

ISSUES OF GEOGRAPHIC INFORMATION SYSTEMS AND THEMATIC MAPPING APPLICATION TO ANALYSIS OF EPIDEMIOLOGICAL SITUATION IN LARGE CITIES

I. Kuznetsov¹, E. Panidi^{1,*}, P. Kikin¹, A. Kolesnikov², V. Korovka³, V. Galkin³

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

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

ABSTRACT:

The paper we summarize and discuss experience of medical statistics data processing and mapping, gained in 2019-2020 in the frames of study devoted to the tuberculosis infection mapping. Basing on this experience, we have formalised a set of research issues, which were elaborated and clarified at the previous stages but demand additional investigations.

Additionally, the paper summarises results of design and prototyping of a Web mapping interface implemented as a part of developed medical Geographic Information System (GIS). Developed GIS is aimed onto mapping and analysis of tuberculosis infection data. The overall structure of elaborated GIS is covered also with respect to the detected research issues.

1. INTRODUCTION

More than 144 million people have been infected worldwide and more than 3 million of them have died (as of mid of April 2021) during Coronavirus (SARS-CoV-2) pandemic of 2019-2021 (accordingly to the Johns Hopkins University – <https://origin-coronavirus.jhu.edu>). The pandemic highlights the need of geospatial analysis implementation (Franch-Pardo et al., 2020) to observe, account and spread prevention epidemic development. While people mobility appears higher and higher in recent decades, monitoring and study of infection spread is applicable (and have to be applied) on different scales from global to city scale. Obviously, the geospatial analysis and mapping of the infection spread and its spatial dynamics (Chistobayev, Semenova, 2013; Gatrell, Bailey, 1996; Gordon, Womersley, 1997; Mayer, 1983; Wei et al., 2020) appear relevant and significant supporting technologies in such a context.

In 2019-2021 we conducted a set of case studies (Kuznetsov et al., 2020) in the domain of GIS-based processing and mapping of infectious diseases statistics data. Conducted studies were aimed onto elaboration of approaches to support of medical administrating and planning of infectious diseases accounting and controlling. The studies were performed on the example of St. Petersburg city located in the North-West of Russia (second largest city in Russia). Medical statistics data were obtained from medical administration of the city. The data were used in depersonalized form accordingly to the Russian law on personal data (Federal law, 2006).

2. METHODOLOGY

Within the study framework, we applied GIS software to support and automate structuring, processing, mapping and analysis of initial medical statistics data collected in St. Petersburg (by city tuberculosis dispensaries) on socially valuable diseases during last decades. The list of the socially valuable (in Russia) diseases is established by Russian Government order (Government of the

Russian Federation, 2004) and incorporates 16 items (including SARS-CoV-2, cholera, malaria, etc.). In our studies, we operated with medical statistics data on tuberculosis (that were main object of interest), hepatitis and human immunodeficiency virus (HIV). During 2019-2020 we designed a methodology for preprocessing of initial medical statistics data, its geocoding and representation in geospatial database implemented in QGIS (<https://qgis.org>) open source desktop Geographic Information System. Basic on-the-map visualisation techniques were discovered also (Kuznetsov et al., 2020). Being focussing onto support of infectious diseases monitoring and analysis made by medical professionals and administrators, now we are providing an effort to implement Web-based GIS interface to operate with collected data and maps. Taking into account that operation with GISs and maps is not standard competence of medicals and administrators, we discover Web interface as maximally user friendly and intuitive. So, the Web interface have to eliminate interdisciplinary barrier for medical colleagues, while desktop GIS have to remain as an interface for involved GIS professionals.

Currently we are working on Web interface prototyping. The interface is developed accordingly to the dashboard development methodology (Dong et al., 2020; Wei et al., 2020; Martorell-Marugán et al., 2021) that appears popular in the Coronavirus study support projects. This methodology assumes complex but intuitively perceptible visualisation of medical statistics data using maps and infographics. To illustrate the approach, such worldwide famous examples as Johns Hopkins University COVID-19 dashboard (<https://coronavirus.jhu.edu/map.html> – Figure 1) and the dashboard of the Regional Office for Europe of the World Health Organisation (<https://who.maps.arcgis.com/apps/opsdashboard/index.html#/a19d5d1f86ee4d99b013eed5f637232d>) can be mentioned, as well as in-Russia famous Yandex (Russian IT-giant) COVID-19 dashboard (<https://datalens.yandex/7o7is1q6ikh23> – Figure 2).

* Corresponding author



Figure 1. Web interface of the COVID-19 dashboard elaborated by Johns Hopkins University.

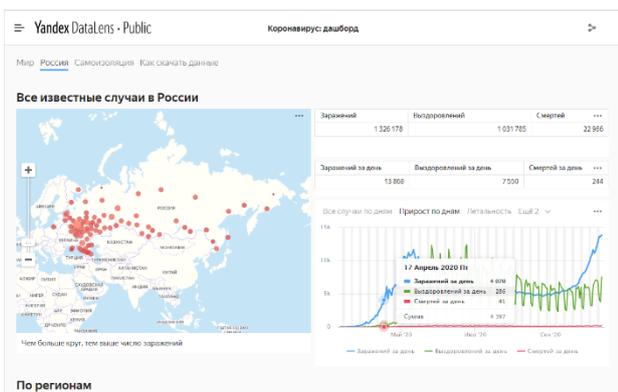


Figure 2. Web interface of the COVID-19 dashboard elaborated by Yandex company.

3. EPIDEMIOLOGICAL MAPPING ISSUES

While the GIS-based automation of data processing and mapping is applied, we faced also a number of issues connected with the data character we are dealing with.

First of most significant issues we detected is the postal addresses geocoding accuracy. We developed the GeoMedica geocoder application (built upon OpenStreetMap Nominatim geocoding engine) to provide mapping of initial medical statistics data georeferenced by postal addresses (Kuznetsov et al., 2020). Geocoding accuracy we gained was above 78% (Figure 3), which is estimated as an appropriate by medical professionals involved into the project. However, the geocoding errors possible to be corrected to gain this accuracy were corrected in manual mode that is very time consuming. Nevertheless, a part of geocoding errors registered during the geocoding of initial medical statistics data cannot be resolved even in manual mode due to the fundamental incorrectness of postal address (absence of a part of address record or fictitious address).

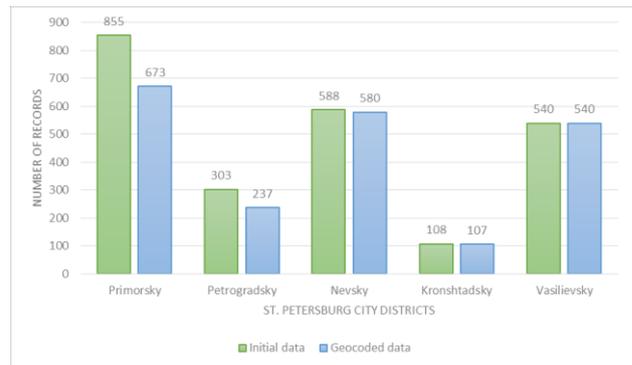


Figure 3. Proportion of the initial medical statistics data records and geocoded data records, by administrative districts.

Second of the issues is the necessity of analysis methodology development and redesign. In a number of cases, working hypotheses on medical statistics and its spatial referencing were faulty. For example, maps of infection cases density (Figure 4) demonstrates different spatial pattern in comparison to the maps of infection cases amounts (Figure 5). So the definition of infection hotspots becomes not rigorous.

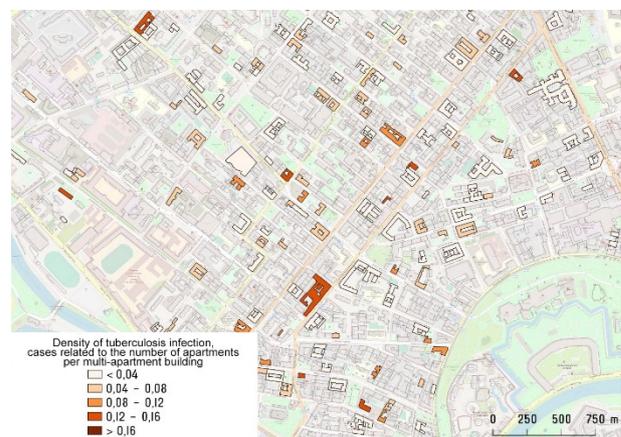


Figure 4. Example of tuberculosis infection density map (cases related to the number of apartments per multi-apartment building).

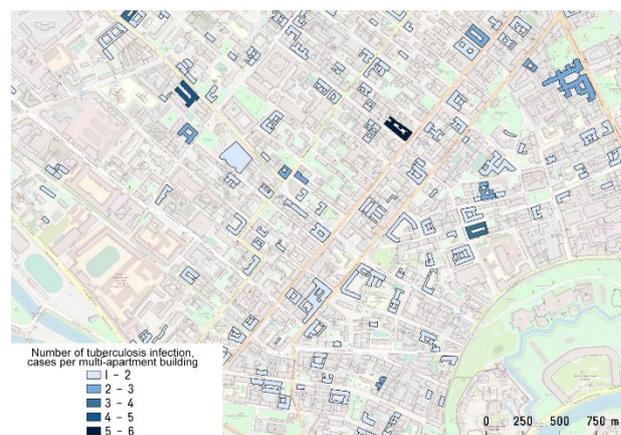


Figure 5. Example of tuberculosis infection cases map (number of cases per multi-apartment building).

Third of the main issues we can mention is the legislation issue. While we operate quite special medical data it is better to have clear rules guided what types of mapped medical statistics and mapped results of its analysis can be (or have to be) published in open access, and what types cannot be. There are no direct legislative restrictions on such maps publishing presented currently in Russian laws. On the other hand, no direct and explicit rules of publishing are presented also. For example, the Federal Law on biological safety (Federal law, 2020) states that informing of local authorities and citizens on the activities provided against the biological hazards (including diseases fighting) is the prerogative of regional authorities; while the same law establishes the need of scientific activities organization in the domain of biological safety. Moreover, the president decree (President of the Russian Federation, 2019) denotes the need of geographic information systems development to support rapid response to biological threats.

4. RESULTS

The data we are operating in GIS are derived through geocoding process from initial medical statistics data collected by city tuberculosis dispensaries (each serves one of administrative districts of the city). All database records related to administrative districts through the geocoded postal address and equipped with time marks marking a period of infection case observation. These features of the database give possibility to estimate and map 9 parameters of the tuberculosis infection spread, which are used by city tuberculosis dispensary when providing operational activities:

1. The population amount in the administrative district (at the end of the calendar year), formed from the Russian federal statistics agency (Rosstat)
2. The tuberculosis infection rate (a relative indicator that reflects a number of newly detected tuberculosis infection cases per 100,000 citizens; the cases newly detected during the treatment or after the death of the patient are accounted)
3. The tuberculosis infection spread rate (a relative indicator that reflects a number of tuberculosis infection cases under the supervision of dispensary medicals per 100,000 citizens)
4. The death rate due to the tuberculosis infection (a relative indicator that reflects the number of death cases due to the tuberculosis infection per 100,000 citizens)
5. The number of newly detected cases of tuberculosis infection
6. The number of citizens detected as newly infected by tuberculosis (including posthumous cases)
7. The number of citizens detected as infected by tuberculosis and HIV
8. The number of patients with low drug resistance
9. The epidemiological burden (the rank of epidemiological danger assigned of the observed territory, it is formed on the basis of the ranks of all the above mentioned indicators)

The geospatial database is operated in QGIS by GIS professional and used basically for mapping purposes to visualise infection spread and dynamics by administrative districts in selected time period (Figure 6, 7).

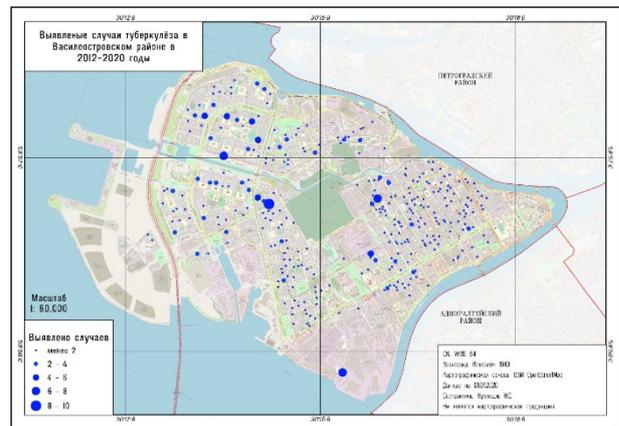


Figure 6. Map of detected tuberculosis infection cases, Vasileostrovsky administrative district of St. Petersburg, 2012-2020.

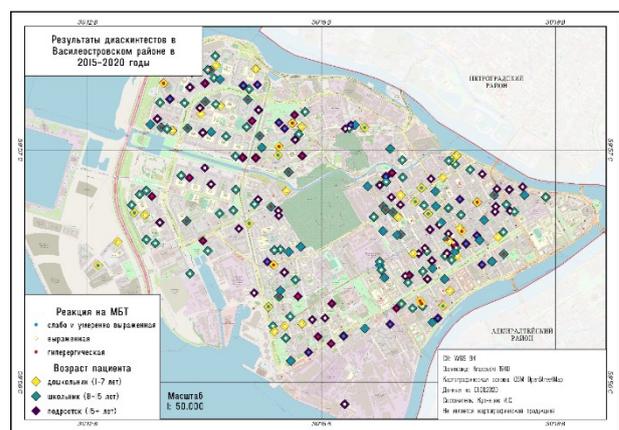


Figure 7. Map of testing results for tuberculosis infection, Vasileostrovsky administrative district of St. Petersburg, 2015-2020.

To ensure Web compatibility of the formed geospatial database we attracted NextGIS cloud platform (<https://nextgis.com>) that enables to visualise maps and separate map layers and publish visualised maps on the Web. In our study, NextGIS has established itself as an excellent QGIS compatible tool for Web GIS (Web component of elaborated GIS) development and implementation. NextGIS is a Russian commercial company that builds its business in the geospatial field upon the open source software, data and methods. The company appears also as one of QGIS contributors, that is why the NextGIS software is QGIS compatible and can be easily integrated in the frames of our research.

Previously, we implemented a pilot Web mapping project using the data collected for the Moskovsky administrative district of St. Petersburg (Kuznetsov et al., 2020). It was entitled as FEBRIS GIS. The Web map incorporated not only data about tuberculosis infection cased in the Moskovsky district area, but also information about subdivision of the district, and on tuberculosis clinics (Figure 8).

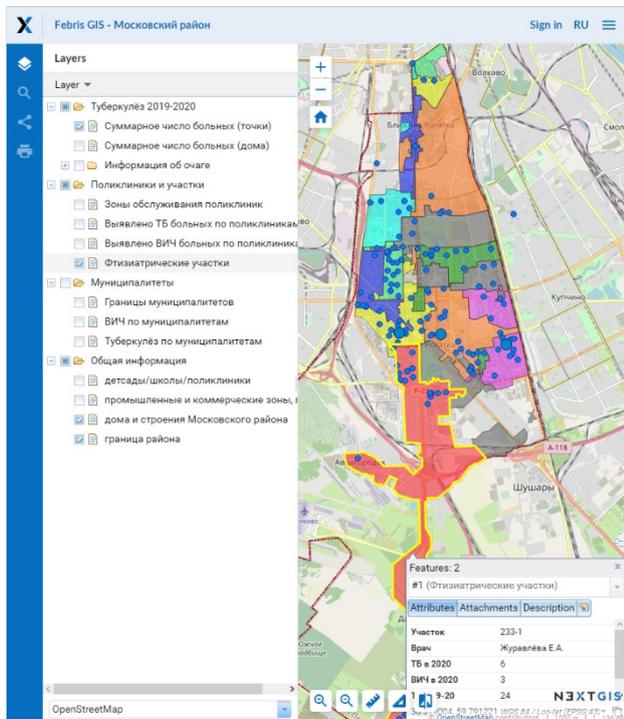


Figure 8. NextGIS/FEBRIS GIS Web interface.

Due to the issue of analysis methodology development and redesign (mentioned in the section 3 of the paper), we selected a dashboard approach for the Web representation of operated data at next stage of our study. To ensure Web representation of related information in addition to map layers itself we used Tilda (<https://tilda.cc>) Web platform and elaborated a prototype of FEBRIS GIS Web site. The Tilda is a commercial platform that provides a virtual Web hosting and WISIWIG Web site constructor. It was selected among similar platforms in result of comparison of the available functionality and service costs.

The Web site was determined to be having 5 sections (pages) to represent all the necessary information about the project. Since the site editor is block-based, we used blocks for each section. Significant feature of the Tilda is a single-page layout of developed Web sites (as Tilda was developed for small companies and landing Web pages used to collect contact information, but not for full-fledged Web sites). Nevertheless, Tilda provides zero blocks use (empty blocks allowing to design a page from scratch, and to embed necessary external components into the page), has convenient technical documentation, and equipped with easily customizable Search Engine Optimization tools. These were the pro arguments to select it as a Website engine.

With respect to new Web interface implementation, data representation chain in developed GIS has become composed of desktop preprocessing and loading into database, uploading to the Web GIS and Web maps visualisation, visualisation and representation of additional data on the Web site (<http://febrisis.ru>). Accordingly to the dashboard paradigm, the additional data incorporates lists, tables, diagrams and detailed map legends. However, being at the prototyping stage currently we implemented static representation of all the additional data to elaborate approach to data composition design (Figure 9). Map visualisation is integrated into the Web page as an external resource (derived from NextGIS/FEBRIS GIS).

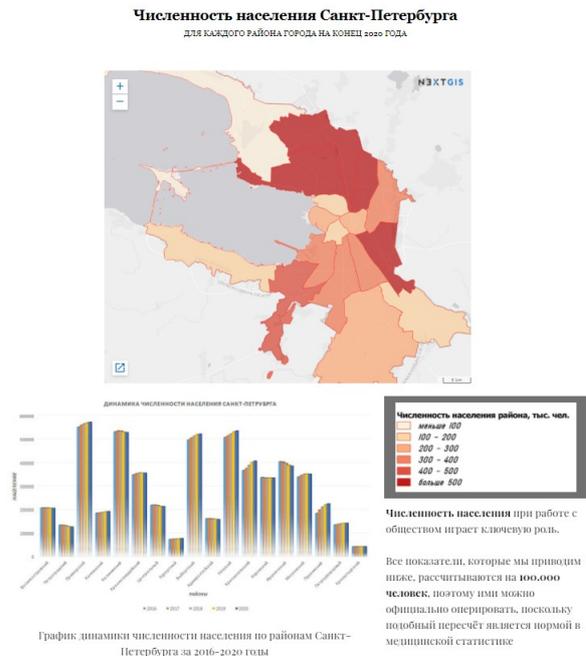


Figure 9. Example of data representation on the febrisis.ru.

It was established that first four of above mentioned indicators will be presented on the Web site currently due to the legislation issue (population amount, tuberculosis infection rate, tuberculosis infection spread rate, tuberculosis infection death rate). These indicators also can be perceived easily with minimal explanations by not qualified user. Visualisation of remaining indicators is available currently on the NextGIS/FEBRIS GIS Web interface. So the FEBRIS GIS application domain list is composed mostly of medical professional activities, and medicals are recognised as able to discover all the available data to resolve professional aims (Figure 10). Ordinary user can discover basic map visualisations on prototyped Web interface.

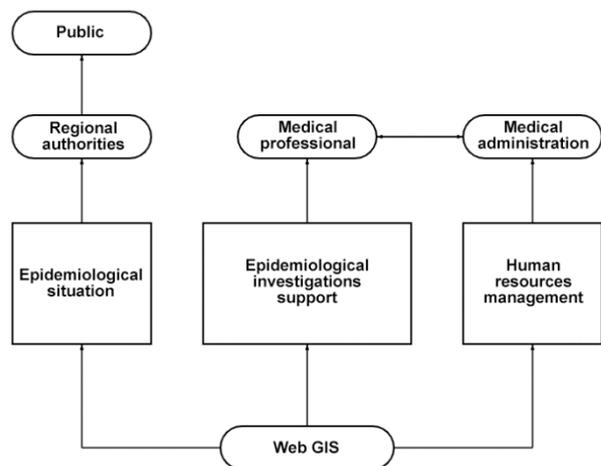


Figure 10. Application domains for the FEBRIS GIS.

5. CONCLUSIONS

As a current result of the work, the Web site was developed that is incorporated as a part of developed Geographic Information System. With the help of this part, it became possible to obtain spatial information about the epidemic situation in St. Petersburg in easy perceptible interface. The main requirements for the data visualisation were met successfully:

1. The Web interface allows to assess the overall epidemic situation in the city using thematic maps (map layers) and additional data
2. The Web interface provides the user with the necessary information about the essence of the conducted (our) research project

Next project stages were elaborated:

1. The initial data on the epidemic situation in the city were processed, all the data is provided in an anonymized aggregated form
2. Map layers representing medical statistics data have been created
3. NextGIS-based Web GIS was developed, filled with data, and configured
4. A Web interface has been prototyped that allows to visualise elaborated maps on the Web alongside the descriptive information and other accompanying resources
5. The project Web site was published on the Web under the permanent URL – <http://febrisis.ru>

Development of similar Web interfaces generally contributes to the expansion of knowledge about socially valuable diseases, allows to support fight against the diseases. Generally, the system (Web interface) is expected to be integrated into the Web site of the relevant medical organization. The software selected as a basic in our study makes it possible. Developed system appears as an excellent basis for interdepartmental cooperation, and allows building a high-quality dialogue between society and the government.

Basing on conducted work and on analysis of gained results we are able to formalize conclusions on currently needed additional elaborations in discovered domain. These conclusions form our research plan for nearest future, and can be used also by other domain-involved researchers to detect research aims.

In the plane of geocoding facilities, we can pose that it is needed to avoid extra conversions of the initial medical data. The key task is the elimination of intermediaries when transferring information from the medical databases to the map. It can be realised through the integration of the GIS subsystem into medical data management system, and through the application of the GIS subsystem to the postal addresses control on initial stages of medical statistics data formation.

In the plane of methodology development, we can conclude that it is needed to explore new metrics for the spatial distribution of infectious diseases to represent this phenomenon as multidimensional, while location only data becomes not enough to estimate spatial pattern of infection. For example, implementation of such parameter as apartment type in geospatial database probably will ensure wider capabilities for spatial patterns allocation.

Finally, in the legislation plane we can pose the need to propose mapping products produced in result our studies as a material for national-level discussion on medical maps standardization and(or) regulation in case of need.

ACKNOWLEDGEMENTS

Web GIS platform: courtesy of NextGIS LLC (<https://nextgis.com/privacy/>).

REFERENCES

- Chistobayev, A.I., Semenova, Z.A., 2013. Medico-geographical mapping in the former USSR and modern Russia. *Vestnik SanktPeterburgskogo Universiteta, Seriya Geologiya i Geografiya*, 4, 109-112. (in Russian)
- Dong, E., Dua, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5), 533-534. doi:10.1016/S1473-3099(20)30120-1
- Federal law, 2006. On personal data #152 of July 27, 2006. (in Russian)
- Federal law, 2020. On biological safety in Russian Federation #492 of December 30, 2020. (in Russian)
- Franch-Pardo, I., Napoletano, B.M., Rosete-Verges, F., Billa, L., 2020. Spatial analysis and GIS in the study of COVID-19. A review. *Science of the Total Environment*, 739, 140033. doi:10.1016/j.scitotenv.2020.140033
- 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.
- Government of the Russian Federation, 2004. The Order #715 of December 1, 2004, On approval of the list of socially valuable diseases and the list of diseases that pose a danger to people. (in Russian)
- Jeefoo P., Tripathi K.N., 2011. Dengue risk zone index (DRZI) for mapping dengue risk areas. *International Journal of Geoinformatics*, 7(1), 53-62.
- Kuznetsov, I., Panidi, E., Kolesnikov, A., Kikin, P., Korovka, V., Galkin, V., 2020. GIS-based infectious disease data management on a city scale, case study of St. Petersburg, Russia. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2020, 1463-1467. doi:10.5194/isprs-archives-XLIII-B3-2020-1463-2020
- Kuznetsov, I., Panidi, E., Korovka, V., Galkin, V., 2020. City-scale GIS-based monitoring of infectious disease – contemporary issues, case study of St. Petersburg, Russia. Proceedings, 6th International Scientific Conference GEOBALCANICA 2020, 791-796. doi:10.18509/GBP.2020.87
- Kuznetsov, I., Panidi, E., Korovka, V., Galkin, V., Voronov, D., 2020. Web-based representation and management of infectious disease data on a city scale, case study of St. Petersburg, Russia. *International Archives of the Photogrammetry, Remote Sensing*

and *Spatial Information Sciences*, XLIV-3/W1-2020, 87-91.
doi:10.5194/isprs-archives-XLIV-3-W1-2020-87-2020

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)

Martorell-Marugán, J., Villatoro-García, J.A., García-Moreno, A., López-Domínguez, R., Requena, F., Merelo, J.J., Lacasaña, M., de Dios Luna, J., Díaz-Mochón, J.J., Lorente, J.A., Carmona-Sáez, P., 2021. DatAC: A visual analytics platform to explore climate and air quality indicators associated with the COVID-19 pandemic in Spain. *Science of the Total Environment*, 750, 141424. doi:10.1016/j.scitotenv.2020.141424

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

President of the Russian Federation, 2019. Decree #97 of March 11, 2019, On the fundamentals of the state policy of the Russian Federation in the domain of chemical and biological safety for the period up to 2025 and beyond. (In Russian)

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

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

Wei, X., Wang, M., Kraak, M.-J., 2020. Where we are in fighting against COVID-19. *Environment and Planning A*, 52(8), 1483-1486. doi.org/10.1177/0308518X20931515