a single quality characteristic and a single AQL*.

**Acknowledgment**

The work was supported by Russian foundation for basic research, project № 19-01-00412 А «Development of models and algorithms of cognitive and neuroevolutionary modeling to support decision - making in intelligent information management systems». 