European Population Conference

Liverpool, UK, 21-24 June 2006

Topic 9: Data and Methods

"Internet and Multimedia Demography: What's New?"

Eugeny Soroko
Center for Demography and Human Ecology, Institute for Economic Forecasting, Russian Academy of Sciences, senior researcher; Demoscope Weekly, site administrator

1. Demographic databases on the Net

Current status of Internet as the tools for demographic data dissemination is well known. A lot of international institutions utilize web-servers for storage and granting access to huge amount of various demographic data.

During the last years many databases were opened and extended on the Web. They include population statistics, mortality indicators, census data, etc. To safe the place, here they are not given, use a reference, for example, Sources of data for the Information-analytical system, placed on the page of Demoscope Weekly http://demoscope.ru/weekly/iase/sources.php .

In dealing with such databases many questions arise. Some of them are the common ones for Internet; some are specific for population studies.

The first and the most frequent one is the problem of search. Where and how to find just the data we are looking for? Are they recent? Are they reliable? Are they precise? Do they cover the period, country, age, indicator, etc. we really need? Unfortunately this one is the universal one to web-search irrespective to our field of demography. However, the task to provide demographers with the web-guide to the most valuable and useful resources is an urgent need. I consider as a very fruitful the suggestion to organize a special session at the European Population Conferences, that includes all the problems included in this presentation and to discuss them.

The second question is how the data found are presented? We meet here several aspects. One is the data format. Is it really comfortable for the user to find some data on a html page, some in pdf file, some – as a plain text, another on a jpg image of graph? Unfortunately no standards for demographic data presentation on the web were developed yet, we have no criteria of convenience we

require to solve that problem. And what is the result? Very often we spend one minute to find some indicator and a dozen to copy-and-paste and re-format what was found. Another side of this problem is the number of decimal places used for the indicator's value. To illustrate such questions, let us consider a well-known 2005 World Population Data Sheet of the Population Reference Bureau (http://www.prb.org/pdf05/05WorldDataSheet_Eng.pdf). This pdf file is very good to be printed, but if you need a set of figures?… In this case the user needs to convert and format the data from that pdf. No doubt, PRB has them as a spreadsheet, has not it? Another question is the values of total fertility given there. The number of decimal digits is only one, thus we have as many as 17 (!) countries and regions that have "equal" TFR value 1.3. Does this correspond to the reality? And why did they do this?

One rather new problem for the databases on the web is the volume of information. Some of them have hundreds megabytes, and some – gigabytes. What to do with them? No question arises if the user really needs all or almost all data to download to his computer. However, the problem arises when the user needs some figure based on the data obtained from the database, but directly absent in it. In this case after downloading the initial data he must make some calculations. There are several aspects in this situation. The first one is whether or not the figures received are well formatted for further data processing. Unfortunately many databases do not give opportunities to obtain the data found in the form or formats matching the needs of data processing on the user's computer. The second point is that sometimes the algorithms of calculation are not so simple like summation, ranking, or e.g. averaging. This may be life table calculation, population projections, standardization, etc. The user needs some worksheets, macros, or another computer software. In its turn this software must be installed, or developed, or at least explored. And thus the user needs some time to do this. Very often the calculations required have big volumes, technically monotonous, non-creative. A great concern of the user may be the case when the data from which he needs to produce only one final figure is huge, and no software required is at hand.

What are the ways to solve the situation described above? The main idea is to provide the database itself not only with the tools for data search and extraction, but with the instruments of data processing on the server side of the database. This includes as a minimum the software for such processing and user interface to control this processing (to specify type of calculation, its parameters, output format). This idea realized for Kannisto-Thatcher Database on Old-Age Mortality (http://www.demogr.mpg.de/databases/ktdb/) was performed at the Max-Planck Institute for Demographic Research a couple of years ago. As to the relations to the database the features equipped serve as a superstructure. As to additional user's opportunities this part of the database may be called "Analysis Toolkit". In more details it is the topic of separate presentation.

What are the main advantages of this toolkit? 1-The user does not need to download to his computer the data from the database that

are “raw” from the point of view of the final figures required. 2-He does not need to develop, or search and install, and/or study and utilize some software. It is being used on the server side by the system. 3-The user obtains the final result required in the format specified: as a set of figures, image of diagram, or file. 4-Significant feature from the point of view of demography science is that algorithms the different users utilize from all over the world are the same since they are located at the database server and guarantee equal results obtained for equal data arrays.

How does it really work? You may see it at the fragment of Kannisto-Thatcher Database on Old-Age Mortality placed at Demoscope Weekly: http://demoscope.ru/weekly/iase/ias04_kt.php?terr=1&ind=44. The information-analytical system gives the opportunity to calculate "on the fly" the life table and other indicators based on Old Age Mortality Database: life expectancy, probability of death, population size, median length of life, and median age of population for selected countries. For the first 4 indicators two age interval parameters are required – the first and last ones. Data aggregation by countries may be performed: 1) using summing up of population and death counts, death probabilities are defined from these aggregates for countries, called as "summing up", 2) using averaging the death probabilities for selected countries for each year, sex, and age. One example of such query results is at http://demoscope.ru/weekly/iase/ias06_kt.php?yb=1980&ye=2000&ltw=1&x1=80&x2=130&ind_kt=5&Submit22=OK&terr=1&ind=44&c2=1&c7=1&c11=1&ye1=2002, where the aggregated life table for Denmark, Norway, and Sweden (1980–2000) was calculated via summing up and displayed.

The author will thank all the readers who can suggest new features of current demographic Internet and the urgent problems of web demography to discuss, that may be sent by e-mail to soroko@demoscope.ru.

Inquire the additional details at http://demoscope.ru/center/so/epc2006.html.

2. Five years of Demoscope Weekly

In January 2006 web demographic newspaper Demoscope Weekly accomplished five years. Just before this data the millionth visitor browsed it. Actually the most of this period it is a hybrid of biweekly academic periodical, popular science journal, and a news bulletin as in a newspaper. The number of issues published exceeds 230. Exposition of Demoscope here is based on the presentation by Anatoly Vishnevsky, its editor-in-chief, he made at the MPIDR in 2004.

The editorial board considers DW as a bridge between the scientific community and the General Public. The main goals of the sire are: 1-To lay stress upon the population problems in the public opinion, 2-To discuss these problems not from time to time in connection with some special events (population census for example), but regularly, as it is common in the case of the problems of economics or politics, 3-To put this discussion under the supervision of the researchers themselves but not the journalists, frequently ignorant and/or politically engaged.

General public for Demoscope Weekly is the intellectual elite. It is not designed for mass audience. "General public" for DW is the people who form today or will form tomorrow the intellectual climate of the country - journalists, the academic sphere, university lecturers and students.

The number of unique DW monthly visitors exceeds now 30 thousand. In 2005 people from more than 100 countries visited it.

How DW visitors use its materials? According to our survey, 55 per cent - in research, 30% in teaching, 16% in development of proposals in social policy, 14% in applied calculations, 14% in preparing material for mass media, 10% in students work, 55% for outlook broadening.

What are the sections of DW? By now they are about 50 and include: Front page, Demographic barometers, Russian news, Contents of the issue, Eurasian news, Cover story, Newspapers write about…, Announcements, Voting, Census, Reading-room, Read books and periodicals, Population projections, Take care of women!, Profession – researcher, Analytics, From the history of demographic thought, What do we know about the fox?, Discussion, Interview, Guestbook, Visiting statistics, Subscription, Archives, About the project, Search, Correspondence, From archive files, Web surfing, Population policy in documents and comments, Translation.

DW inaugurated from the following principles: high scientific level, no political tendency, no forbidden grounds, broad coverage of the subject, high information density, access free of charge, interactivity.

Main topics of DW include: Population size, composition and space distribution; Family and fertility; Health and mortality; Internal and international migration; Population and economy; Population and policy; Quality and standard of living; Population and environment; Employment; Status of women, children, youth, elderly; Population history; History of demography

One of the significant parts of DW is Annex. It is a source of data and instrument for population analysis. Currently it consists of the sections: 40 industrially developed countries of the world (24 population indicators in the form of tables, interactive graphs, including 50-year time series), 15 newly independent states (21 indicators), databases, interactive maps, population clocks, software for demographic projection, basic population indicators for all countries of the world, tables of the USSR and Russia censuses. Statistical annex is continuously being developed, extended, and added new features.

The database of population indicators is periodically updated. However, of course, it is not a specific feature of DW. The principle we try to follow is to extend the set of instruments for applied demographic analysis and to shorten the distance between the demographic data itself and presentation of it in the form useful in the analysis. The examples of them are: calculation of relative indicators (http://demoscope.ru/weekly/iase/ias04_02_05.php?terr=1&ind=40), plotting of graphs "on the fly" (http://demoscope.ru/weekly/iase/ias05.php?tim=0&cou=2&terr=1&ind=4&Submit=OK), drawing of

sex-age pyramids (http://demoscope.ru/weekly/iase/ias05_02_pi.php?terr=1&ind=40&gor=3), ranking (http://demoscope.ru/weekly/iase/ias05_02_09.php?terr=1&ind=47&tim=1&cou=0&sort=3). We extend the list of sources for reliable database refilling on a regular basis as well.

Inquire the additional details at http://demoscope.ru/center/so/epc2006.html.

3. Sketch of the Hypertext Demography Guide

1. Introduction

An ordinary textbook of demography represents usually a book sized about several hundreds of pages. It is equipped with tables, graphs, pyramids, and a list of references. However this form significantly restricts its opportunity to present new methods of research, results of recent investigations, arrays of census and survey data, etc. Most of the mentioned have a computer form and thus could hardly be included in a book. One of the means to remove these limitations could be a development of such a textbook in a rather different form – as a **hypertext multimedia guide of demography**. You could find below the problem description and basic ideas clarifying the requirements to its structure, contents, and development.

2. Problem statement

One can rarely meet a remarkably new textbook of demography satisfying him in all extents. It becomes completely unapproachable for all lecturers and professors. Why is it so? Among the major reasons we could find the following:

1- Volume limitation. A student would not carry in his briefcase a book heavier than a brick.
2- New methods can be hardly satisfactory reflected in a study book since only few people are involved in writing it. Up to now the most textbooks do not cover rather significant fields. Some examples are: stochastic population projections, household structure analysis, biodemography, micro simulation modeling, event history analysis, historical demography.
3- Subjectivity in selection of topics covered in a textbook that is quite natural since the authors may have different research and teaching experience.
4- Extent to which different sections are presented in a textbook may significantly vary due to many different programs of demography courses in various universities. It may depend on the country, faculty, and profession taught.
5- Even the most modern textbook once published becomes soon the outdated one – many new research methods and study results appear every day, and a certain delay lays from an inventory to publishing it in book with educational purposes.
6- The recent decade shows that very often the results of research fundamentally cannot be represented in the paper form. They include for example rather big data files, live images, and sophisticated algorithms available only in the computer form.

3. Ways to solve it

What can be suggested to solve the problem stated? This is a hypertext guide of demography (HGD). Its contents should satisfy the following principles:

1- HGD is a computer-oriented set of texts, tables, graphs, and other components organized as a single whole.

2- HGD may be represented in two variant forms – as a CD, or as a web-site.
3- HGD contains demographic texts explaining basic concepts, theories, terms, etc. in any available forms: HTML, PDF, DOC, or others, divided into appropriate chapters, paragraphs, or sections.
4- HGD texts are equipped with the links referring to another texts, electronic tables, databases, terms, files, and web sites according to the notions used there.
5- HGD uses computer programs for educational purposes.
6- HGD uses arrays of demographic data in the best appropriate form like databases, Excel, or ASCII files. They include both the demographic macro-indicators, and micro-data of censuses, surveys, and population registers.

4. Principles of HGD development

Since the size of HGD assumes to be rather large, it cannot be created at once, and some procedure of its development should be presumed.

1- At first stage some small version may be done. It should contain a set of chapters from some ordinary textbooks. This would serve as a skeleton of HGD. We may call this a principle of **gradual development**.
2- Then several sections starts to be transform into the hypertext ones: if some paper is referred to, it should be added to the HGD; if some computer software or data arrays are described in a text, they are included in HGD together with documentation, test and educational examples. Thus a principle of **entirety** is performed.
3- The work on the system development would need rather big team of authors. It must contain at least editors, writers, web-masters, and moderatos. This team may consist not only of the permanent staff, but also temporary invited persons and volunteers wished to contribute. It is a principle of the authors team **breadth**.
4- Multi-level (multi-program) approach. Naturally the requirements to the level of knowledge in demography significantly depend on the profession taught. Thus for example officers of population registers and civil registration bureaus, physicians, sociologists, and applied demographers need different depth and particularity of their education. To provide this, HGD should presume the following approach: a required nucleus (text of lecture with basic terms, formulas, etc.) supplied by links to more detailed sections, demography classics and history, and other files. Some of the latest may be declared as optional. Let us call this a **multilayer** principle.
5- Only for students or for everybody? HGD should be created to satisfy not only the studying (teaching) process (students, PhD post-graduates, post-doctoral ones, their lecturers and masters), but also research scientists wishing to refresh, deepen their knowledge or glance some new fields. Thus any block needs to be at the maximum frontline of modern science. The **maximal scientific depth** principle must also provide the level of utilizing the demographic knowledge in popular-scientific literature and mass media.
6- Principle of **thematic divisibility**. Due to volume restriction of CD, or to different readiness of topics to be covered it is reasonable to presume issue of thematic block CDs covering this or that field, topic, class of methods or problems, such as:

Preparing, conducting, data processing and analysis of population censuses,
Marriage, family formation and dissolution,
Data collection on deaths and life table calculation,
Biodemography – basic concepts, theories and research results,
Methods of population forecasting, UN and national population projections,
FFS: full set of publications, survey questionnaires and data sets, etc.

7- Language of HGD. To provide the largest size of HGD users, at least at the beginning, it is quite natural to start with **English**. This principle however will not remain later when may be presumed the components (papers, chapters, computer software, etc.) in other languages and translation of the ready ones into them.
8- **Feedback** principle. To continuously improve and widen the HGD contents it is useful to establish email, post, fax feedback where the users may give any remarks, opinions, suggestions, which must be analyzed and taken into account.

5. What is needed for educational purposes?

In terms of the educational process, HGD provides two main stages: **lectures** (systematic exposition of basic terms, methods, concepts, theories and results of analysis) and **seminars** (practical exercises with demographic data, their processing, presentation of results, analysis). What is needed to complete these stages? Surely, **exams**! HGD must also include computer tests allowing the reader to check whether or not and to what extent the knowledge given is well griped.

Except for the properly demographic theories, formulas and data, a scholar often needs to deal with some **supplementary** data. They can include geographic, national and regional social, economic, legislation information that the system should also include. Examples of these may be: reference of countries territorial units; time series of labor force participation, GDP, medical care expenses, etc. at country level; International classification of diseases and causes of death; legislation and practice of national family, dependency, and maternity allowances; and so on.

One another very useful component is a collection of typical students' and scholars' errors, speculation mistakes, mass media bugs that we want to escape.

6. HGD navigation

To provide navigation within HGD several different ways should be presumed.

Basic internal thematic navigation. Gives hyperlinks in the texts to another texts, terms, maps, programs, tables, images, spreadsheets, files, etc.

Pop-up menus and comments, help windows, tooltips that provide reading and moving through HGD more live, clear, quick, and flexible.

Terminology dictionary. Contains thesaurus of all the demographic terms used in HGD, their short description together with cross-references to other related terms, more general and more particular ones ("See also"-like).

Multilingual demographic dictionary (English-German-French-Russian).

Directories of the HGD components. Each type of files and data contained in HGD is supplied by a thematic catalogue and (or) ABC index such as:

- Lectures index
- Maps catalog
- Databases index
- List of computer tests (exams)
- Directory of programs for various demography courses
- Who is Who in demography
- Directory of Internet resources
- Computer software index
- Book and multimedia libraries index, and other if necessary.

Search engine providing an opportunity of looking for information within HGD by author, type, form, content, date, and other descriptors.

7. Conclusions

HGD, draft of which is sketched here, if created, might have the following advantages:

- Main problems stated in section 2 may be successfully solved: Volume limitation practically disappears or spreads by at least a hundred times. Statement of demography becomes rather full and up-to-date.
- The users who begin to study a subject do not need to spend time on searching books and papers in libraries, data in computers or archives, etc. They immediately obtain access to them.
- The user is supplied by both the instruments (theory, methods, formulas, software) and basic materials (demographic data, electronic tables and other files) together with supplementary ones.
- The users directly deal with the "bowels" of the modern science – watch them, as well as try, exercise, train and estimate them.
- Distance demographic education gains additional chances to widen and deepen.
- The demography knowledge becomes better structured and organized adequately to challenges and opportunities of the computer epoch.

One could easily sight a lot of other positive features.

The author will thank all the readers who can suggest any new other features, aspects, and components of HGD, participate in its development, or at least discuss it, that may be sent by e-mail to soroko@demoscope.ru.

Inquire the additional details at http://demoscope.ru/center/so/epc2006.html.