

# Electronic System of Storage, Representation, and Analysis of Epidemiological Information on a Federal District Scale

DOI: 10.17691/stm2017.9.4.21

Received May 3, 2017



L.A. Soltsev, PhD, Chief of the Laboratory of GIS-Technologies and Bioinformatics<sup>1</sup>;  
 N.N. Zaitseva, MD, PhD, Head of the Privolzhsky District Center on Prophylaxis and Struggle against AIDS<sup>1</sup>;  
 E.I. Efimov, MD, DSc, Professor, Director<sup>1</sup>; Head of the Department of Microbiology and Immunology<sup>2</sup>

<sup>1</sup>Blokhina Scientific Research Institute of Epidemiology and Microbiology of Nizhny Novgorod, Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor), 71 Malaya Yamskaya St., Nizhny Novgorod, 603950, Russian Federation;

<sup>2</sup>Nizhny Novgorod State Medical Academy, 10/1 Minin and Pozharsky Square, Nizhny Novgorod, 603005, Russian Federation

**The aim of the study** was to work out an epidemiological atlas, the functional core of which is the database of the epidemiological process of actual infectious diseases in Privolzhsky Federal District (PFD).

**Materials and Methods.** The epidemiological atlas consists of the following components: a data bank, topographic base, direct data access tool, and visualizer of the epidemiological situation. The topographic base and data bank form a database, while the visualizer and direct access tools form a graphical shell for working with the database. The project was implemented as a website. Reports of PFD subjects filled-in in compliance with form No.2 "Information about infectious and parasitic diseases" served as the material for creating the database. The database is controlled by MySQL system, which provides the possibility to store the information in the form of linked tables. Using the PHP language the information is converted into pivot tables, diagrams of time dynamics, and schematic maps.

**Results.** To organize the process of acquisition, storage and maintaining the epidemiological atlas database in the running condition, a protocol of data exchange was designed and implemented into practice, the necessary software was created for semi-automatic and automatic report processing and inserting into the database. XLS (XLSX) format of the electronic tables was chosen as an exchange format.

The following functions are realized in the epidemiological atlas: automatic generation of reports from the database; the analysis of intra-annual incidence dynamics; evaluation of the current monthly dynamics of the morbidity rate compared to the retrospective information.

All generated materials can be exported in the form of XLS (CSV) tables or diagrams as PNG, JPEG, PDF files.

At present, the atlas represents a constantly functioning and regularly updated system accumulating data on infectious morbidity rate over all 14 PFD subjects with a spatial resolution to the administrative region subject and a time resolution of 1 month.

**Conclusion.** At present, there are no analogs to the "PFD Epidemiological Atlas" in the Russian Federation. Our project is a complete geoinformation system, where the possibilities of the initial data access, analytical and prognostic information acquisition, and export of the results are combined.

**Key words:** geoinformation systems; epidemiological atlas; epidemiological information database; epidemiological process monitoring.

The most important criterion of the population health is a morbidity rate (incidence). Incidence indices are registered on the regular basis with a high degree of detailing in all medical organizations of the Russian Federation, which report to the organs and institutions of Rospotrebnadzor (the Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing) (Rosstat order No.52 of January 28, 2014; Rospotrebnadzor order No.1203 of December 2, 2014).

The collected information meets the spatial data requirements [1]:

has a spatial tie-in (a settlement, administrative area, region, federal district, where the morbid event (case) was registered), and a time localization (the day, month, year of the registered event);

has an attributive part, which contains information about cases of registering this or that infection in the given object and/or any other information (population size, water treatment quality, localization in the natural and geographical zone, etc.);

as a "substrate", has a map of the examined (monitored) territory, the separate elements of which

**For contacts:** Leonid A. Soltsev, e-mail: solntsev.l.a@nniem.ru

(settlements, administrative areas, regions, federal districts) are the objects of surveillance.

Consequently, there are no theoretical and methodological obstacles to using geoinformation systems (GIS) and a spatial data model as the tools for epidemiological analysis. The development of such systems gives the opportunity to perform semiautomatic and automatic analysis of the captured information by means of the methods and algorithms used in the practical work.

**The aim of the study** was to work out an epidemiological atlas, the functional core of which is the database of the epidemiological process of actual infectious diseases in Privolzhsky Federal District (PFD).

To achieve the goal, it was necessary to solve the following tasks:

1. To develop a protocol of obtaining initial data from the Hygienic and Epidemiological Centers (HEC) and/or Rospotrebnadzor territorial offices in the PFD subjects.

2. To design and realize the database structure to store the information.

3. To create a graphic shell to work with the base.

#### Materials and Methods

**Initial data.** Presently, the territorial organs and institutions of PFD Rospotrebnadzor apply various software for morbidity registration: “Analysis of population morbidity” (APM) program; the program AS “Statistica Rospotrebnadzor” (KRISTA); original program solutions created by the organization specialists. Data in these software products are stored in their own formats, therefore they can be accessed using the software, in which they were developed. However, the research work implies addressing the initial data, and application of automated algorithms of data processing imposing definite requirements on the format of the data themselves. Such a format will be further called “machine-readable” (i.e. information is stored in the form of uniformly structured tables, and so on). Now, the principle of storage and presenting data in the

machine-readable form is rarely realized in practice in the Rospotrebnadzor system of morbidity registration to the full scope. It should be noted that information can be exported into these programs only in standardized forms. This fact, considering the character of the software and the tasks it faces, is not a drawback, however, it demands creation of some intermediate means for conversion of the data from standardized reports to the machine-readable form.

Reports filled-in in compliance with form No.2 “Information about infectious and parasitic diseases”, being one of the tasks of HEC, served as an initial material for creating the epidemiological atlas database. The reports were converted into the format suitable for the direct export into the base.

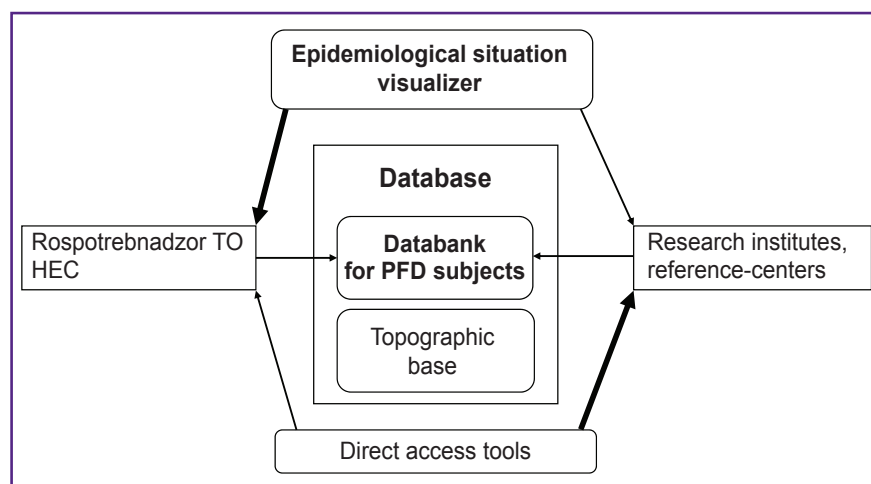
**System architecture.** The epidemiological atlas is composed of the following components: a data bank, topographic base, tool of the direct access to the visualizer information about epidemiological situation. The topographic base and data bank form a database. The visualizer and the direct access tool make a graphic shell for working with the base. A general atlas architecture is presented on Figure 1.

The data bank is filled with the information presented by HEC and Rospotrebnadzor territory offices as well as by scientific organizations and reference-centers (for specialized sections of the epidemiological atlas). The functional of the epidemiological situation visualization is oriented to the Rospotrebnadzor surveillance bodies. The direct access tools allow obtaining the information in the machine-readable form.

The functional core of the epidemiological atlas is the database controlled by MySQL system, which provides the possibility to store the information in the form of linked tables. Using the PHP language the information is converted into pivot tables, diagrams of time dynamics, and schematic maps. The user is not limited by the types of analysis and data presentation created by designers, he has always the opportunity to get access to the initial information.

**Figure 1. General atlas architecture**

Paths of information movement are designated by arrows: thick arrows show the paths that are most eagerly sought by Rospotrebnadzor territorial organs (TO) and research institutes respectively. HEC — Hygienic and Epidemiological Center; PFD — Privolzhsky Federal District



PFD_DENAIL		
Column	Type	Comments
SHAPE	geometry	Field for object geometry storage
adm3_name	varchar	PFD subject name
adm4_name	varchar	Administrative area name
adm_lvl	varchar	Numeric code of administrative subordination level
loc_kod	int	<b>General unique territory identifier</b>
okato_name	varchar	Territory name according to RCOAD
rcap	varchar	PFD subject capital name
rkod	int	<b>PFD subject code</b>
rname	varchar	PFD subject name

AREA_T		
Column	Type	Comments
okato_name	varchar	Territory name according to RCOAD
adm3_name	varchar	PFD subject name
adm4_name	varchar	Administrative area name
adm_lvl	varchar	Numeric code of administrative subordination level
rkod	int	<b>PFD subject code</b>
rcap	varchar	PFD subject capital name
global_kod	int	<b>General unique territory identifier</b>
loc_name	varchar	Territory name according to the reports of PFD subject
loc_cod	int	Local numeric identifier

ILL_T		
Column	Type	Comments
GR_ID	int	Identifier of a disease group
GR_ID_DESC	varchar	Name of a disease group
MKB10	varchar	ICD10 code
ID21	int	<b>Internal disease code (from reports)</b>
ID16	int	External disease code
NAME21	varchar	Disease name (from reports)
NAMEX	varchar	Disease name

R21INF		
Column	Type	Comments
yr	decimal	Year
mon	decimal	Month
area	int	<b>General unique territory identifier</b>
dies	decimal	<b>Unique disease identifier</b>
s1a	decimal	Number of diseased ("Total population" group)
s21a	decimal	Number of diseased ("under 17 years" group)
s3a	decimal	Number of diseased ("under 14 years" group)
s4a	decimal	Number of diseased ("under 1 year" group)
s5a	decimal	Number of diseased ("from 1 to 3 years" group)
s7a	decimal	Number of diseased ("from 3 to 6 years" group)
s10a	decimal	Number of diseased ("Rural population" group)
s23a	decimal	Number of diseased ("Rural population (under 17 years)" group)
s18a	decimal	Number of deaths ("Total population" group)

R21NAS		
Column	Type	Comments
yr	int	Year
area	int	<b>General unique territory identifier</b>
s1r	decimal	Population size ("Total population" group)
s21r	decimal	Population size ("under 17 years" group)
s3r	decimal	Population size ("under 14 years" group)
s4r	decimal	Population size ("under 1 year" group)
s5r	decimal	Population size ("from 1 to 3 years" group)
s7r	decimal	Population size ("from 3 to 6 years" group)
s10r	decimal	Population size ("Rural population" group)
s23r	decimal	Population size ("Rural population (under 17 years)" group)
s18r	decimal	Population size ("Total population" group)

Figure 2. Scheme of interconnections for data tables in the module of a separate Privolzhsky Federal District (PFD) subject (shown for Republic of Chuvashia as an example)

PFD\_DETAIL — spatial data along the borders of the subject administrative units; AREA\_T — PFD subject administrative units directory; ILL\_T — infections reference book; R21INF — data on incidence rate; R21NAS — data on population size

The epidemiological atlas is a complete system, i.e. in addition to the function of information storage and presentation it has an interface for inputting new data, which are automatically processed in the system and are considered when generalized materials are prepared.

Schematic diagram of data organization in the base with reference to a separate PFD subject is shown on Figure 2.

### Results

**Data input and storage.** A protocol of data exchange between the territorial organs and Rospotrebnadzor institutions and Blokhina Scientific Research Institute of Epidemiology and Microbiology of Nizhny Novgorod was developed, the necessary software tools for semiautomatic and automatic processing of reports and their input to the database were created to organize the acquisition, storage and maintenance of the epidemiological atlas base in the running condition. XLS (XLSX) format of electronic tables was chosen as an exchange format. This format is accessible as an export tool in all program products used by the PFD territorial administrations, and is convenient for the automatic data processing and conversion. As of April 2017, archival data sets (up to 2010) and monthly reports, which became later the main form of information exchange, were obtained and processed. Presently, the process of data obtaining and processing is already a routine procedure not requiring any additional correction. Besides, a double system of input correctness control is used: a formal compliance of the report structure with the initially set criteria (a lay-out, margin names, etc.) is verified, and a sampling checkback is performed, i.e. the values obtained after report conversion are compared with the initial ones. In addition, a tool was developed for a pointwise correction of already input data, as well as a tool for converting the obtained tables in SQL transactions.

**Thematic mapping.** The main way of data presentation is thematic mapping and automatic territory classification according to the attributive sign (morbidity rate, and so on). And here, it is important to choose a classification scale (for a more precise highlighting of the problem territories it is necessary to pass on from the level of PFD subjects to the level of the administrative units of these subjects) and the method of combining the territorial entities into groups. If one attribute is used, it is optimal to apply one of the simplest methods of classification. In the epidemiological atlas, a user can choose one of the most common classification methods: equal, quantile, geometric intervals, natural breaks. In those cases when classification should be made by a set of attributes (both qualitative and quantitative), more complex methods are used, e.g. discriminant analysis, computer-aided instruction, etc. A combination of several parameters on one schematic map is also possible. For example, the borders of the supposed focus can be marked (in compliance with the data presented by HEC),

accompanying information displayed, and the areas grouped by the level of incidence/prevalence of some natural focal disease.

**Data analysis.** Currently, the following capabilities are realized in the epidemiological atlas:

- automatic generation of reports from the database (for example, comparative analysis of the current and last year events, etc.) with modification of the obtained result in accordance with the approved reporting form;

- analysis of the intra-annual incidence dynamics. For example, data about the contribution of separate months to the total number of the diseased per year can be presented in the form of a "thermal map" or a line graph. One more variant is the analysis of the intra-annual morbidity rate [2] and highlighting separate functional components of the morbidity dynamics series [3, 4]. The last method is known to be rarely applied in the epidemiological practice in Russian Federation;

- assessment of the current monthly dynamics of the morbidity rate compared to the retrospective. It allows prediction of the morbidity rate development for the current year orienting to the spread of the given month values for the preceding years.

All generated materials can be exported in the form of tables in the XLS(CSV) format or diagrams in PNG, JPEG, PDF formats. The graphic material is sufficient for using in accounts and reports.

**Registration and implementation of the database into practice.** The database of the epidemiological atlas passed the procedure of the state registration (Certificate of State Registration of the database "Electronic epidemiological atlas of Privolzhsky Federal District" No.2015621143 of 24.07.2015), the request was satisfied by the department of registration of computer programs, databases and integrated microcircuit topologies of the Federal Institute of Industrial Property of the Federal Service for Intellectual Property (Moscow). The application of the epidemiological atlas in the practical work was approved by Rospotrebnadzor order No.283 of 13.04.2015 "On realization of the board solution of 06.03.2015 No.2 "On the activity of the FBES "Blokhina Scientific Research Institute of Epidemiology and Microbiology of Nizhny Novgorod" of Rospotrebnadzor for 2012–2014"; by Rospotrebnadzor order No.1129 of 23.10.2015 "On conduction of the regional meeting "The study of the positive experience of using electron epidemiological atlas of PFD in organization and conduction of epidemiological surveillance of actual infections"; by Rospotrebnadzor order No.1209 of 16.11.2015 "On creating a regional (district) center to monitor PFD infectious diseases on the basis of FBES "Blokhina Scientific Research Institute of Epidemiology and Microbiology of Nizhny Novgorod" of Rospotrebnadzor"; by the protocol of the meeting "On additional measures of realizing assignments of the Governmental commission on human health protection regarding prevention of HIV-infection spreading in the Russian Federation" (8–10 December 2015, Suzdal).

Implementation of the epidemiological atlas into practical work is documented by the acts of HEC Rospotrebnadzor administrations in the PFD subjects (Bashkortostan, Mordovia, Tatarstan, Republics of Mari El, Udmurtia and Chuvashia, Nizhny Novgorod, Penza, Saratov, Samara, Ulyanovsk, Kirov, Perm regions).

**Discussion.** There are systems in the Russian Federation, which are introduced into daily practice and are provisionally available to specialists.

ZikaMap cartographic program for acquisition and analysis of information on measures for monitoring and regulation of the mosquitoes population, the carriers of viral diseases including Zika fever, developed by Stavropol Anti-Plague Institute of Rospotrebnadzor [5], is used presently on the Krasnodar territory. This program is oriented mostly to the input of data of entomological monitoring and its visualization rather than the analysis. Besides, the system is dedicated only to a specific purpose.

Rostov-on-Don Research Anti-Plague Institute of Rospotrebnadzor [6] has developed a geoinformation portal of infectious diseases (cholera, anthrax, tick-borne hemorrhagic fever). It simply visualizes the registered cases of infections from the list.

The development of the Russian Research Anti-Plague Institute "Microbe" of Rospotrebnadzor [7] — portal "Sanitary measures in Gorno-Altaysk natural focus of plague" — is the most elaborated system. The portal map is provided with a legend, which contains information about the layers used. The following can be found on the electron map: name, age, status, date of the disease onset, the contact date, probable place of contamination, disease form, laboratory confirmation, method of laboratory confirmation, date of laboratory confirmation.

Additionally, the portal contains information about 189 medical settings in Republic Altai including their address, the number of medical staff (including specialists in extremely dangerous infections), availability of boxes.

Among the foreign systems, WHO [8] and CDC [9] portals can be noted. These portals are designed for presenting global information (to the country level in [8] or exclusively for USA administrative units [9]). Foreign systems are small-territory-oriented systems, their tools for data input, storage, and visualization are unknown to the authors. These projects are likely to be conducted by specialized organizations and represent closed industry solutions.

The GIS project "Epidemiological atlas of Privolzhsky Federal District" (<http://epid-atlas.nniem.ru>) created on the basis of Blokhina Scientific Research Institute of Epidemiology and Microbiology of Nizhny Novgorod of Rospotrebnadzor is a unique publicly available development of the similar class on the territory of the Russian Federation. At present, the atlas represents a constantly functioning and regularly updated system accumulating data on infectious morbidity over all 14 PFD

subjects with a spatial resolution to the administrative region subject and a time resolution of 1 month. The archive range of coverage is from 2010 to the present days. The functional of the epidemiological atlas enables realization of all stages of the epidemiological analysis: from obtaining separate data from the district subjects to the presentation of the consolidated analytical and prognostic information.

The epidemiological atlas may serve as the basis for creating the whole class of more dedicated systems (surveillance of natural focal diseases, monitoring of antiepidemic measures, etc.). At present, a specialized section on HIV-infection monitoring is functioning ([http://epid-atlas.nniem.ru/a03\\_data\\_main\\_aid.html](http://epid-atlas.nniem.ru/a03_data_main_aid.html)).

Implementation of the epidemiological atlas into practical work of research and surveillance establishments of Rospotrebnadzor makes it possible:

- to decrease the number of inquiries to the territorial organs;

- to reduce the time of data reception by the organizations concerned;

- to automate the process of preparing data for analysis;

- to use off-the-shelf-solutions in performing standard analyzing procedures.

The developed GIS project meets the requirements of the public policy of introducing new technologies into practical work of Rospotrebnadzor institutions and health authorities.

**Conclusion.** At present, there are no analogs to the "Epidemiological atlas of Privolzhsky Federal District" in the Russian Federation. Our project is a complete GIS system, combining the possibilities of the initial data access, analytical and prognostic information acquisition, and export of the results.

**Acknowledgement.** The authors thank administrations of Rospotrebnadzor PFD subjects and Hygienic and Epidemiological centers in PFD subjects for their assistance in preparing initial materials for the database.

**Study Funding and Conflicts of Interest.** The work was not supported by any source, and there are no conflicts of interest related to the study.

## References

1. *Geoinformatika* [Geoinformatics]. Pod red. Tikunova V.S. [Tikunov V.S. (editor)]. Moscow: Izdatel'skiy tsentr "Akademiya", 2010.
2. *Osnovy retrospektivnogo analiza infektsionnykh zabol'evaemosti* [Fundamentals of retrospective analysis of infectious diseases]. Pod red. Shkarina V.V., Rakhmanova R.S. [Shkarin V.V., Rakhmanov R.S. (editors)]. Nizhny Novgorod: Izdvo NizhGMA; 2009.
3. Cleveland R.B., Cleveland W.S., McRae J.E., Terpenning I. STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* 1990; 6(1): 3–73.
4. Filatova E.N., Solntsev L.A., Utkin O.V. The impact

of seasonal factors on the dynamics of incidence rate of infectious mononucleosis in different age groups (on the example of Nizhny Novgorod). *Epidemiologiya i infeksionnye bolezni* 2017; 22(2): 79–85.

5. *ZikaMap*. URL: <http://snipchi.ru/page.php?326>.

6. *Geoinformation portal*. URL: <http://gis.antiplague.ru/index.php>.

7. *Sanitary measures in Gorno-Altai natural focus of plague*. URL: <http://www.microbe.ru:6655/altaymap/>.

8. *WHO*. URL: <http://www.euro.who.int/en/health-topics/disease-prevention/vaccines-and-immunization/publications/surveillance-and-data/who-epidata>.

9. *CDC*. URL: <https://www.cdc.gov/surveillancepractice/data.html>.