

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

I. Kuznetsov¹, E. Panidi^{1,*}, V. Korovka², V. Galkin², D. Voronov³

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

³ Saint Petersburg Central Tuberculosis Dispensary, St. Petersburg, Russia

Commission IV

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

ABSTRACT:

In 2019-2020, we conducted a set of case studies devoted to the investigation and design of a methodology for GIS-based support of medical administration and planning on a city scale when accounting and controlling infectious disease. The studies were conducted for the administrative territory of St. Petersburg city (Russia), and were based upon the medical statistics data collected and accounted by St. Petersburg medical administration. The statistics included data on tuberculosis, human immunodeficiency virus and hepatitis infection. All the medical data used in the study are impersonalized. GIS-based MDMS prototype was developed upon the QGIS software.

Moving forward in the previously formed study direction, now we are working on MDMS interface redesigning to facilitate its usability. Current activities are focussed on incorporation of the Web interface into previously developed MDMS prototype. The paper discusses development of the Web GIS interface prototype, and poses feature research and development aims. First feedback collected from medicals makes it possible to pose a Web-GIS-based MDMS as more flexible and easy to use, in comparison to the desktop-GIS-based.

1. INTRODUCTION

Geographical Information Systems (GISs) are demanded as one of valuable technologies by medical geography. Medical cartography formed as an overlap of medical geography and cartography domains (Chistobayev, Semenova, 2013) attracts GISs as a mapping automation toolkit, while medical geography itself attracts it as a geospatial analytics and data management infrastructure. Coronavirus pandemic appeared in 2019-2020 showed clearly that geospatial analysis is extremely valuable when analysing epidemics development and infectious diseases spread (Franch-Pardo et al., 2020). While people migrations appear extremely intensive (when studying on all scales, beginning from global and ending by city scale), infection spreads with according intensity (more than 42.8 million people have been infected worldwide and more than 1.1 million of them have died as of end of the October 2020, accordingly to the Johns Hopkins University – <https://coronavirus.jhu.edu>). The automation is critically needed also to account medical statistics data and produce geospatial estimations in relevant time. Despite this, GIS technology is still not everyday instrument for medical administrations in many regions of the world, due to a number of aspects, including its costs and its relative complexity from the point of view of medicals.

For more than a year, our research group formed by researchers from universities and medical institutions is working on the study of GISs application to accounting and modelling (first of all, mapping) of the spread of socially valuable diseases (Russian Government, 2004) in the administrative area of St. Petersburg city (Russia). During this time, we have gained

essential experience in the integration of real-life medical statistics data into GIS (Kuznetsov et al., 2020). Currently collected data on tuberculosis for the last 5 years cover all 18 administrative districts of St. Petersburg. In 6 out of 18 districts, a comprehensive study and mapping were conducted, with attraction of the additional medical statistics data on human immunodeficiency virus (HIV) and hepatitis, and estimations of the work quality of local hospitals and clinics. Data on HIV and hepatitis were needed to estimate the distribution of so-called “mixed infection”. The mixed infection is understood by medicals as joint spread of abovementioned infections.

Results gained on previous stages of our investigations incorporated into the prototype of GIS-based medical data management system (MDMS), particularly we developed a set of techniques for medical statistics data accounting, storage and mapping. The MDMS was built upon the OGIS software as this universal desktop GIS reduces costs being full stack open source GIS software. We designed data storage model applicable for the medical statistics data we were dealing with, and developed QGIS module used as data management interface. Additionally, algorithms and corresponding program code were elaborated to ensure geocoding of the medical statistics data, and a set of maps were produced to reflect spatial distribution of collected data and to test appropriate mapping techniques.

Interaction with medicals shows clearly that implementation of GISs helps them to highlight and visualise diseases spread in the city. We have not gained official incorporation of our ideas into system of infectious diseases accounting and monitoring in St. Petersburg currently. However, communication with

* Corresponding author

colleagues show us that the study prompted local authorities and a number of specialized institutions (located not only in St. Petersburg) to start considering GISs as a fundamental tool helping to fight against socially valuable diseases.

We can mention the City Tuberculosis Dispensary No. 3 (Primorsky district of St. Petersburg) as an example. Specialists of the dispensary detected an outbreak of tuberculosis infection in early July 2020 in one of the new built-up areas of the district, basing also on the data accumulated and visualized in the frames of our research activities. Throughout the summer, constant monitoring of the territory was carried out. Mobile teams of medicals monitored and examined population of the area directly at homes. According to the information provided by medicals, there were about 10-12 new (unregistered previously) cases of tuberculosis infection. This case demonstrates the possibility of GIS incorporation as an supporting technology when organizing a preventive fight against the disease in the area.

However, despite all the advantages, GISs still cannot become a reliable basis for diseases detection. GIS tools are operated easily by GIS domain specialists, but are poorly suited for operation by medical professionals. This context raised the question of further development of the project. One of the promising directions for us is the incorporation of Web GIS facilities into developed toolkit.

2. OPERATED MEDICAL STATISTICS DATA

Currently, we have accumulated a significant number of different map layers and embedded these layers into developed GIS/MDMS prototype. We store GIS data in the SHP format, to preserve restructuring simplicity of accumulated data massive at the ongoing experimenting and prototyping stage of the study.

The data is depersonalized accordingly to the Russian laws (Federal law on personal data). Here we have to mention that postal addresses of patients (registered infection cases) are preserved in the data we operate, and this not contradicts to the law but may seem strange. In any case, taking into account that the St. Petersburg city is built-up by multi-floor buildings (usually 5-floor and more) we may conclude that deanonymizing of patients is almost completely excluded.

Postal addresses are needed to geocode and map initial medical statistics data. About 14,000 addresses are processed already. This gives a possibility to visualize development of tuberculosis both in a particular administrative districts and in the entire city (Fig. 1 and 2).

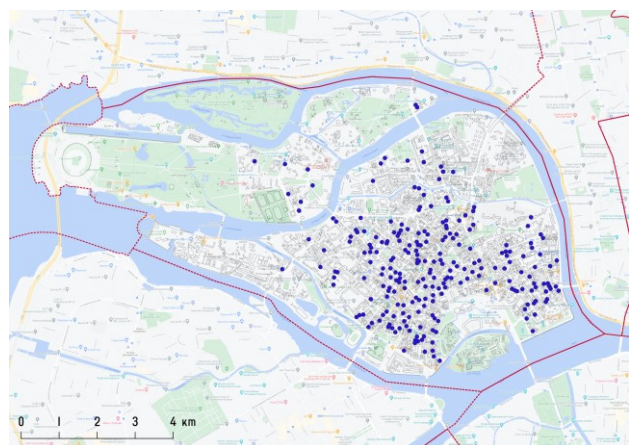


Figure 1. Tuberculosis infection cases observed in the Petrogradsky administrative district of St. Petersburg

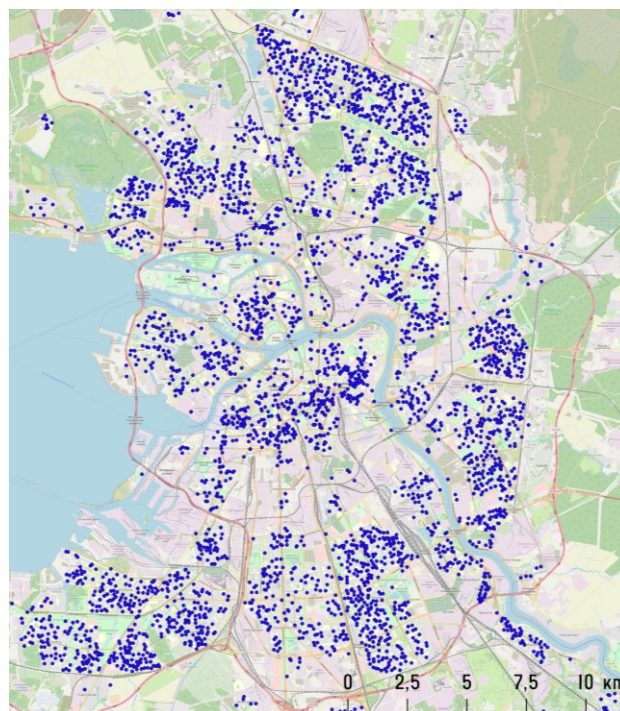


Figure 2. Tuberculosis infection cases observed in the area of St. Petersburg city

One of valuable problems of medical statistics data accounting in Russia is its partial decentralizing. Large cities (this is our case also) are divided into a number of operating zones with their own separated socially valuable diseases monitoring services. Statistics data collected in the operating zones are reported to the city's Central Tuberculosis Dispensary (<http://tubercules.org/>), but the content of these reports is declared as intended for informational purposes only. In turn, the Central Tuberculosis Dispensary maintains its own statistics. Finally, this leads to some inconsistencies in statistics data collected and stored at different administrative levels. Figure 2 represents reworked data of Central Tuberculosis Dispensary.

Abovementioned decentralizing of medical statistics data leads also to the inconsistencies in structuring of data presented by zonal divisions. In our study, we tend to collect all available data to observe infectious situation in whole and to detect internal contradictions of medical statistics. In this way we designed unified template for the initial data structuring that includes currently:

1. Infection type
2. Dates of its registration and deregistration (due to the healing or death)
3. Structured postal address of infection observation (Kuznetsov et al., 2020)

However practically, the template can be used currently only as a data unification tool at the stage of incorporation of the initial medical statistics data into GIS. Every zonal division has its own data management approach and cannot implement currently neither redesign the approach, nor fill an additional reporting form (first of all, due to the overloading of involved medicals and, as a consequence, their physical inability to fill out any additional documents).

3. REPRESENTATION OF MEDICAL STATISTICS DATA

Currently we represent accumulated data in desktop (QGIS-based) interface. Incorporation of Web GIS tools into developed GIS/MDMS prototype may help to transform it into distributed system and to transfer initial statistics accounting tools directly to the zonal divisions of diseases monitoring system through the Web browser interface. Potentially, it may help zonal divisions to optimise accounting process and to unify the form of accounting results representation. In this case, we may automate also the formation of GIS database and move from centralised database accumulation to collaborative, ensured directly by zonal divisions.

Another one significant problem that can be resolved with the help of Web GIS interface is the data over-accumulation and its rapid obsolescence. By request of local authorities and medical executive units, individual datasets can be extracted from the GIS and represented separately for individual territories. However, the data is constantly updated and the corrections are made on a weekly basis. So the special-case datasets extracted manually become outdated even before any decisions are made in a particular end-user organization. The Web interface should help solve this problem also. Users have to be able in this case to view the most updated information in self-service mode without going into the details and complexity of GIS operating, as the Web interface is maximally intuitive for almost any user.

Web interface have become popular in recent years for publication of geospatial data and for it's delivery to consuming end users. Web maps and Web GIS solutions can be discovered currently as well-known GIS tools, but remains developing GIS domain, both in the plain of attracted Web technologies, and in the plain of application ideology and use cases.

One of special classes of the Web GIS solutions composed of Web interfaces called usually as dashboards. Such solutions are built upon a toolset that incorporates Web mapping, Web infographics and backend computation and analytics tools. Geospatial data are visualized in this case in multimedia mode (in parallel in the forms of maps, diagrams, lists, tables and texts), that reflects complexity of geospatial data (in general case) and facilitates visual analysis of the data.

Coronavirus epidemic of 2019-2020 have pushed application of dashboard solutions into research and development trends when visualizing medical geospatial data (Dong et al., 2020; Wei et al., 2020; Martorell-Marugán et al., 2021). Well-known examples in this case are the Johns Hopkins University COVID-19 dashboard (<https://coronavirus.jhu.edu/map.html> – Fig. 3) and the dashboard of the Regional Office for Europe of the World Health Organisation (<https://who.maps.arcgis.com/apps/opsdashboard/index.html#/a19d5d1f86ee4d99b013eed5f637232d>), both built upon ArcGIS Online platform (<https://www.arcgis.com>). Similar solution is developed by Yandex (Russian IT-giant). This dashboard is built upon its own online analytics and Web map engines (<https://datalens.yandex/7o7is1q6ikh23> – Fig. 4) and uses Johns Hopkins University medical statistics data for entire world and official data of Russian authorities (<https://xn--80aefpbgmfb1c0a.xn--p1ai>) for the territory of Russia.

Dashboard approach perfectly solves several tasks at once: it informs citizens about possible risks, provides data for the media, and allows the authorities to provide somehow control of disease development. The approach can be used as a best practice when designing and developing GIS/MDMS interface.



Figure 3. Web interface of the COVID-19 dashboard from the Johns Hopkins University

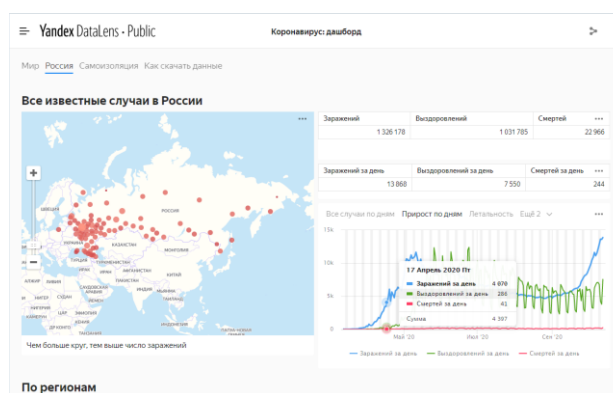


Figure 4. Web interface of the COVID-19 dashboard from the Yandex

4. PROTOTYPING OF THE WEB ACCESS TO THE MEDICAL STATISTICS DATA

After reviewing and analysing solutions available on the market, we came to the conclusion, that there is no ready-to-use Web interface engine optimal to be incorporated into our GIS/MDMS prototype, and we have to develop our own dashboard using different open source components as building blocks. We made this decision based on the following factors:

1. Economic feasibility. The desktop GIS prototype is built upon open source software, so the integration with commercial Web platforms will rise the costs significantly
2. A large array of processed information. Not critical, but in a number of cases server-side data storage at commercial platforms is limited or demands extra payment
3. Low level of GIS competence of the medicals. It is needed to ensure design and architecture of the system maximally flexible and applicable to be reworked or replaced due to the specific needs of Web interface end users

Maximal available flexibility in Web GIS interface development can be ensured when using PHP and JavaScript Web development, and building the interface from scratch. Such a way allows to elaborate any architecture of system backend and to optimize data processing operations. This way however assumes time spending for basic Web GIS functionality development (data loading, database operating, etc.). Therefore, despite the obvious advantages of pure PHP/JavaScript use, we decided to use Python Django

(<https://www.djangoproject.com/start/overview/>) as a Web GIS interface engine. The Python scripting corresponds also with use of QGIS/Python for the development of the desktop GIS/MDMS subsystem.

First “must have” building block is the map visualisation (or simply Web mapping) engine. A number of Web mapping engines are presented on the market, but for prototyping we selected two distributed by Russian geospatial IT companies, NextGIS and Geosemantika. The selection was based upon the possibility of easily available support in native language, and on the legal aspect (as we operate some kind of specific data, it is better to leave it in the legal framework of Russia).

NextGIS (<https://nextgis.com>) is a Russian commercial company that builds its business in the geospatial domain around open source software, data and techniques and appears one of QGIS contributors. NextGIS software stack incorporates QGIS compatible tools for Web representation of geospatial data and Web GISs development. A pilot Web mapping project called “Febris GIS” (<http://febrisis.nextgis.com/resource/14/display?panel=layers>) use the data we collected for the Moscovsky administrative district of St. Petersburg. Web map incorporates not only data on patients in the area, but also information about clinics and operating zones (Fig. 5).

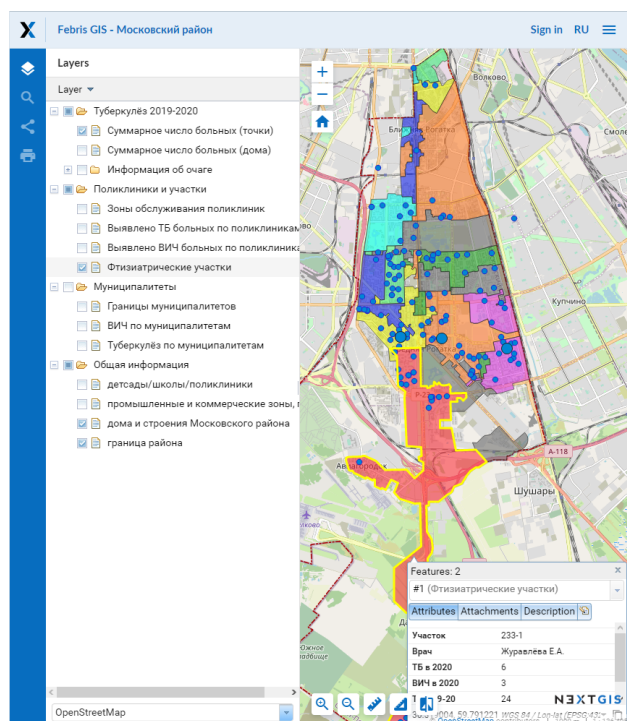


Figure 5. Web interface of the Febris GIS project

Geosemantika (<https://geosemantika.ru/>) is a Web platform for spatial data publication, which includes sharing access rights, cataloguing and maps publication itself. Standard Web interface of the mapping platform is much more diverse. For example, a developer can add a timeline to the map to operate time dependant data that is extremely important for our study. Commercial account at the Geosemantika platform costs almost 2 times cheaper to operate in comparison to the NextGIS.

5. CONCLUSION

Web publication of generalised medical statistics data and derivative maps was supposed at the early beginning stage of our study. However, the elaboration of medical statistics data preprocessing and structuring was needed to be done before. We see huge potential in this way. Web representation of the medical statistics data can be not only information source for citizens, but also have to facilitate the work of medical professionals. Even now, at the stage of desktop GIS/MDMS prototyping, we receive positive feedback from both the medical managers and ordinary specialists. Implementation of the Web component into the GIS/MDMS will ease interaction with the data and maps for the medicals.

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.org/10.1016/S1473-3099(20)30120-1
- Federal law on personal data #152 of 27.07.2006 (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.org/10.1016/j.scitotenv.2020.140033
- 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. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B3-2020, 1463-1467. doi.org/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 GEOALCANICA 2020*, 791-796. doi.org/10.18509/GBP.2020.87
- 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.org/10.1016/j.scitotenv.2020.141424
- Russian Government, 2004: The Order of the Russian Federation Government of December 1, 2004 No. 715, On approval of the list of socially valuable diseases and the list of diseases that pose a danger to people (in Russian)

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