

INTERIM REPORT IR-98-043/June

Population Projections: New Opportunities for Software Development

Sergei Scherbov (s.scherbov@fsw.rug.nl)

Approved by
Wolfgang Lutz (lutz@iiasa.ac.at)
Leader, *Population Project*

Contents

Introduction	1
WWW Technology: A New Direction in Software Development	2
HTML-Based Multiplatform Interfaces	3
How Could a Web-Based Population Projection System Work?	5
Prototype of the Demographic Server Based on Web Technology	6
Use of Demographic Site (Fields of Application)	8
Conclusion	9
Appendix: Brief Glossary of Terms	10

Abstract

The major principles of the development of platform independent software for population projections based on WEB technology are discussed, and a prototype WEB-based demographic server is developed. The possibilities of using the WEB approach for broader issues of demographic analysis and training are discussed. The paper was initially presented by the author at the workshop "Rethinking International Population Projections" organized by IIASA (6-8 June 1996, Laxenburg, Austria). Although almost two years have since passed and WEB technology has become much more powerful, some of the ideas discussed in the paper still await implementation.

About the Author

Dr. Sergei Scherbov is from the Population Research Centre at the University of Groningen, the Netherlands.

Population Projections: New Opportunities for Software Development

Sergei Scherbov

Introduction

The way population projections are performed computationally has changed considerably during recent decades. Many of these changes occurred due to overall developments in computer and software in general. People who are able to make projections have also changed. Two decades ago usually only the developers of the software were able to make the projections, or at least such a person always needed to be present. Most of the projections were performed “through” the programmers unless they were performed using a hand calculator and paper. Some 20-25 years ago many “non-specialists” in computers were simply afraid of computers. The reason for that was that to run a program, one should have some knowledge of the operating system of a particular computer that was mostly command driven and far from being user friendly. Knowledge about computer peripherals was also required.

Dissemination of programs between different types of computers or even computers of the same type but running different operating systems (for example UNIX and VMS on VAX computers) was impossible without source codes. But even then it was limited because source codes were dependent upon a particular operating system of the computer and availability of a compiler. Most of the peripheral devices like plotters, printers, etc., were incompatible and required a lot of codes to be additionally programmed or reprogrammed.

With the arrival of personal computers in the 1980s and the development of user-friendly software and graphical user interface (GUI) the number of people who were willing and able to use computers increased dramatically. Text processors and spreadsheet programs were among the most widely used applications within the scientific community. That was partly due to user friendly, simple interfaces. Dissemination and accessibility of the programs became much easier: programs developed for PCs running DOS could be executed on other PCs running the same operating system. But still there were problems in using peripheral devices, screens, etc.

The next step was the appearance of Windows which became the most popular GUI environment. Windows made the programs developed by the programmers no longer dependent (or at least less dependent) on a particular configuration of computer. The Windows operating system took care of most external and internal devices. Program developers could concentrate more on the substantial parts of their programs. But still it was impossible to use the existing codes on computers running other operating systems: applications written and compiled on a Macintosh cannot run on a Windows PC or UNIX. Usually transferring a program from one computer to another requires skills in the use of computers. Additional programming to adjust

the program to the available operating system was required even for programs written in high level languages. For the end user, fewer computer skills were necessary once the computer was running on an operating system with GUI.

Another burst in computer usage came at the beginning of the 1990s with the development of the World Wide Web (WWW) which linked Internet resources through a system of hypertext. Development of user-friendly, graphical interface Web browsers, simplicity of their usage and availability for almost all existing computer platforms led to almost universal social acceptance of the Web.

Further development of Web technology brought tremendous opportunities beyond simple document retrieval: from searching libraries, databases and informational resources on the Internet to real-time weather observation, on-line mortgage calculations, delivering multimedia information, etc.

This chapter will address the question of how to use Web technology for the needs of demographers and researchers from other fields, not limiting them to browsing documents, but also using demographic models and software, and in particular, how such technology could be applied to perform live population projections and present the results on the Web. There is definitely a need for that. Potential users of projections and projection models are not bound to the demographic community. That is what makes population projections so unique. There is great interest in this subject in many other fields. Specialists in social, political, environmental and many other sciences need to have access to data on population projections and may even wish to experiment with population projections, testing their own assumptions.

In the following sections we will briefly introduce the current Web approach that is used to provide the user with static and dynamic information taken or produced at the Web site; summarize some of the major advantages and disadvantages of such an approach; present a structure of demographic site based on modern Web technology; describe major steps required to make projections using this technology; and discuss possible use of demographic sites. It is also important to mention that such sites have not yet been built (although many components have already been created or tested). The technology is there; we have to use it for our needs.

WWW Technology: A New Direction in Software Development

Today the World Wide Web is the most popular example of distributed information and electronic publishing. It is based on the concept of hypertext--documents linked to other documents, allowing us to follow related topics from one place to another. Instead of being restricted to one file, the documents can be spread across files, computers, companies, and countries. The initial purpose of the Web was to publish hypertext documents in a graphical format. Today Web browsers are taking many functions from computer-operating systems.

To function, WWW must rely on three components: the user's computer, network and other computers which are acting as servers. Invisible to most users, Web servers take most of the burden of the Internet. Each Web page that follows Internet traffic comes from a Web server. Without Web servers, there would be no Web.

Web servers operate in the following way. Web browsers communicate with Web servers using Hypertext Transfer Protocol (HTTP), a simple request/response protocol for information transfer using Transmission Control Protocol/Internet Protocol (TCP/IP). The Web server receives the request, retrieves the file, sends it to the browser, and then tears down the

connection. The graphics in the page are serviced the same way. Browsers then display the downloaded HTML document.

Although Web servers usually store HTML pages, they can store and serve any kind of file, including text or data files. Today, with the exception of forms filled out by users, most Web traffic is one-way--browsers read files from the Web server. But there is also a way to write files to a Web server. This allows Web servers to process the user's data that are stored in a file while additional parameters could be collected using standard form approach. Of course in this case responsibility for the correct structure of the submitted file lies with the user. Filling out the forms can always be made more protected against errors.

The tremendous growth of the World Wide Web, in both Internet and intranet environments, has led to innovative techniques for extending the capabilities of Web servers. The information collected from filled-out forms or received files may be processed by the server using special scripts and programs written by Web developers. Users who run browsers may link into databases, display real-time information, place orders, and perform any number of functions that Web servers by themselves normally cannot do. Thus Web servers can run applications and provide the user with dynamic data as well as static text and graphics. And these are the features that will enable the use of demographic software in general and population projection programs in particular on the Web.

Today one of the most common Web applications providing dynamic data are database search engines. An excellent example from the field of demography is the database search of the International Division of Bureau of Census (<http://www.census.gov/>).

There are several ways to enable Web servers to provide dynamic data. Standards like the Common Gateway Interface (CGI), Internet server application programming interface (ISAPI), and scripting languages like JavaScript and VBscript, as well as full programming languages like Java (Java runs disconnected from the Web server) and Visual Basic, are being used for this.

CGI is a specification allowing Web servers to run external applications that generate HTML documents and send them back to the browser. CGI programs can interact with the user's browser, accept data, and even communicate with an external resource server such as a database server. Developers can write CGI scripts in almost any language native to the Web server, including the popular Perl language, C, C++, Basic, Pascal, or Windows CGI. Thus many of the previously written programs may be re-used.

HTML-Based Multiplatform Interfaces

CGI programs allow to serve dynamic data. They can react to user input and return a resulting HTML page based on this input. The combination of CGI (or ISAPI) and HTML-based front end makes it possible to create platform-independent interface and software. This means that whatever computer platform is being used (DOS, UNIX, Windows 95, Macintosh, OS/2, etc.) the user can still use a particular program or model that was initially created for another platform. Usually platform independence is achieved by using expensive cross-platform compilers and expensive cross-platform GUI libraries and software packages (for example, ZIN Application Framework). Using the HTML approach, the developer does not need to write complicated codes for drawing windows, generating forms, handling input fields, etc. Instead he can write very simple statements using HTML and the job of generating windows, buttons, inputting fields, etc., will be completed by the Web browser. This allows the developer of the software to concentrate more of his efforts on making the application.

Usually HTML-based interface contains two major parts: a form to be filled out by the user and sent to the server, and a program or programs that will process the data collected from that form and send the results of computation back to the user's browser as another HTML document.

Web browsers exist today for virtually every operating system. Using HTML as a front end will give the developer an opportunity to use his software on almost any platform. If an executable code could be produced for the computer running the server, the results will be reachable from any other platform for which browsers exist.

What does the combination of Web technology and HTML-based front end provide? Web-based HTML interface running an application has a number of advantages for the users of an application over a traditional approach of executing a program on a particular computer with a particular operation system.

Easiest and most popular interface. HTML-based interface is widely used, easy to use and known to many Internet and intranet users on any operation system. People are not prejudiced towards this interface. It is socially acceptable. Most of the users do not need to get accustomed to it. And most importantly, up to now most browsers are free.

Platform independence. Applications that are developed for a certain platform can be accessed from any other platform for which a Web browser is available. For example, the user working with a Macintosh can access applications developed for DOS. The only thing that is needed is a Macintosh computer attached to Internet or intranet, and a web server running on the computer capable of executing DOS applications. The fact that Web browsers and servers now exist for virtually all popular operating systems makes the application universally accessible.

Flexibility in presentation of results. The results of computations are returned in an HTML-formatted document. The results may include formatted text, tables and graphs that can be either directly printed from the browser or copied to the text processor, updated, and then printed. Large amounts of data may be sent to the user in the form of compressed files.

Provision of help information. Documents sent to the user may include links to help information concerning particular applications, data, results produced by the application, etc. Through HTML links, the user may learn more about a particular subject without breaking his connection to the application.

Spatial accessibility. Applications as well as help information may be distributed anywhere across the WWW. Execution of applications available at a site in Europe may be initiated in any part of the world.

One of the **disadvantages** of using Web-based approaches compared to conventional ones may be slow access to an application during peak hours.

From an application-developer's point of view, the Web approach also has a number of strong points. As mentioned earlier, there is no need to program GUI or to use expensive GUI-generating software packages. Some of the limitations in selecting controls that were used to build an interface a few years ago (compared to software developed using programming languages like C or Visual Basic) no longer exist. Many of the interface controls are already built in the browsers; they are easy to invoke; and more of them will be coming in the future with the arrival of tools like Java and ActiveX controls. There is also no need for expensive cross-platform compilers.

Advantage could be taken from the fact that resources may be **distributed** on the Internet. For example different tasks may be accomplished by different servers that may be

distanced from each other by thousands of kilometers. And if access is fast, this will not affect the user.

The user may be provided with access to an application without actually having it on his computer. Thus there will be no need for complicated dissemination procedures and later losing track of a program. This is extremely important if there are new releases of application programs due to an improved methodology or updates due to discovered bugs. Any updates in the software reach the user immediately. Improved feedback between software developers and the users may also contribute to more efficient approbation of the software. Comments and suggestions on software improvements may immediately reach developers.

Restricted access to the use of certain applications and data sets as well as the creation of commercial sites are also easy to establish by simply forbidding access for non-registered users through a system of user names and passwords.

Taking into account that potentially any Internet user may have access to any Web server presents a challenge for software developers, in particular when the programs they develop create dynamic data as a response to a user's request. As mentioned earlier, the use of demographic sites will not be limited to demographers. Specialists in social, political, environmental and many other sciences may be especially interested in demographic projections, which makes it even more important to put a lot of effort into developing error-protected and user-friendly software.

How Could a Web-Based Population Projection System Work?

Figure 1 presents a simplified diagram on how projections may be performed on the Web. First, we briefly describe the general approach, and then take a more detailed look at what stages of projections are involved and how they may be implemented.

The following steps list the sequence of interaction between the client (user browser) and a server.

The user sends a request to the server for a document that initializes the communication by simply providing a certain URL address. The server responds to the request directly by sending an HTML document(s) or a form(s) to be filled out. Parameters for population projections are collected from the user via these HTML forms. After the form is filled out the client submits it to the server. The server determines that the request is not for a document it can deliver, and creates a CGI process. In a special manner the parameters from the forms are passed to the gateway programs. Using these parameters, the gateway formulates a request for population projection to a demographic server. The demographic server passes a request to the projection program, and requests information from the database if necessary. Before the projection is performed, the demographic server may request certain additional calculations to be performed either with user-supplied data or with data received from the database. For example, some additional procedures may be applied to these data. After the projection is performed, the projection server returns the results to the gateway program, which formats the results into an HTML file and passes the name of the file to the Web server. The Web server then passes the HTML file containing the results of the projection back to the Web browser for analysis by the user.

This scheme allows Web browsers to serve as platform-independent front ends to live population projections. This approach is now used widely for database publishing on the Web.

Results of the projection may also contain different comments as well as descriptions of the methodology used, in the form of hypertext.

Prototype of the Demographic Server Based on Web Technology

Let us now follow each step that is usually involved in making population projections, and see how it can be implemented using HTML-based approaches. In Figure 2 a typical sequence of steps to perform a projection is presented. We will briefly discuss each step.

The session starts when the user requests from the demographic server¹ an index or home page. This is done by directly entering the URL address, clicking the link from some other document, or using his bookmark. The demographic site sends the client its index page which provides options to select different services. Performing a population projection is among them. Once this option is selected the user may learn more about the subject by clicking on links that refer to different documents with help information. From the same menu the user may select an option to use data available at the server side or to make projections using his own data.

Data input. Most of the parameters for the model could be entered using the form interface. If the user decides to perform projections using his own data he may receive a form similar to the one presented in Figure 3. After he fills in parameters, as presented in Figure 3, and probably answers several more questions that will allow specification of what type of data is available, the server generates an input form to be filled out that may look like the one presented in Figure 4. Another option may be to submit the file with all initial data required for projections directly to the server in an exact format requested by a projection program (the program may be selected later). In this case responsibility for the accuracy in the data structure lies completely with the user. If the error occurs due to incorrect structure of an input file the user may be informed of this and may be given some ideas of what went wrong in the file. Information on the structure of input files should be available from help files. This option may be a choice for an experienced user.²

If the user decides to use data available at the server side, he may receive a form similar to the one presented in Figure 5. Data sets may be selected from a scrolling list. After the user submits a form he may receive additional information about the data he selected, such as the number of age groups, the number of regions (if data correspond to a multi-region system) and some other parameters that characterize the data set, including the sources of information. The user may always return back to the Data Selection Form. On the same form the final projection year is selected.

Formulation of assumptions or scenarios. After baseline data have been defined the user receives a form similar to the one presented in Figure 6. By the time the user receives this form, he has already defined baseline data and the name of the region(s) he wants to work with. Thus he may check if there are scenarios available at the server database for this region. If the baseline data file was picked up from the server, the chances will probably be high. If scenarios

¹ The demographic server, together with other servers, may be a part of a certain Web site. We call it a demographic site to stress that it serves the purposes of demographic analysis. But because most of the demographic functions are served through the demographic server, we will sometimes (if there is no confusion) not make a distinction between demographic server and demographic site.

² In certain cases baseline data on demographic processes like fertility and mortality may not be required, if the user wants to use scenario variables that do not depend on initial model rates.

are available the user could learn more about them. If baseline data were supplied by the user, then he will most likely have to define his own scenarios.

By selecting one of the predefined scenarios for fertility, mortality, and migration that can be stored in the server database, the user has an opportunity to combine total scenarios out of available ones. In fact such an approach may be the choice for those who might have difficulty formulating scenarios in terms of demographic variables. A mixed approach, e.g. a predefined scenario for some variables and a user-defined scenario for others, may also be used. Thus, after studying what is available at the server side, the user fills in the tables as shown in Figure 6, submits it to the server, and receives a new step-by-step form for each scenario variable. Depending upon the previous selection, the form will request either selection of a scenario from the list of scenarios or filling in the actual values for scenario variables with corresponding times.

There are many ways to set up scenarios for fertility, mortality, and migration. For example, a scenario variable that defines the level of mortality may be expressed in terms of life expectancy, rate of decrease in age-specific mortality rates in each age group, change in the values of parameters that define the curve which may fit age-specific rates, etc. If life expectancy is selected as a scenario variable there are again many options to generate age-specific rates: use of model life tables, relational models, etc. This very interesting topic may deserve a separate study which is out of the scope of this paper. For us here it is important that the client fills out the form or forms with the assumption on future developments of fertility, mortality, or migration. Those may be direct or indirect measures. Later they will have to be translated into parameters required by the projection model to actually perform the projections. Depending on the type of scenario variable there may be several interactions with the server.

Scenarios may be stored in final forms that are sent for approval in HTML format to the user by the server. The user may store these files on his computer. The next time he wants to start the projections, he will have to load this file and submit it to the server. All the information will be there. But it is probably more efficient to assign a user identification number and store it on the server side. Each time the user enters the server he will have the option of using his previously-defined scenarios.

Selection of projection model. This step is important for an experienced user, who can understand the differences between the various projection models. Demographers inexperienced in population projections or users from other fields may select the default model. This form should contain comprehensive help information describing available models and the differences between them. Although the next step (translation of assumptions into model rates) does not require interaction with the user, computational algorithms will depend on the selection of the projection model.

Translation of assumptions into model rates. Translation is performed on the server side. Using the information obtained at the previous step, a proper model is used to transform the information on parameters or indicators into projection model age-specific rates. For example if a mortality scenario was defined in terms of model life tables, then a particular table with a particular life expectancy should be generated. Whether age-specific rates or survivorship proportions are later used depends upon what particular projection model was selected, because one model may require age-specific rates and another survivor proportions. Once the data and projection model have been selected, this step does not require interference from the user, and is performed completely on the server side.

Simulation. In general this step does not require any action by the user unless he is informed that a certain error occurred during simulation.

Presentation of results. Depending on the selected projection model, the choice of different indicators to present may be very different. Depending on which projection model was selected, the list of possible output variables may be submitted to the user by the server for selection (Figure 7). After the selection is made, the server provides the user with an HTML-formatted file that will contain requested indicator(s) (Figure 8). It would be possible to provide certain graphic information. In this case the graph is created on the fly by the server and submitted to the browser. The user may request that the results of the projection are delivered as a spreadsheet file (WK1 format).

Use of Demographic Site (Fields of Application)

Research. Several institutions are involved in preparing long-term global population projections today: United Nations, World Bank, U.S. Bureau of the Census, and International Institute for Applied Systems Analysis (IIASA) among others. Here we encounter several problems that potential users of these projections may confront, and try to suggest solutions.

The user may not be satisfied with the set of scenarios that an organization is using. For example, the UN makes only three fertility and one mortality scenarios. One of the reasons for this is that too much information would have to be published and thus the final volume would become enormous in size. Publishing the data on the Web may be a solution to this problem. The most important variants from the UN's point of view may still be published, and the users could be referred to the Web site for the other scenarios. The U.S. Bureau of the Census has an excellent database for their projections, but unfortunately, at least at the time of writing this paper, they provide the user with only one variant.

It is also important to give the user the possibility to repeat projections done by others. In certain cases it may not be feasible (at least today) due to very time-consuming procedures, as in the case of probabilistic projections. But in many other cases it may very well be done. For example, it should be relatively easy to use the approach developed in this chapter to let the user select a country, request the baseline data from a particular database (UN when it becomes available, or U.S. Bureau of the Census, which is available now) and to repeat the projections using the available baseline data and projection program but with a user-defined scenario and time horizon of projection. To do such an exercise for one particular country at a time would not take a lot of computing resources (that is how most projections are done anyway).

The next step would be to allow the user to fill in his own baseline data if he so wishes. If the user-defined data are worth to be stored in a database of the demographic server, with the agreement of the client, such a procedure could be easily achieved.

In fact collecting data or expertise from clients is another area where the Web-based approach may prove to be extremely valuable. Notified experts using restricted access can fill in HTML forms that are sent to the WEB server. This information is stored in a database. Later it can be used to perform conventional or probabilistic projections. A lot of attention should be paid to the design of the form to be filled in, in order to facilitate this process. There should definitely be a lot of help information attached to such forms.

Education. Internet provides great opportunities in education. One of the greatest features of the Web is that it provides almost global accessibility of resources available on different sites in different places. If population data, projections programs and other demographic resources would become available on the Internet, this might considerably contribute to facilitating demographic training at schools, universities and different training programs. The teacher and students would no longer confront the problem of compatibility of training software and available hardware, installation of software and database. This is

especially important for the organization of (UN-type) short courses. Of course population projections may constitute only one part of many different topics for training of demographers with the help of WWW.

Parts of training programs may be developed by different collaborating centers physically located in different regions or even countries. The use of hypertext links could bind their documents and produce a merged course that teachers in all institutions could use in their work. Such courses could have public access or limited access. Intranets may be used to limit access at the organizational level.

General use. As mentioned earlier, specialists from fields other than demography may become frequent visitors of a demographic site. Many government, international and national organizations around the world use population projections in their activities. Specialists from social, political and environmental sciences may also contribute to developing demographic scenarios.

Planning. Statistical offices and people who are involved in national population projections, especially in the most developed countries, may also become interested in using projections from demographic sites, although most likely for reference purposes. Usually statistical offices use their own programs to make projections on a very detailed level. Often these programs use specifics of national data.

In this paper we discussed mostly the application of population projections. But a demographic site may contain many other tools for demographic analysis. The United Nations Software Package for Mortality Measurement (MORTPAK) is a candidate to become available on the Web. Figures 9 and 10 show the type of interface that can be used to work with MORTPAK. In this example, the user has selected the program MATCH (Figure 9); the form to run the program has to be filled out (Figure 10). The output is on Figure 11.

Conclusion

Web technology changes extremely fast. By the time this paper is published, there will be many more new tools enabling more efficient use of demographic software. In this paper we discussed only one of many possible approaches to use demographic software and models on the Web. We did not touch, for example, the approach of using Java language. Many of the demographic models and software could be developed in Java language; Java applets, rather than being executed on the server side (the case of CGI), are sent to the user and executed on the user's computer. Having the user do some of the processing minimizes the amount of traffic between the user's computer and the servers on the Internet, resulting in quicker access. This new technology is becoming wildly popular.

Demographic developments attract more and more attention from society. The reason is clear. Too many areas are affected by future changes in population size: environment, food, social security, etc. It is the responsibility of demographers to give people a chance to learn more about the current status of the population and its future growth. All the necessary technology to do this is now available. Demographers should use this opportunity.

Appendix: Brief Glossary of Terms

ActiveX The new name for the Object Linking and Embedding (OLE) technology that was developed by Microsoft and is part of the Windows operating system.

CGI (Common Gateway Interface) A standard that allows Web servers to run external applications such as complex database searches, custom Web page construction, etc.

FTP (File Transfer Protocol) An Internet protocol that allows a user on one host to transfer files to and from another host over a network.

Home page The first page of a Web site or of a logical group of HTML documents.

HTML (HyperText Markup Language) The language with which World Wide Web documents are formatted. It defines fonts, graphics, hypertext links, and other details.

HTTP (HyperText Transfer Protocol) The protocol that negotiates document delivery to a Web browser from a Web server.

Hypertext A way of presenting information in which there are links from one document to another. In a Web document, the link is a URL pointing to another Web page or other resource.

Intranet A private network that uses Internet software and standards.

ISAPI Internet Server Application Programming Interface.

Java An object-oriented language, developed by Sun Microsystems, for writing distributed Web applications.

TCP/IP (Transmission Control Protocol/Internet Protocol) The collection of transport and application protocols used to communicate on the Internet and other networks.

URL (Uniform Resource Locator) A standardized character string that identifies the location of an Internet document.

Web browser Client software that requests and displays HTML documents and other Internet or intranet resources.

Web server A server that stores and retrieves HTML documents and other Internet or intranet resources using HTTP. Also called an HTTP server.

Web site A collection of web pages. A web site contains a “home page” as well as a set of other web pages that can be linked from the home page.

World Wide Web (Web) The Internet’s worldwide, HTML-based, hypertext-linked information system.

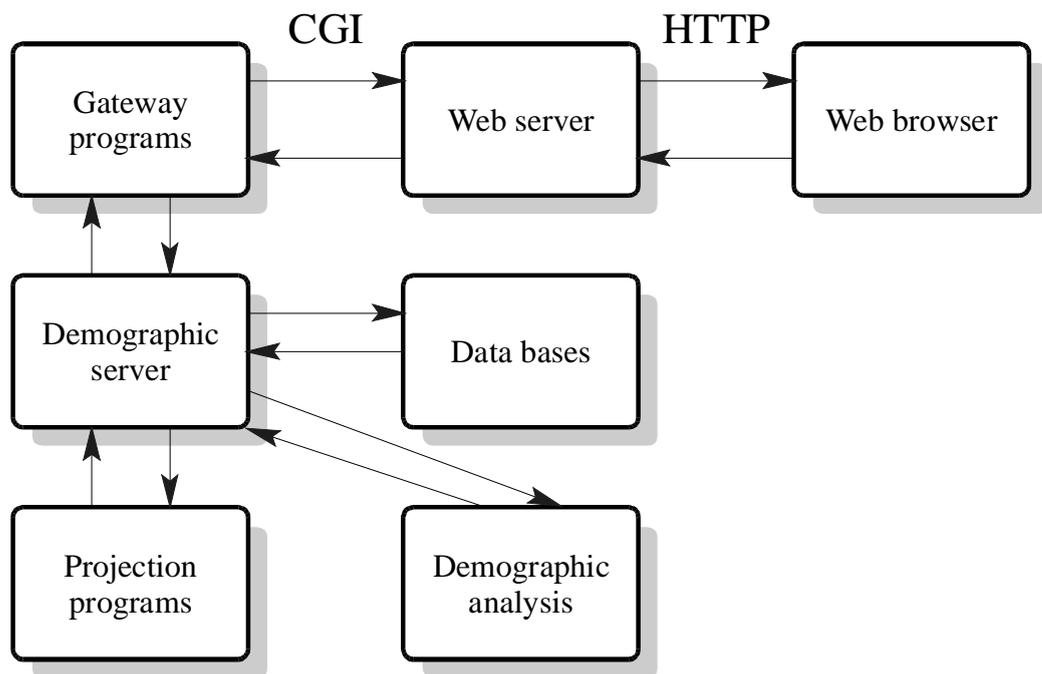


Figure 1. Major components of the demographic site.

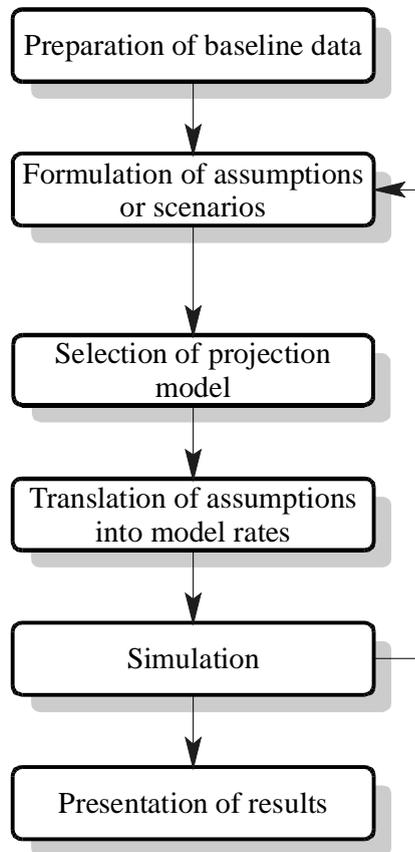


Figure 2. Steps in making population projections.

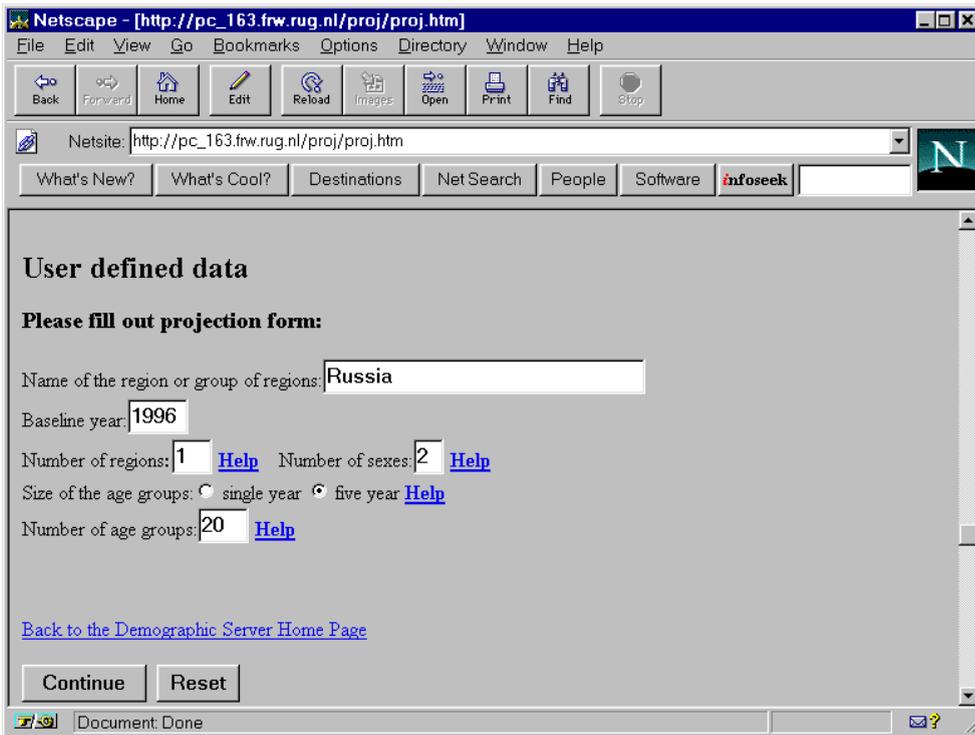


Figure 3. Form for entering basic parameters of the model.

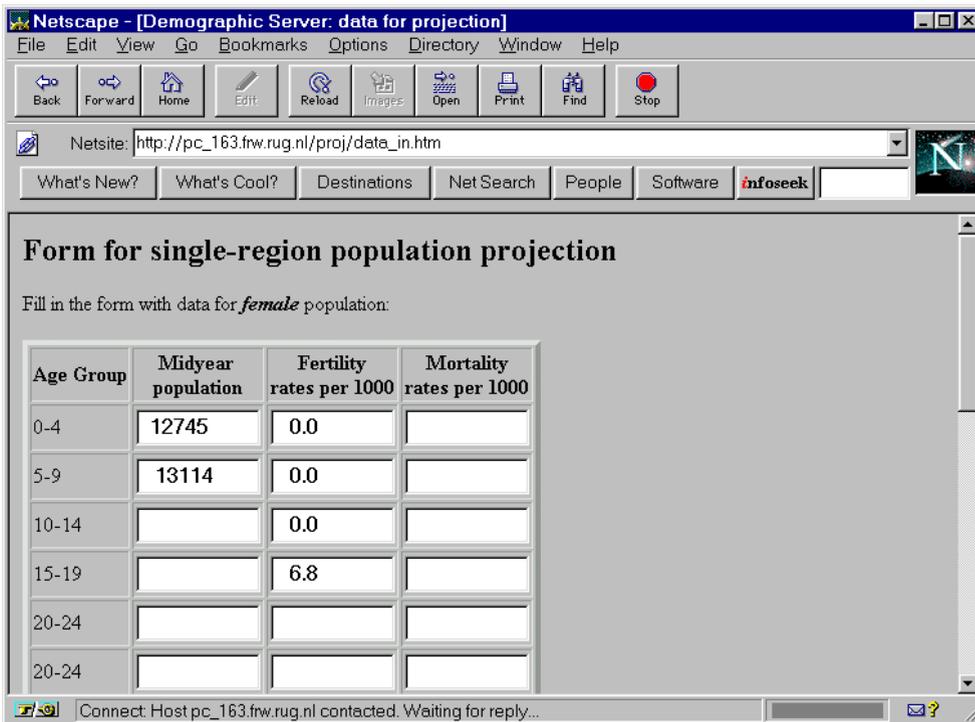


Figure 4. Form for entering baseline data.

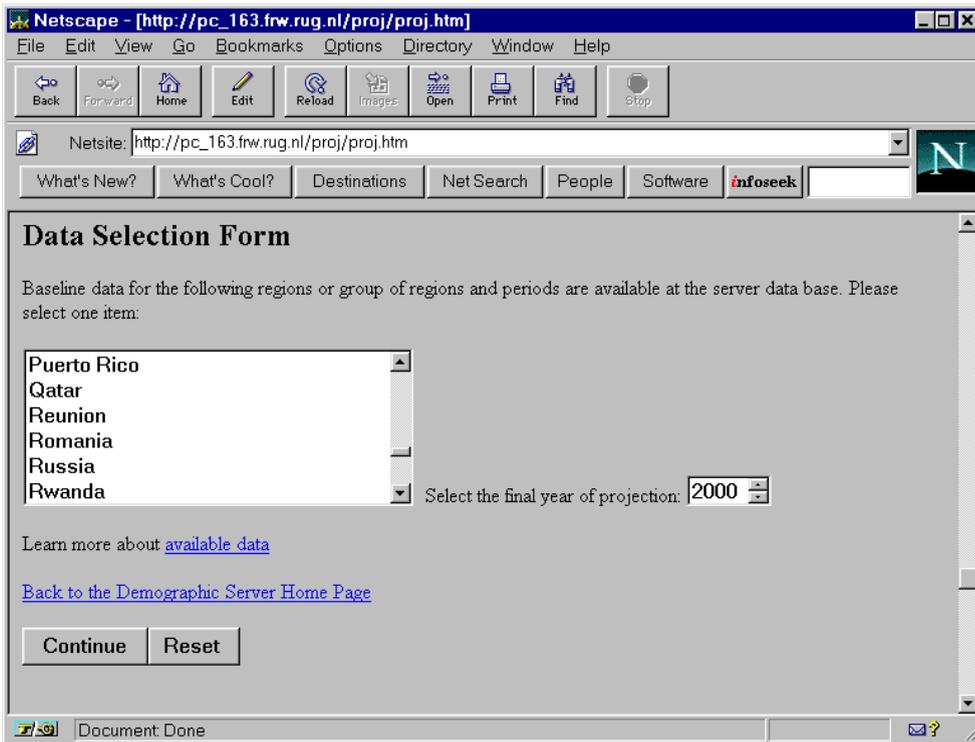


Figure 5. Selection of existing baseline data.

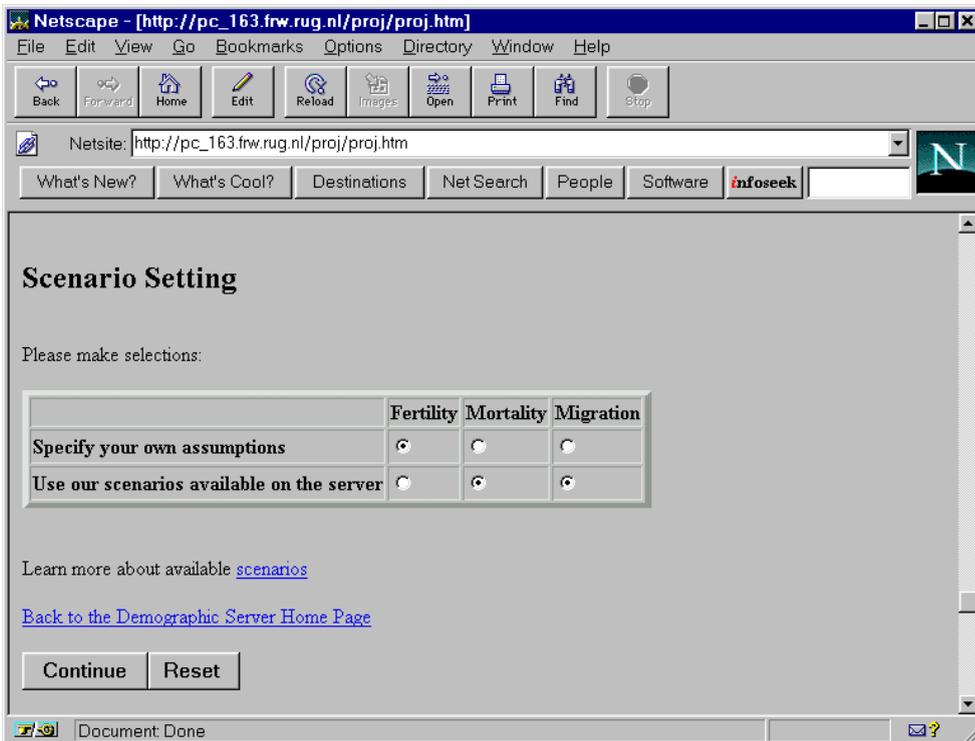


Figure 6. Scenario definition form.

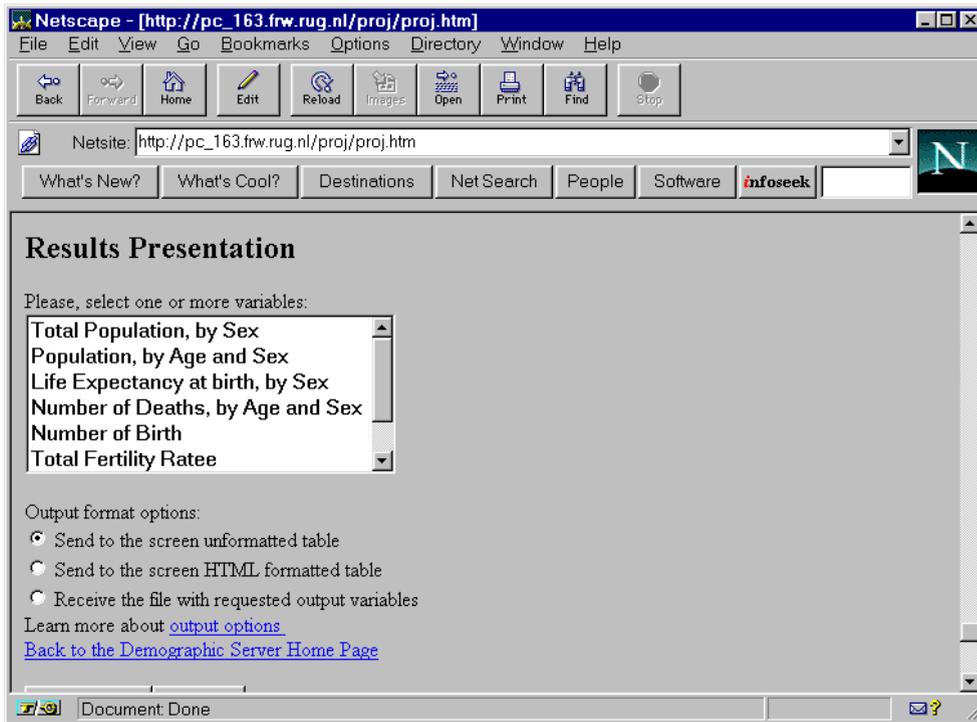


Figure 7. Selection of output variables.

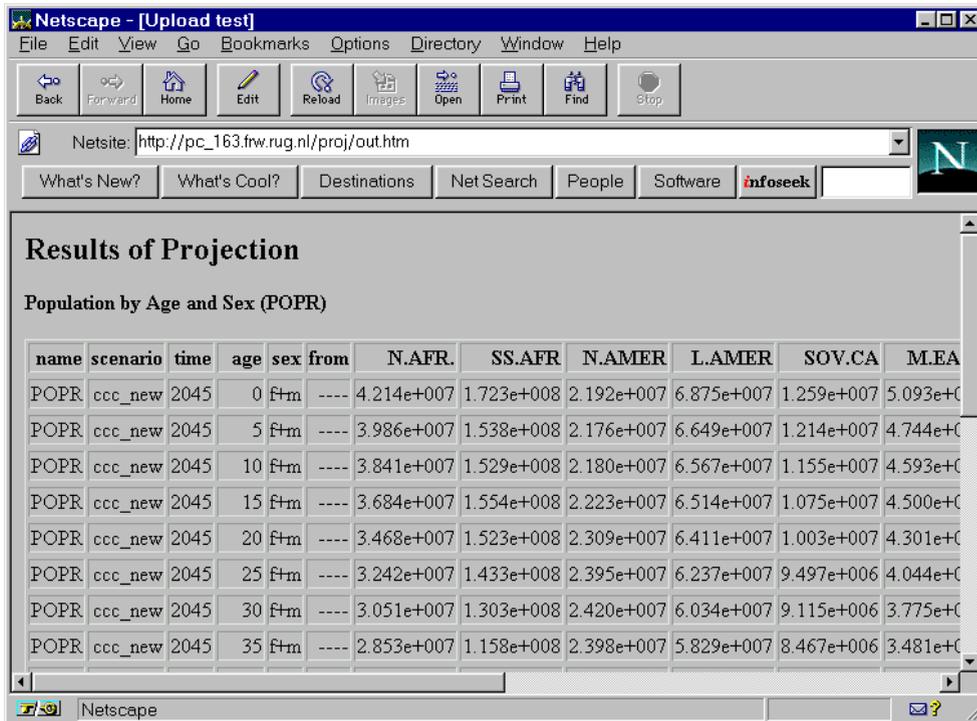


Figure 8. HTML formatted output results.

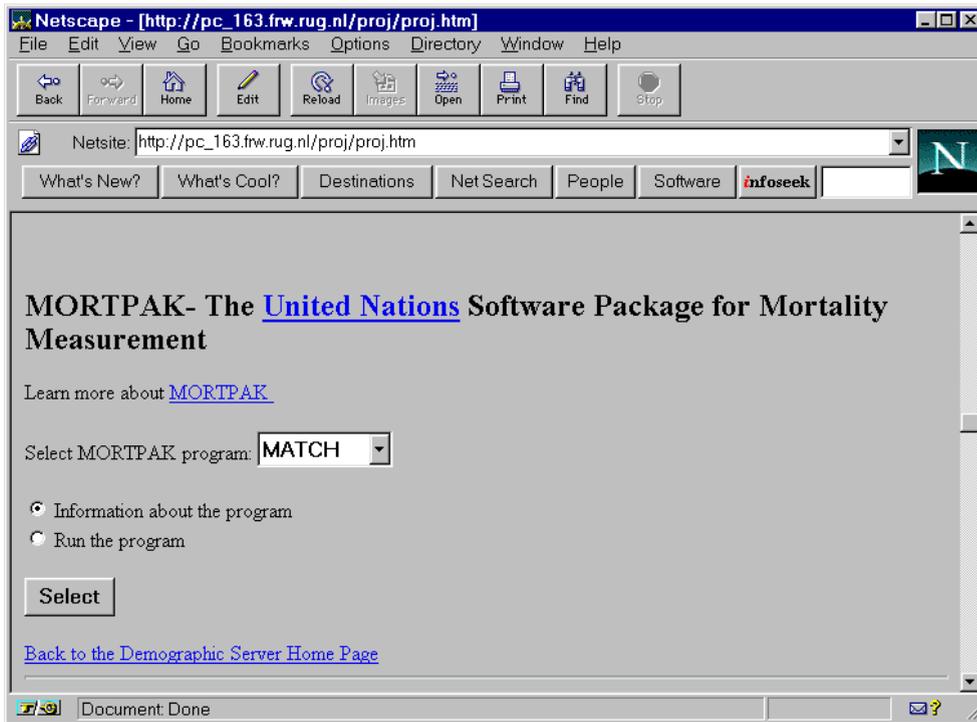


Figure 9. MORTPAK procedure selection form.

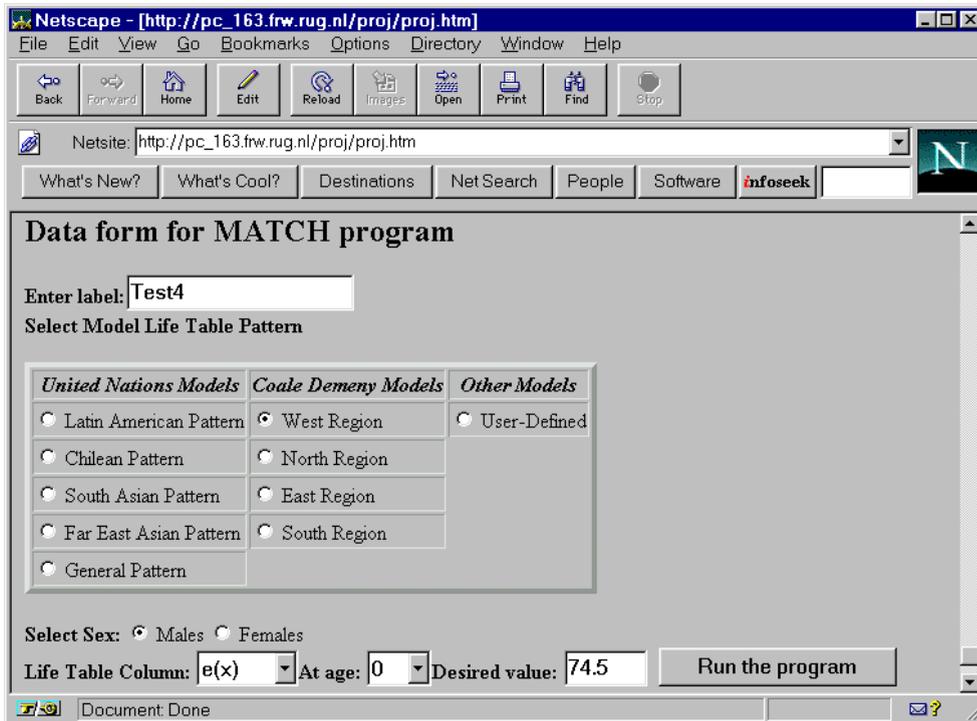


Figure 10. Form for entering basic parameters of MATCH program.

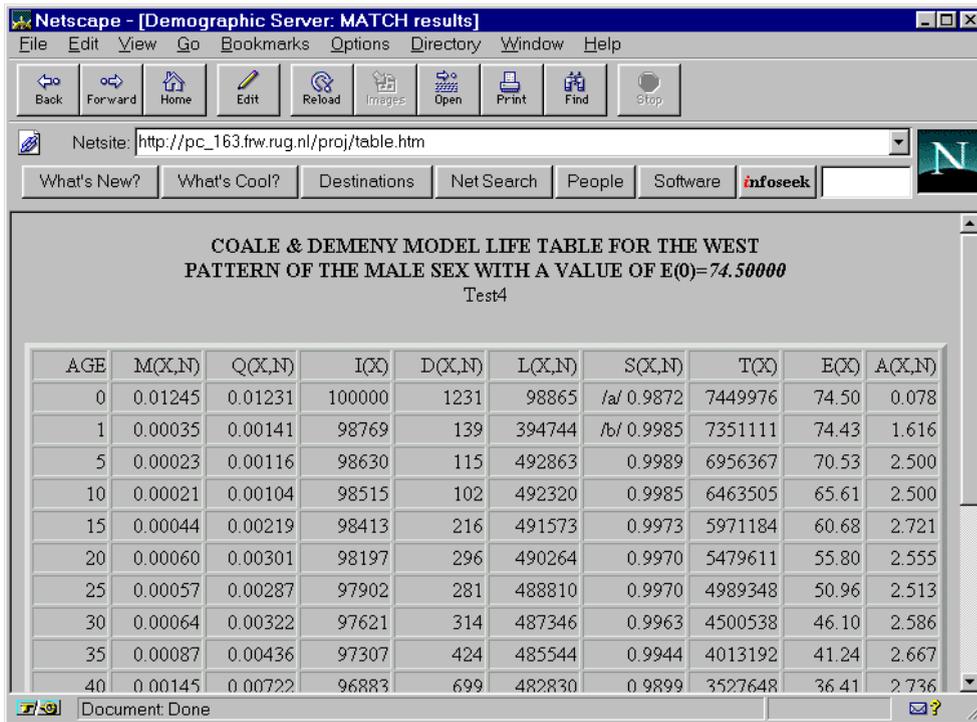


Figure 11. Output of MATCH program.