

GIS-BASED INFECTIOUS DISEASE DATA MANAGEMENT ON A CITY SCALE, CASE STUDY OF ST. PETERSBURG, RUSSIA

I. Kuznetsov¹, E. Panidi^{1,*}, A. Kolesnikov², P. Kikin³, V. Korovka⁴, V. Galkin⁴

¹ Department of Cartography and Geoinformatics, Institute of Earth Sciences, Saint Petersburg State University, St. Petersburg, Russia - panidi@ya.ru, e.panidi@spbu.ru

² Siberian State University of Geosystems and Technologies, Novosibirsk, Russia

³ Peter the Great St. Petersburg Polytechnic University (SPbPU), St. Petersburg, Russia

⁴ Saint Petersburg State Research Institute of Phthisiopulmonology, St. Petersburg, Russia

KEY WORDS: Medical Cartography, Infectious Disease, GIS-based Data Management, GIS-based Monitoring

ABSTRACT:

Medical geography and medical cartography can be denoted as classical application domains for Geographical Information Systems (GISs). GISs can be applied to retrospective analysis (e.g., human population health analysis, medical infrastructure development and availability assessment, etc.), and to operative disaster detection and management (e.g., monitoring of epidemics development and infectious diseases spread). Nevertheless, GISs still not a daily-used instrument of medical administrations, especially on the city and municipality scales. In different regions of the world situation varies, however in general case GIS-based medical data accounting and management is the object of interest for researchers and national administrations operated on global and national scales. Our study is focused onto the investigation and design of the methodology and software prototype for GIS-based support of medical administration and planning on a city scale when accounting and controlling infectious diseases. The study area is the administrative territory of the St. Petersburg (Russia). The study is based upon the medical statistics data and data collection system of the St. Petersburg city. All the medical data used in the study are impersonalized accordingly to the Russian laws.

1. INTRODUCTION

At the beginning of the 21st century, infectious diseases remain a major global problem. Governments try to fight diseases, but don't succeed always. Recent incidents have shown that densely populated regions can become the source of global and dramatic epidemics. In particular, the COVID-19 epidemic has already caused huge damage to most countries. More than 3 million people have been infected worldwide and more than 200,000 of them have died as of end of the April 2020, accordingly to the Johns Hopkins University¹.

It is becoming obvious that the fight against infectious diseases assumes improving of sanitary conditions, increasing of medical specialists' skills and qualifications, and renewing of medical equipment. One of expected trends is the development of medical technologies applied to monitoring and forecasting of socially dangerous diseases. Forecast of the disease growth significantly saves money and other resources. It is much easier to prevent an epidemic than eliminate devastating consequences.

Russia cannot be an exception in this context. However, while the COVID-19 can be defeated, only by mid-summer 2020² (as it expected on spring 2020, accordingly to the TASS Russian state news agency), the fight against other diseases will may take decades.

On December 1, 2004, the Government of Russia approved the list of dangerous diseases (Russian Government, 2004). This list includes tuberculosis, human immunodeficiency virus (HIV), hepatitis, genital infections, diabetes, cancer, and a number of mental illnesses. The Annex to the document states that GIS technologies can be used for monitoring and forecasting tasks.

At the moment, GIS-based analysis of infectious disease data is actively discovered by researchers (Jeefoo, Tripathi, 2011;

Huang, Wang, 2012; Malkhazova et al., 2016). However, in the case of St. Petersburg city there are no GIS tools implemented to solve the tasks of diseases monitoring. Integration of GIS applications into medical administration processes would be a great add-on to existing medical information systems and databases used to track and fight diseases.

2. APPROACHES, DATA AND METHODS

Medical geography (Gatrell, Bailey, 1996) and medical cartography (Chistobayev, Semenova, 2013; Schweikart, Kistemann, 2013; Stampach et al., 2010) are quite popular research domains. Both, retrospective (Gordon, Womersley, 1997; Richterich, 2017; Lesnykh, Mel'nikova, 2019) and operative (Mayer, 1983; Qi et al., 2018) analysis of medical statistics data are studied. However, implementation of mapping technologies and GISs into health care system remains poor, while demanded at the level of specialists involved into diseases monitoring. Fig. 1 shows an attempt to elaborate map of infection cases, and similar examples we met at a number of medical institutions. This trend demonstrates serious concern of medical specialists and administrators in finding convenient tools for spatial tracking and analysis of diseases.

The study is conducted at the area of St. Petersburg city (European part of Russia, 60°N, 30°E). Medical statistics data on tuberculosis, HIV and hepatitis were selected as the study object, as these infections are discovered as associated (dependant one on another) by medicals, and are relevant for all the big settlements.

Taking into account the need to implement GIS-based mapping tools into not GIS-qualified domain, we established the study aim

* Corresponding author

¹ <https://coronavirus.jhu.edu/map.html>

² <https://tass.ru/obschestvo/8138947>

as prototyping of a GIS application able to support monitoring of infectious diseases.

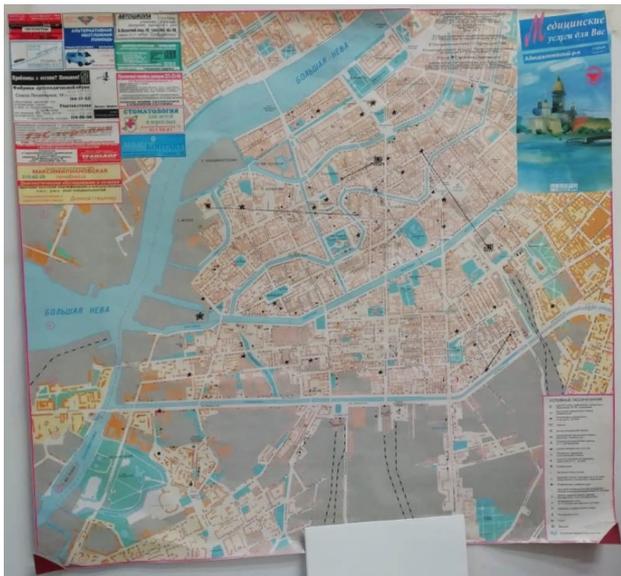


Figure 1. Tourist map is used to show spatial distribution of tuberculosis patients (Admiralteysky district of St. Petersburg)

The requirements to the application were elaborated collaboratively with medical practitioners as follows:

1. Friendly interface; training of a medical staff is not expected, so the interface should be clear and informative
2. Free of charge; budgets of medical institutions do not include funds for such projects currently
3. Integration with some universal GIS; to ensure wider capabilities of data analysis
4. Data accumulation and mapping facilities implemented accordingly to the currently used structuring of medical data

Basing on the abovementioned approaches, we used QGIS³ as a basic software platform and implemented the application in the form of QGIS module.

Base map data were derived from OpenStreetMap⁴. Map data were used as underlying map when visualising the medical statistics and, additionally, as geometries source when medical statistics data geocoding.

3. GEOCODING OF THE MEDICAL DATA

To incorporate medical statistics data into GIS we have to provide its geocoding. As these data originally are not coordinated and cannot be mapped directly. The data are associated by postal addresses, and due to this can be coordinated using traditional postal-address-based geocoding techniques in GIS software.

Despite the availability of a number of geocoding modules for QGIS (e.g., MMQGIS⁵, RuGeocoder⁶), we decided to integrate geocoding functionality directly into the developed application (module). It helped to minimize data processing chain, and eliminated some functionality problems presented in available modules. For example, MMQGIS do not allow control of the

address correctness directly during the geocoding process before adding processed data to the geocoded results. RuGeocoder on the other hand, was developed in Python 2.0 programming language, that is not supported in current versions of QGIS.

As we need both, to ensure geocoding of the already collected medical statistics data, and to provide possibility for direct collection and geocoding the data is GIS interface, geocoder interface was prototyped to support these two options (Fig. 2). It supports loading data from file for batch geocoding of retrospective data, and manual postal address filling for new coming data.

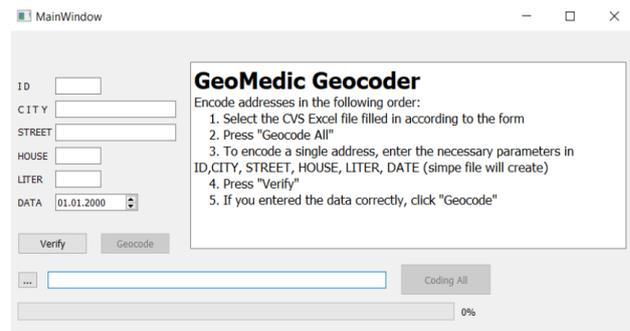


Figure 2. Graphical user interface of prototyped geocoding module for QGIS

Official address database in Russia is compiled and managed by the Federal Tax Service. It is named as Federal Address Information System⁷ (FIAS – as in Russian language the word Information is written before the word Address) and available publically and freely on the Web. A street address is composed of region name, region district name, city (or town/village) name, street name, house number, house block number, and letter (that also can be used for house blocks numbering). Last two items are not obligatory.

Initial data were composed of Microsoft Excel table filled by medical data, buildings Shapefile derived from OpenStreetMap, and address database file derived from FIAS. We used FIAS data as an etalon, and provided linking of medical data to the map geometries through the FIAS data (so in fact, both medical statistics table and geometries table were linked to the FIAS table). To harmonize our geospatial database with official (FIAS) addressing system, we used PostgreSQL⁸ for linking addresses in medical statistics dataset with addresses attributed to building geometries on the map. After this, the addresses were geocoded and attributed with geographic coordinates using the prototyped module.

The module plays interfacing and organizing role, while the geocoding itself (linking of the FIAS table to map geometries) was made by means of Nominatim⁹ search and geocoding engine for OpenStreetMap. At the first step, point geometries were generated from geocoded coordinate pairs, and the points were intersected (joined spatially) with building area geometries at the second step. At the end, we harvested point and area datasets attributed with medical data, such as patient age, sex and infectious status with time marks (Fig. 3).

³ <https://qgis.org>

⁴ <https://www.openstreetmap.org>

⁵ <https://plugins.qgis.org/plugins/mmqgis/>

⁶ https://plugins.qgis.org/plugins/ru_geocoder/

⁷ <https://fias.nalog.ru>

⁸ <https://www.postgresql.org>

⁹ <http://nominatim.org>



Figure 3. Geocoded map layers of point (red markers) and area (green polygons) geometry

Thus, we implemented one-window application able to provide geocoding in manual mode (one-by-one address) with correctness control on each iteration, or in automated mode with final control of geocoding errors. The application is built upon free and open source Nominatim geocoding engine and needs internet connection to access OpenStreetMap data server when geocoding.

The only restriction is applied to geocoding process, that is one geocoding operation for one postal address per one second is possible. This limit is established by Nominatim usage policy¹⁰. All the medical data were impersonalised before we got access to it accordingly to the Russian Federal law (Federal law #152, 2006). This data feature has to be commented separately. It may look that presence of postal address associated with sex and age in database item can be enough to deanonymize a patient. However, databases of people living addresses are not public in Russia, so from formal point of view this deanonymizing is not possible. Additionally, bearing in mind that the study area is a city inhabited by more than 5 million of people migrating time by time, and built up with multi-floor and multi-apartment houses, we may conclude that deanonymizing process cannot be easy in this case (even when accessing to the living addresses database).

4. MAP VISUALIZATION OF MEDICAL DATA

As the infectious disease data has time dimension, we provided dynamic map visualisation with the help of QGIS Time Manager¹¹ module. In fact, this module is used as a second part that composes designed application prototype, alongside with the developed geocoding module. Time Manager makes it possible to turn on and turn off geometries on the map accordingly to their time marks stored in the attribute table, and to slice the time dimension into needed segment. It adds also the time slider bar (Fig. 4) to the QGIS graphical user interface to provide map animation control.

¹⁰ <https://operations.osmfoundation.org/policies/nominatim/>

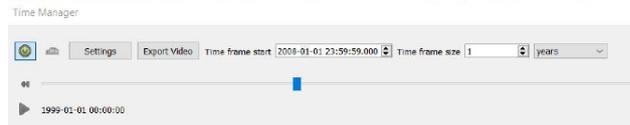


Figure 4. Time slider bar of the Time Manager module

To collapse doubled geometries (in the cases when more than one infection case were registered at the address) we processed our geocoded shapefiles and computed the number of patients at each address for each year. Initially, we elaborated data for the Admiralteysky district of St. Petersburg. In this area, we had data on HIV, tuberculosis, hepatitis B and hepatitis C collected over 20 years. This dataset is up to the most complete in terms of medical statistics.

Basing on computed attributes and time slicing capabilities of the Time Manager, we produced basic map series for every disease to highlight its time dynamics (Fig. 5 and Fig. 6). The map series is implemented in two forms, as a series of static maps, sliced for every year, and as the dynamic animated map, controlled through the time slider bar of the Time Manager.

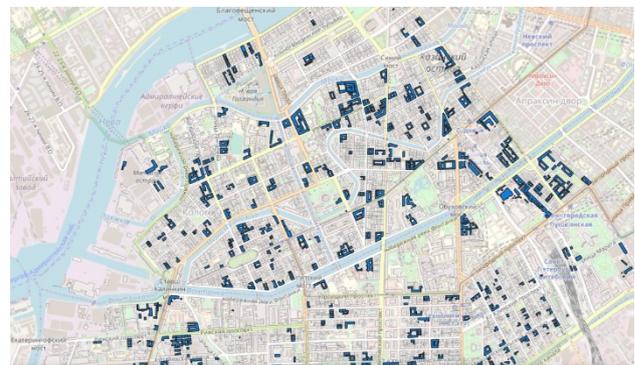


Figure 5. Visualization of tuberculosis cases reported in 2008 in Admiralteysky district of St. Petersburg (base map © OpenStreetMap contributors)

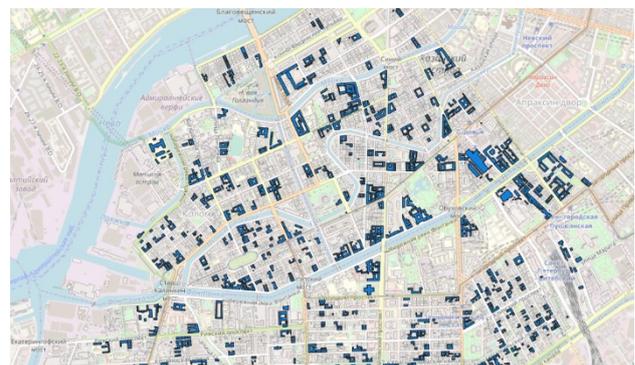


Figure 6. Visualization of tuberculosis cases reported in 2018 in Admiralteysky district of St. Petersburg (base map © OpenStreetMap contributors)

Additionally, while studied infectious diseases are discovered as associated and common visualisation of (for example) HIV and tuberculosis is valuable for medical staff, we composed these data on another one map series (Fig. 7).

¹¹ <https://plugins.qgis.org/plugins/timanager/>

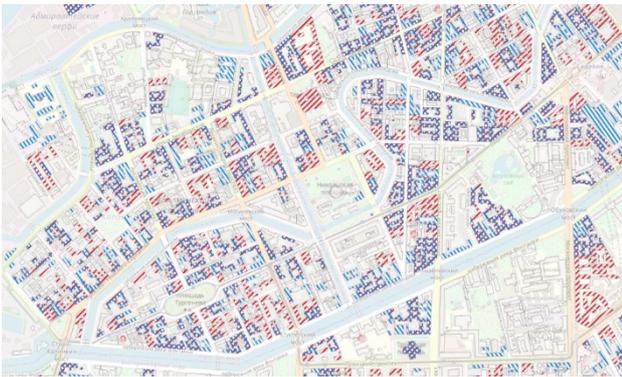


Figure 7. Joint mapping of HIV (red hatching) and tuberculosis (blue hatching) cases reported in 2019 in Admiralteysky district of St. Petersburg (base map © OpenStreetMap contributors)

These maps have a potential to be involved when forecasting the epidemic situation. Accordingly to the World Health Organisation the tuberculosis is one of leading causes of death of the HIV infected people¹². In years of monitoring, it is observed also that people infected by HIV are 34 times more likely to be infected by tuberculosis. Produced map visualisations make it possible to study spatial features of these interdependences in mapped area, and provides additional opportunity to automate estimation of multiple infecting rate.

At next stage, we processed data for the Primorsky district of St. Petersburg (Fig. 8 and Fig. 9). While the Admiralteysky district composes historical center of St. Petersburg, where old buildings with dorm-type apartments are presented frequently, Primorskiy district has a lot of new multi-apartment housing and larger population, but lower population density. Medical statistics data in this case were collected over 2006-2019. These maps show clearly the differences in spatial patterns of infectious disease in historical center and peripheral districts of the city.

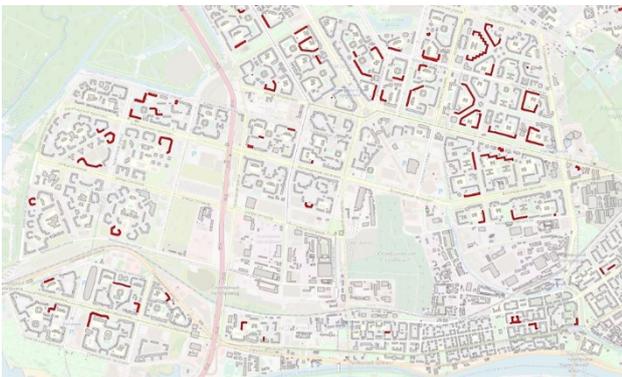


Figure 8. Visualization of HIV cases reported in 2010 in Primorsky district of St. Petersburg (base map © OpenStreetMap contributors)

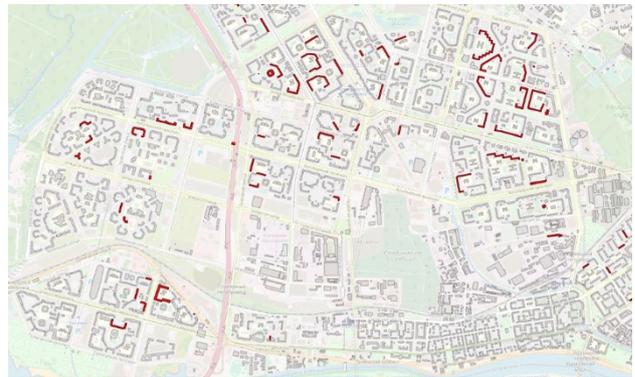


Figure 9. Visualization of HIV cases reported in 2018 in Primorsky district of St. Petersburg (base map © OpenStreetMap contributors)

To visualize the number of infected people over all the years (or other period) of monitoring, point geometries can be used. In the Fig. 10 size of the point markers depend on the number of infected people at the address.

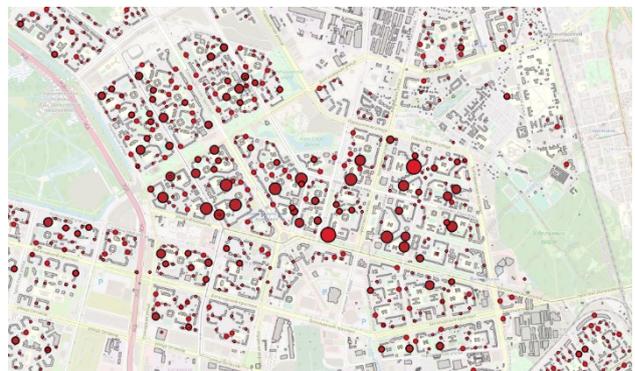


Figure 10. Visualization of HIV cases reported in 2006-2019 in Primorsky district of St. Petersburg; size of markers correspond with the amount of reported HIV cases (base map © OpenStreetMap contributors)

5. CONCLUSION

At the current stage of the study, we collect first feedback on our prototype. Our results produce mixed reaction from various experts, as being spatially visualised, these data can be perceived terrifying, especially by nonqualified person. Despite this, we need to cover the entire city of St. Petersburg in our geospatial database to judge on its potential for real-life use. Some preliminary conclusions were made basing on gained results in the medical plane also:

1. District-scale mapping cannot confirm the theory of so-called "floating focuses of the disease" (proposed by medicals), i.e. cannot confirm displacement of disease focuses in space and time; Whole the area of the city have to be mapped in the long term
2. Joint mapping of tuberculosis and HIV allowed to determine the centers of high epidemiological danger (while HIV-infected people are the most tuberculosis-susceptible)
3. Retrospective mapping cannot be applied in practice to prevent the development of infectious diseases; In

¹² <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>

particular, specialists in Phthisiology claim the need of investigations in the plane of disease forecast and predictive mapping

4. Medical data statistics have to be represented also in relative values (density per square kilometre, and etc.); There are no formulations in the field of sanitary and epidemiological legislation in Russia related to absolute indicators

Russian Government, 2004. The Order of the Russian Federation Government of December 1, 2004 No. 715, On approval of the list of socially significant diseases and the list of diseases that pose a danger to people (in Russian)

Schweikart J., Kistemann T., 2013: Mapping health and health care [Kartographie der Gesundheit]. *Kartographische Nachrichten*, 63(1), 3-11 (in German)

Stampach R., Konecny M., Kubicek P., Geryk E., 2010: Dynamic cartographic methods for visualisation of health statistics. *Lecture Notes in Geoinformation and Cartography*, 199089, 431-442. doi:10.1007/978-3-642-03294-3_27

Finally, future work have to be devoted to the QGIS interface customizing aimed on simplification for potential end users. This will help also to design out-of-the-box product by excluding all not-needed functionality of graphical user interface. Another one future work direction is the integration of cartographic analysis tools into existing medical information systems.

REFERENCES

Chistobayev A.I., Semenova Z.A., 2013: Medico-geographical mapping in the former USSR and modern Russia. *Vestnik Sankt-Peterburgskogo Universiteta, Seriya Geologiya i Geografiya*, 4, 109-112 (in Russian)

Federal law on personal data #152 of 27.07.2006 (in Russian)

Gatrell A.C., Bailey T.C., 1996: Interactive spatial data analysis in medical geography. *Social Science and Medicine*, 42(6), 843-855. doi:10.1016/0277-9536(95)00183-2

Gordon A., Womersley J., 1997: The use of mapping in public health and planning health services. *Journal of Public Health*, 19(2), 139-147. doi:10.1093/oxfordjournals.pubmed.a024601

Huang X., Wang J., 2012: 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*, 37(SUPPL.2), 70-73

Jeefoo P., Tripathi K.N., 2011: Dengue risk zone index (DRZI) for mapping dengue risk areas. *International Journal of Geoinformatics*, 7(1), 53-62

Lesnykh S.I., Mel'nikova, O.V., 2019: Generation of databases and visualization of current epidemiological information for purposes of a medical-ecological monitoring of a region. *Geography and Natural Resources*, 40(2), 115-121. doi:10.1134/S1875372819020033

Malkhazova S.M., Mironova V.A., Pestina P.V., Orlov D.S., 2016: Emerging and re-emerging infections in Russia: A medico-geographical aspect. *Vestnik Moskovskogo Universiteta, Seriya 5: Geografiya*, 5, 24-32 (in Russian)

Mayer J.D., 1983: The role of spatial analysis and geographic data in the detection of disease causation. *Social Science and Medicine*, 17(16), 1213-1221. doi:10.1016/0277-9536(83)90014-X

Qi Y., Guo K., Zhang C., Guo D., Zhi Z., 2018: A VGI-based foodborn disease report and forecast system. Proceedings of the 4th ACM SIGSPATIAL International Workshop on Safety and Resilience, a18. doi:10.1145/3284103.3284124

Richterich A., 2017: Digital health mapping: Big data utilization and user involvement in public health surveillance. *Geographies of Digital Culture*, 144-185. doi:10.4324/9781315302959_10