

CITY-SCALE GIS-BASED MONITORING OF INFECTIOUS DISEASE – CONTEMPORARY ISSUES, CASE STUDY OF ST. PETERSBURG, RUSSIA

DOI: <http://dx.doi.org/10.18509/GBP.2020.87>
UDC: 616.9:[528.93:004.6(470)]

Ilia Kuznetsov¹
Evgeny Panidi¹
Vladislav Korovka²
Vladimir Galkin²

¹ Saint Petersburg State University, **Russia**

² Saint Petersburg State Research Institute of Phthisiopulmonology, **Russia**

ABSTRACT

Medical geography and medical cartography have a long history. These domains cover both the tasks of retrospective monitoring and analysis (e.g., human population health monitoring, medical infrastructure development and availability assessment, etc.), and operative identification and management of emergencies (e.g., monitoring of epidemics development and infectious diseases spread). The same problems can be attributed also as the classical issues of Geographical Information Systems (GISs) implementation into applied domains. At the same time, implementation of particular geospatial facilities and technologies when such monitoring is performed on a city or municipality scale has its own specific, and always turns out to be significantly dependent on the architecture and technology of the health management system in a given city (or municipality).

Our research is devoted to the investigation of the possibilities for GIS-based support of medical administration and planning on a city scale and performed on the example of Saint Petersburg city (Russia). Basing on the medical data collection system developed and used by Saint Petersburg medical administration, we are working on the design of a GIS-based monitoring system prototype that can be used for infectious diseases monitoring, and can allow accumulation of georeferenced medical data, study of spatiotemporal dynamics of disease rate, and disease patterns identification. Currently we are identifying challenges and problems allocated at the system conceptualizing and design stage. Medical data integration into GIS database is discovered in the paper.

Keywords: Medical Cartography, Infectious Disease, GIS-based Monitoring, Geocoding, Geospatial Databases

INTRODUCTION

Possibilities for geographic interpretation and geospatial (GIS-based) analysis of medical data in general [1], [2], [11], [12] and infectious disease data in particular [4], [5], [7] for decades compose an one of popular research topics, which is discussed and discovered by geographers, cartographers and GIS specialists. There is great potential in GIS application to the domain of health care. However, graphical (and particularly cartographic) facilities for medical data representation remains limited and undeveloped, due to the complexity of medical data (while a long list of health parameters can be used generally to characterize situation, and the statistics is dynamical in time and space). Many authors recognize GISs as an effective instrument for retrospective [3], [6], [10] and operative [8], [9] medical data analysis and complex representation, but medical

(health) map usually appears to be quite simple in the meaning of used cartographic techniques.

On the other hand, simplified data representation can be valuable to ensure decision support made by non-cartographer (medical specialist). In this plane, it is an issue how to synthesize and generalize initial medical data series to provide clear and valuable data visualization. This issue assumes that data have to be well-structured to be applicable for deep analysis, complexing and consequent estimation of synthetic metrics for medical or epidemiological (for instance) conditions.

Our study is at an early stage and aimed on support of epidemiological monitoring of the Saint Petersburg city territory, while the geospatial analysis and cartographic representation generally not used and are almost not familiar to involved medical specialists. Preliminarily provided experiments were conducted on city district medical statistics dataset (Admiralteysky district data were used), which incorporated depersonalized data on tuberculosis, hepatitis and HIV (Human Immunodeficiency Virus) incidents referenced by carrier ID, time and postal address (included the house numbers, but not flat numbers). Experiments shown that the issue of common cartographic representation of these three infectious diseases data on the city (city district) scale is accompanied with a number of obstacles, due to the dynamics in time and spatial inhomogeneity. In addition to this, probably more significant issue is the data reprocessing (at the into-GIS loading stage) and structuring, while results of these operations are extremely valuable in the meaning of correctness and flexibility of cartographic representation refinement.

Regarding this, initial problem consists in georeferencing of the medical data. Problem context is composed by diversity of initial data structuring, need of address geocoding while integrating into GIS, and ambiguity of geometry type selection that is used to visualize data on the map. Medical staff involved in initial medical documents formation, data depersonalization and preparation for into-GIS loading, generally may have no clear understanding of full data processing chain, especially in the meaning of postal address structuring. So in general case, dataset collected from medicals cannot be processed and incorporated into GIS directly.

Most common errors are:

- 1) Grammatical errors (unavoidable for large amounts of data);
- 2) Addressing errors (addressing system can be not clear in some cases, and the errors occurred due to manual address entry may appear and lead to incorrect allocation of data particle);
- 3) Data structuring errors (appear due to the unsystematic documents filling in some cases).

DATA AND METHODS

Preliminary experiment was focused testing of the address geocoding capabilities of the medical statistics used in our study. We used address database publically available on Federal Informational Address System (FIAS) Web site (<https://fias.nalog.ru>). This database being an official informational resource was used as reference standard of city street addresses. Initial table composed of observed inflectional disease incidents was joined to the FIAS tables basing on address fields presented in the initial table. Additionally, the GIS vector layer composed of geometry and attributes of city buildings was joined to the FIAS tables (also basing on street address fields). Unjoined records of the initial medical statistics, and equally records of vector layer were then filtered and

checked for the address description correctness, with consequent correction of found errors.

All the joining and filtering operations were organized in PostgreSQL (<https://www.postgresql.org>) database, as it is GIS compatible free and open source database management software. To provide the operations a number of queries was designed, which helped to correct errors partially in automated mode through the flexible joining rules.

Finally, the database composed of three abovementioned data sources was exported into GIS layers (SHP format was used, data were presented in QGIS - <https://qgis.org>), where geometry was derived from initial GIS layer, street address – from FIAS database, and medical data from initial table of medical statistics. This approach helped to prepare maximally correct dataset, but was extremely time consuming due to the need of queries formation, test and adjustment, and to the need of manual correction of detected errors.

In parallel, we discovered two available in QGIS geocoding modules, but both were recognized as not effective in the project context. RuGeocoder module (https://plugins.qgis.org/plugins/ru_geocoder/) is built upon the version 2.0 of the Python programming language that is not supported in actual versions of QGIS. MMQGIS module (<https://plugins.qgis.org/plugins/mmqgis/>) had no options to create errors report easily, and then became inapplicable as no initial data could be lost. Additionally, the number of geocoding errors produced by the module was up to 30% (Fig. 1), which was recognized as unsatisfactory.

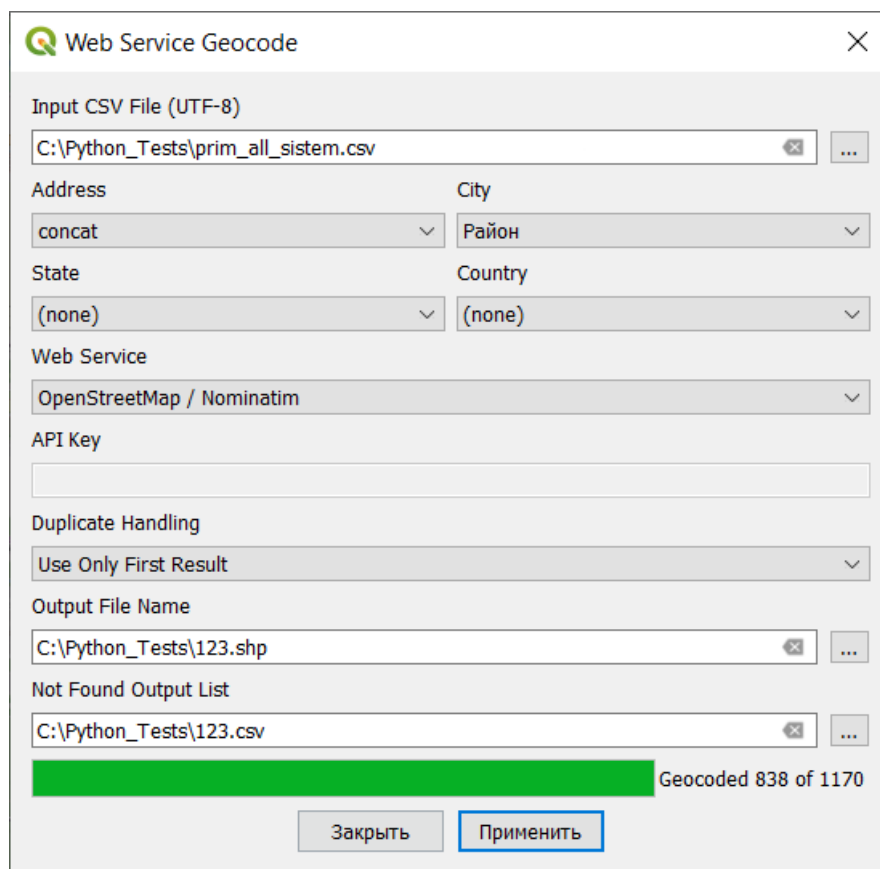


Figure 1. Results of test dataset geocoding by MMQGIS module, error rate is 28% (332 of 1170 records)

To reduce geocoding time and make geocoding process flexible in the meaning of project needs, we implemented prototype of geocoding module for QGIS designed accordingly to these needs (able to generate error reports and to provide geocoding in batch and single-record modes). Nominatim service (<http://nominatim.org>) was used as geocoding engine, which operates OpenStreetMap data to recognize street address locations.

RESULTS AND DISCUSSION

Developed geocoding module incorporated graphical user interface (Fig. 2) to ensure files opening and saving, and designed to read and write data from/to CSV-formatted text files. When executed, the module reads CSV file, parses the data into array and separates address date using Slice operation. One address can be geocoded in one second, as it is a limit set by the OpenStreetMap user agreement. Geocoded records are supplemented with coordinates-containing fields and written into new CSV file. If geocoding error occurred, an additional CSV file is created and stored, where all the error record are collected. This allows to control geocoding results, correct errors and re-geocode remaining particles of data.

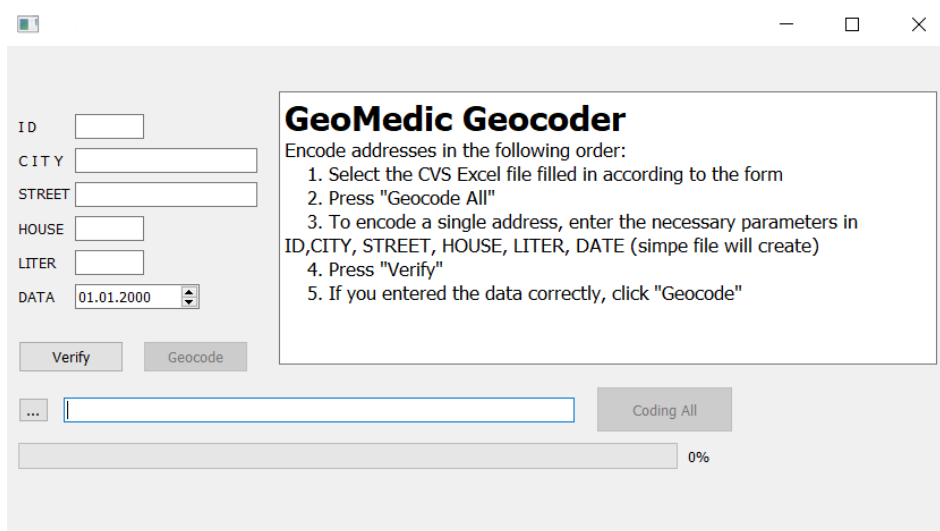


Figure 2. Graphical user interface of the developed geocoding module prototype

Medical statistics dataset can be geocoded directly through reading from CSV file, or manually (record by record) using graphical interface (when error correcting, for example). Produced dataset can be read directly in QGIS and drawn on the map using coordinates stored in addition to medical data.

Therefore, the generated data can drawn on the map as a point geometry layer. However, in a number of cases it can be needed to highlight the buildings on the map where the incidents were observed. To provide this opportunity, we derived vector layer of buildings from OpenStreetMap dataset and run vector intersection operation for point incident and polygon building geometries. This operation helps to provide additional address control, as in some cases points can be allocated not within building geometry (Fig. 3). Additionally, this operation makes it possible to structure medical data accordingly to the GIS needs, as initially different incidents observed at one address are mapped by a number of points. Then the intersection operation is accompanied by joining of the point geometry attributes to polygon geometry in many-to-one manner, and joined buildings can be drawn on the map (Fig. 4).

processing, but have to be accompanied by double control, as at different stages data structuring errors can appear.

Provided experiments and prototyping have shown that the lack of compatible ready to use processing tools are presented in desktop GIS software. Future work have to be devoted to the enhancement of geocoding facilities, through embedding visual geocoding (made by manual pointing on the map) and possibly combining OpenStreetMap-based geocoding and FIAS-based address control into one processing chain.

Also, the experience of dynamic visualization of produced data highlights the needs of specialized analog of Time Manager module development, as it has some restrictions in data visualization and a number of not demanded functions and interface elements. While the data model used to organize the data and built accordingly to the Time Manager demands have to be redesigned, as being able to visualize data accordingly to the time measure, we are not able currently to do the same using number of incidents as a measure, or by combining time and number of incidents measures.

REFERENCES

- [1] Chistobayev A.I., Semenova Z.A. Medico-geographical mapping in the former USSR and modern Russia. *Vestnik Sankt-Peterburgskogo Universiteta, Seriya Geologiya i Geografiya*, 2013, No. 4, pp. 109-112. in Russian
- [2] Gatrell A.C., Bailey T.C. Interactive spatial data analysis in medical geography. *Social Science and Medicine*, 1996, Vol. 42, Issue 6, pp. 843-855. doi:10.1016/0277-9536(95)00183-2
- [3] Gordon A., Womersley J. The use of mapping in public health and planning health services. *Journal of Public Health*, 1997, Vol. 19, Issue 2, pp. 139-147. doi:10.1093/oxfordjournals.pubmed.a024601
- [4] Huang X., Wang J. Analysis on the learning hot spots about AIDS evidence-based medical evidence abroad based on SPSS knowledge map. *Geomatics and Information Science of Wuhan University*, 2012, Vol. 37, Issue SUPPL.2, pp. 70-73.
- [5] Jeefoo P., Tripathi K.N. Dengue risk zone index (DRZI) for mapping dengue risk areas. *International Journal of Geoinformatics*, 2011, Vol. 7, Issue 1, pp. 53-62.
- [6] Lesnykh S.I., Mel'nikova, O.V. Generation of databases and visualization of current epidemiological information for purposes of a medical-ecological monitoring of a region. *Geography and Natural Resources*, 2019, Vol. 40, Issue 2, pp. 115-121. doi:10.1134/S1875372819020033
- [7] Malkhazova S.M., Mironova V.A., Pestina P.V., Orlov D.S. Emerging and re-emerging infections in Russia: A medico-geographical aspect. *Vestnik Moskovskogo Universiteta, Seriya 5: Geografiya*, 2016, No. 5, pp. 24-32. in Russian
- [8] Mayer J.D. The role of spatial analysis and geographic data in the detection of disease causation. *Social Science and Medicine*, 1983, Vol. 17, Issue 16, pp. 1213-1221. doi:10.1016/0277-9536(83)90014-X
- [9] Qi Y., Guo K., Zhang C., Guo D., Zhi Z. A VGI-based foodborn disease report and forecast system. *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Safety and Resilience*, 2018, Article number a18. doi:10.1145/3284103.3284124
- [10] Richterich A. Digital health mapping: Big data utilization and user involvement in public health surveillance. *Geographies of Digital Culture*, 2017, pp. 144-185. doi:10.4324/9781315302959_10
- [11] Schweikart J., Kistemann T. Mapping health and health care [Kartographie der Gesundheit]. *Kartographische Nachrichten*, 2013, Vol. 63, Issue 1, pp. 3-11. in German
- [12] Stampach R., Konecny M., Kubicek P., Geryk E. Dynamic cartographic methods for visualisation of health statistics. *Lecture Notes in Geoinformation and Cartography*, 2010, Issue 199089, pp. 431-442. doi:10.1007/978-3-642-03294-3_27