

Alfred Nizard (INED)

Vladimir Pshenkin (Moscow State University, Faculty of Economics)

Alain Monnier (INED)

Alexandre Avdeev (Université de Strasbourg 2 « Marc Bloch »/Moscow State University)

Using the Internet technologies to explore huge databases with changing structure: the example of the World Health Organization International Mortality Database (WHO IMD).

1. Nowadays it comes into particular prominence that, in comparison with many other scientific areas, the population studies are backward with respect of using Internet resources and modern information technologies. Actually there are only a few public accessible Internet resources providing users with reliable demographic information. This is especially surprising because for the long time the population data is one of the most standardized and structured all over the world. It could sound paradoxically, but the foregoing advantages of demographic data were probably the main cause of a gap in applying of modern information technology to their exploration. The explanation is simple, though. Most of national, as well as international databases were developed at the time when the Internet technologies were still in “embryo” status. Their creation was very costly in terms of material and human resources, their functioning satisfies users’ requirement so far, so the idea of adaptation of the existing databases to the new technologies does not meet with a very enthusiastic support among their creators or owners. As a consequence, the practical use of the information contained in the existing databases is quite limited.

2. The WHO IMD could be given as one of the examples of extremely rich source of information whose exploitation is not easy for users. This well-known database contains information about numbers of deaths by age, sex and cause in 241 countries and covers a period since 1950 up to nowadays. Today the database consists of 2 158 538 records or about 500 Megabits of plain text. But such an impressive size is not the only difficulty of database exploitation; more serious problem is that 7-dimensional information: country, calendar year, International Classification of Diseases and Causes of Death (ICD) revision, cause of death, sex, age and age type, is represented in a usual 2D table. Thus, one column may contain information about several age groups (there are 8 types of age classification in the database). An identification of causes of death represented in rows is even more complicated, since database contains information on the last four ICD revisions (7th to 10th) and their modifications, i.e. 12 different lists of causes of death, which were in use since 1950. For example, in France causes of deaths were coded according to version 7A of ICD in 1950-1967, 8A in 1968-1978 and version 9B in 1979-1999; since 2000 it is a 4-digit code of version 10. At the same time, periods of use of one or another ICD version differ from one country to another. Therefore, to use effectively this database, one has to get an access to the source file and to know perfectly its structure as well as the codes of countries, causes of death, age groups etc. Finally, one needs to have certain computer skills in order to extract the information from the database. Regular update of information makes this task even more difficult.

3. The aim of our communication is to present the IT system providing the opportunity of exploitation of the WHO IMD to everybody who has an access to the Internet and is able to understand English, French or Russian. There is no need for special knowledge in neither programming, nor special software (except of Internet browser) or huge hard disk in user’s computer. The system facilities allow calculating, in automatic regime and real time, the basic demographic indicators necessary for the international comparisons, such as age-specific and standardized mortality rates, as well as the general and cause-specific life tables.

4. In the future perspective, the development of the given approach would allow to create a global expert system of demographic information which would provide a quality control of the benchmark demographic data based on the international comparisons and some usual statistical analysis. In practice, participation in such a global networking would allow governments and policy-makers to define the priorities in improvement of their national systems of the population information.